Docket #0608 Date Filed: 9/11/2023

UNITED STATES BANKRUPTCY COURT
DISTRICT OF MASSACHUSETTS

In re: TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Reorganized Debtors.	Chapter 11 Cases 14-40987-EDK 14-40988-EDK 14-40989-EDK Substantively Consolidated
STEPHEN B. DARR, TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Plaintiff, v. FRANZ BALAN, A REPRESENTATIVE OF A CLASS OF DEFENDANT NET WINNERS, Defendants.	Adversary Proceeding No. 16-4006
STEPHEN B. DARR AS TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Plaintiff, v. MARCO PUZZARINI AND SANDRO PAULO FREITAS, REPRESENTATIVES OF A CLASS OF DEFENDANT NET WINNERS, Defendants.	Adversary Proceeding No. 16-4007

OPPOSITION OF STEPHEN B. DARR, AS TRUSTEE OF THE ESTATES OF TELEXFREE. LLC, TELEXFREE, INC. AND TELEX FINANCIAL, INC., TO CLASS REPRESENTATIVES' MOTION TO EXCLUDE TESTIMONY OF DR. CAMERON FREER AS INADMISSIBLE UNDER *DAUBERT*

1440987230911000000000014

TABLE OF CONTENTS

I. BACKGROUND 1

A. The Freer Aggregation Methodology 7

1. Step 1: Selecting for Data Quality and Transforming the Fields 7
2. Step 2: Cleaning the Fields 8
3. Step 3: Deterministic Step 8
4. Step 4: Blocking 9
5. Step 5: Rubric Application to Calculate Gammas 10
6. Step 6: Calculate Match Probabilities 10
7. Step 7: Compute Clustering for Several Fixed Thresholds 10
8. Step 8: Dynamic Clustering 11
9. Step 9: Convert the Clustering of Account Groupings into a Clustering of User Accounts 12
10. Step 10: Tabulate Total Net Equity for Each Cluster 12
11. Results 13

II. DR. FREER'S AGGREGATION METHODOLOGY MEETS ALL ADMISSIBILITY REQUIREMENTS UNDER FEDERAL RULE OF EVIDENCE 702 13

A. Dr. Freer's Proposed Testimony is the Product of Reliable Principles and Methods 15

B. Dr. Freer's Scientific, Technical, or Other Specialized Knowledge Will Help the Trier of Fact to Determine a Fact in Issue 17

1. Dr. Freer's statistical and computer science background make him more than qualified to opine on a probabilistic approach to linking User Account belonging to the same Participants. 17
2. The output of Dr. Freer's aggregation methodology identifies each Net Winner. 18

C. Dr. Freer's Aggregation Methodology is Based on Sufficient Facts and Data 19

1. Dr. Freer specifically addresses the sufficiency of data quality in the TelexFree database. 20
2. Dr. Freer specifically addressed Mr. Dennis's unfounded critiques of his data quality assessment. 23
3. Dr. Freer did not rely on an excluded Net Equity summation, and, even if the calculation changed it would have no impact on the reliability of the Freer Aggregation Methodology. 24

D. Dr. Freer's Opinions Reflect a Reliable Application of The Principles and Methods to The Facts of the Case 25

1. Dr. Freer's estimated error rates are considered better than acceptable results in the data science community. 25
2. Dr. Freer's methodology does not rely primarily on name. 26

III. MR. DENNIS'S OPINIONS ARE NOT ADEQUATELY SUPPORTED 28

A. Class Representative's Motions Lack a Scientific Foundation Because Mr. Dennis Lacks the Requisite Education, Training and Experience, and His Opinions are not the Product of Reliable Scientific Principles and Methods. 29

1. Dennis Lacks the Requisite Education, Training and Experience to Opine on Dr. Freer's Report, or Probabilistic Data Linkage Methodologies Generally. 30
2. Mr. Dennis has Minimal Prior Experience with Probabilistic Modeling. 31
3. Mr. Dennis Has No Experience Working with the Fellegi-Sunter Model that Lies at the Heart of the Freer Aggregation Methodology..... 33
4. Mr. Dennis's Critique of Dr. Freer's Aggregation Output Lacks any Statistical Support and Thus Fails the "More Probable than Not" Standard under FRE 702 34

IV. CONCLUSION..... 39

EXHIBIT LIST

Exhibit No.	Exhibit Description
1	Developing Standard Tools for Data Linkage
2	April 29, 2022 Expert Report of Dr. Cameron E. Freer
3	Advisory Committee Note to 2000 Amendment to FRE 702
4	Advisory Committee Note to 2023 Amendments to FRE 702
5	July 31, 2020 Rebuttal Expert Report of Joshua W. Dennis
6	November 24, 2020 Transcript of Evidentiary Hearing
7	Using the EM Algorithim for Weight Computation in the Fellegi-Sunter Model of Record Linkage
8	An Application of the Fellegi-Sunter Model of Record Linkage to the 1990 U.S. Decennial Census
9	Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records
10	May 22, 2023 Reply Report of Dr. Cameron E. Freer
11	The Intuition Behind the Use of Expectation Maximisation to Train Record Linkage Models
12	May 9, 2023 Deposition of Joshua W. Dennis
13	Fuzzy Matching and Deduplicating Hundreds of Millions of Records with Splink
14	Modern Scientific Evidence, Vol. I
15	The Role of Expert Judgment in Statistical Inference and Evidence-Based Decision Making
16	April 14, 2023 Supplemental Rebuttal Expert Report of Joshua W. Dennis

OPPOSITION OF STEPHEN B. DARR, AS TRUSTEE OF THE ESTATES OF TELEXFREE. LLC, TELEXFREE, INC. AND TELEX FINANCIAL, INC., TO CLASS REPRESENTATIVES' MOTION TO EXCLUDE TESTIMONY OF DR. CAMERON FREER AS INADMISSIBLE UNDER *DAUBERT*

The *Domestic & International Class Representatives' Motion To Exclude Testimony Of Dr. Cameron e. Freer as inadmissible Under Daubert* [A.P. No. 16-04006, Doc. 442] (the "*Daubert* Motion") should fail for three reasons. First, the Trustee's expert, Dr. Cameron E. Freer is a preeminent data scientist whose work in this case meets all of the requisites for admissibility under Federal Rule of Evidence 702, including under the 2023 amendments to the rule. Dr. Freer's use of the Fellegi-Sunter Model with Expectation Maximization (the "FS-EM Model") for large set data linkage is appropriate, and his education, training and experience permits him to apply the FS-EM Model to the TelexFree user database to arrive at a reliable determination of linkage between individual net winner accounts. Second, the Class Representatives' motion relies on the opinion testimony of Joshua W. Dennis. Mr. Dennis has little to no experience in the science of complex data linkage, and he relies on limited *ad ho*c examples of perceived inconsistencies to suggest set-wide data inconsistencies, rather than adhering to rigorous a statistical review. His opinion testimony should be afforded little, if any, weight by the Court. Third, any remaining account linkage questions should be resolved in a follow-on Claims Portal process similar to that used in determining net losers. That process will allow individual net winners to challenge Dr. Freer's net winner cluster results at the individual user account level, where appropriately supported by other evidence.

I. BACKGROUND

TelexFree was a massive Ponzi and pyramid scheme that ensnared more than a million individuals in multiple countries. While TelexFree was ostensibly in the business of selling voice

over internet protocol (VOIP) services, its real business was the sale of membership plans, where membership fees from new members ("Participants") were used to pay the "profits" of existing Participants.

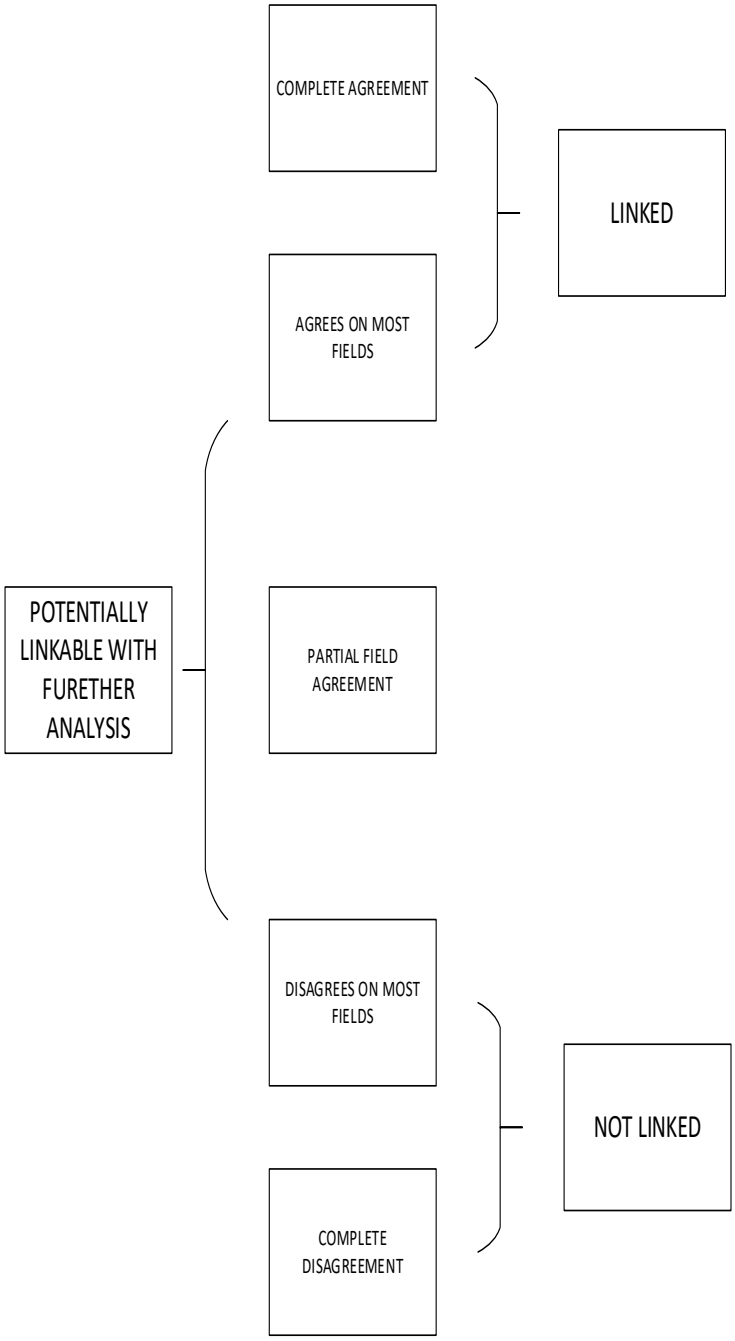
Ympactus (an affiliate) operated a related Ponzi/pyramid scheme in Brazil, using the same business model and maintaining its records on the TelexFree database. There were approximately 17 million User Accounts in the TelexFree database. The Trustee has previously determined that, of that 17 million individual accounts, approximately 11 million were attributable to TelexFree, 4 million were attributable to Ympactus, and 2 million involved User Accounts for which no membership fee was ever paid.

When the TelexFree scheme collapsed, some Participants received more money than they had paid into the scheme (Net Winners), whereas the majority of Participants had contributed more into the scheme than they were paid out (Net Losers). Many Participants maintained multiple User Accounts. However, the TelexFree database treated each of the approximately 17 million User Accounts as a standalone record that was not internally linked to other User Accounts, even when User Accounts were opened by the same Participant. Additionally, each User Account could contain up to 44 different fields of data, including name, phone number, email address, street address, etc. Some Participants at times entered information that was inconsistent across these fields in their various User Accounts. For example, two User Accounts opened by the same Participant may have entered a different street address, or variations of their name.[1]

[1] Many Telex participants had multiple names. By way of example, individuals of Portuguese extraction use both personal and family names, including up to two from the mother's side, two from the father's side and two from the husband. See https://en.wikipedia.org/wiki/Portuguese_name. The same user might enter some or all of their given names in a given account.

As set forth in more detail below, Dr. Freer relied on the FS-EM algorithm, a peer-reviewed data linkage methodology for linking data without the need for a separate control set of verified data. The Freer data linkage output resulted in one "cluster" of User Accounts for each Participant, so each User Account appears in only one cluster.

Data linkage is the science of joining disparate records based on matched identifiers. It can involve merging and linking data from two or more separate databases, or as in this case, linking separate user accounts contained in a single database. In each instance, the goal is to link data based on matching variables.



By way of example, consider the following abbreviated table drawn from Durham County, North Carolina voter records containing seven fields of data:

	NAME	ADDRESS	AGE	SEX	RACE	BIRT.	PARTY
1	Domineck Q. AAshard Jr.	914 Monmouth Ave #3	26	M	B	_	LIB
2	Domineck Q. AAshard Sr.	1408 Auburndale Dr.	55	M	B	NY	DEM
3	Xiomira A. Martinez	1715 Cole Mill Rd.	31	F	O	HL	REP
4	Xiomira A. Martinez	2923 Forrestal Dr.	31	F	O	HL	_
5	Jacqueline D. Fuller	141 Bagby Ln.	54	_	_	_	DEM
6	Jacqueline D. Fuller	905 Cook Rd.	56	F	B	NC	DEM

Accounts 1 and 2 share similar names, but the ages, addresses and party affiliations suggest a father/son relationship. Accounts 3 and 4 are likely the same individual, even though the party affiliation for one account is missing (null). Accounts 5 and 6 could be the same person with an age error in one account, but it is not clear, and requires further analysis.

Data linkage can be performed as a *deterministic* exercise or as a *probabilistic* exercise. The deterministic approach involves the exact matching of data field values to determine if a pair of records *are or are not* a match, while the probabilistic approach involves calculating a weight for each field that indicates for any pair of records how *likely* it is that they refer to the same person if there is a match in that field, or aggregation of fields. Deterministic matching is generally used where there is a high-quality identifier (e.g., a unique ID number for each user) available, or where multiple fields can be matched and concatenated to reach a high confidence match. Matching determinations are typically done after the data has been cleaned. Cleaning involves conforming individual field data to a more uniform state with losing important

information, e.g., removing multiple spaces within a name, or converting fields to all upper or lower case.

Once cleaned, matching is done on a one-to-one basis in deterministic matching. The result is one of two outcomes: the field data compared across two records is either a match or a non-match. Probabilistic linkage, on the other hand, assigns a match probability to each field compared; the higher the probability, the more likely the match. Additionally, weights may be assigned to each field considered, with higher weights assigned where the probability that a match across that field indicates a true overall record match (the *m*-variable) outweighs the probability of match across that field for an overall record non-match (the *u*-variable). For example, referring to the table above, a likely match across two records in party (affiliation) should be giving less weight that a likely match in street address when determining if two records should be linked.

The FS-EM Model used by Dr. Freer has been developed specifically to address large scale record linkage where there is no "gold standard" database[2] separately available to test *m*- and *u*-variable output. FS-EM is a peer-reviewed, unsupervised, iterative machine learning algorithm used to create data linkage even where data fields are incomplete or incorrect. There are multiple advantages of using the EM algorithm to calculate *m* and *u* probabilities for probabilistic matching, including no requirement of an external "gold standard" to teach the algorithm before it can be used. A useful description of the FS-EM Model can be found at:

[2] A gold standard database is an existing previously verified database against which probabilities can be calculated that two records in the database at issue will agree on a matching variable in truly matched pairs (the m-probability), compared to the probability that the same matching variable will agree in non-matched pairs (the u-probability). FESM was specifically designed to allow for estimating m- and u-probabilities in the absence of a gold standard database.

https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/developingstandardtoolsfordatalinkagefebruary2021, also attached as Exhibit 1.

A. The Freer Aggregation Methodology

Data linkage is typically done in stages. First, the data is prepared for linking. This includes evaluating data quality to select which fields to use for comparison, along with data cleaning, parsing, and standardizing. Next, linkage is the bringing together of pairs of records for comparison and identifying matches, i.e., belonging to the same person. Finally, the results are checked to determine their accuracy. This includes evaluating the error rate of record linkage created by the model.

The Ten-Step Aggregation Methodology developed by Dr. Freer (the "Freer Aggregation Methodology") for clustering accounts in the TelexFree database includes these stages, including a data quality review to confirm that the Telex data was suitable for linkage. A brief overview of each step is provided below. A detailed description of each step is included paragraphs 121 through 208 of his April 29, 2022, opening Report (the "Freer Report"), attached as Exhibit 2.

1. Step 1: Selecting for Data Quality and Transforming the Fields

The TelexFree database includes 44 fields in each User Account Record. Dr. Freer assessed each field across all records to identify which data fields were of sufficiently high quality to use for reliably assessing matches. The dimensions he assessed include completeness (how often the field values are present vs. missing), consistency (how consistent the values used for matching are), accessibility (inclusion of different kinds of non-overlapping fields such as name, address, phone, email, etc.), and believability (how credible or plausible the actual values are).

Dr. Freer found 11 TelexFree user account fields to be of sufficiently high quality to warrant inclusion in his Aggregation Methodology: *rep_nome* (name), *rep_email* (email), *rep_cel* (cell phone), *rep_fone* (phone), *rep_end* (street name), *rep_numero* (house number), *rep_cep* (ZIP code analogue), *rep_pwd_secondary* (hashed password), *rep_cpf* (SSN analogue), *rep_datanasc* (date of birth), and a truncated version of *rep_login* (login name). In addition, the field *rep_id*, a unique identifier assigned in increasing order to each record by Telex, was used to define each unique User Account record. Dr. Freer's extensive data quality testing is described generally at paragraphs 22 – 27 of the Freer Report, Ex. 2, and in more detail in paragraphs 85 – 110.

2. Step 2: Cleaning the Fields

Next, Dr. Freer applied standard cleaning techniques to the values of each field to ensure that the data in each selected field across all records is more consistent and uniform. The data cleaning techniques used for each field varied, but were generally designed to apply industry-accepted transformations, such as transforming all characters to lowercase and eliminating spaces and "garbage" (meaningless) characters, while retaining the informational content of the field to the maximum extent possible. See Ex. 2, Freer Report, ¶¶ 126-135.

3. Step 3: Deterministic Step

To make it computationally reasonable to perform probabilistic linkage methods on a dataset as large as the TelexFree dataset, Dr. Freer next performed a deterministic matching step; User Accounts that matched exactly across eight cleaned fields were placed into the same account groupings[3]. 7,563,961 User Accounts had at least one "exact" match to another User Account across all eight cleaned fields. This step reduced the number of records for subsequent

[3] The data in the remaining 3 fields were combined into sets of values in the account grouping to preserve the data.

probabilistic analysis by more than half: 10,987,617 User Account records were aggregated down to 4,367,728 account groupings[4] that were treated as equivalent to a separate record in the blocking and probabilistic linkage steps described below (account groupings were expanded back into their constituent User Accounts in Step 9). Importantly, this step also reduces the number of record pairs that were required to be analyzed in subsequent steps by more than 4-fold. See Ex. 2, Freer Report, ¶¶ 136-143.

4. Step 4: Blocking

Because probabilistic matching algorithms consider the potential matches across each pair of records being analyzed, the approximately 4 million account groupings (from the step above) would result in approximately 10 *trillion* pairs of possible account groupings for consideration. Because the computation run time for comparting 10 trillion separate pair combinations is prohibitive, Dr. Freer used a "blocking" scheme to reduce the number of pairs for consideration. Blocking relies on a set of rules that each pair must pass in order to be considered in the probabilistic matching phase, in this case requiring a high level of agreement between the two pairs across the name field and at least one cleaned other field other than the cleaned version of *rep_numero* (house number).[5] The final result of this step was a blocking set consisting of 117,563,883 pairs of account groupings for later probabilistic analysis[6]. See Ex. 2, Freer Report, ¶¶ 144-160.

[4] Some account groupings were a "grouping" of just one record while the largest account grouping contained 3,536 records.

[5] The full rule is set forth in ¶¶ 150-153 of the Freer Report, Ex. 2.

[6] For computational reasons, 62,725 User Account pairs that included a name value of "John Williams," email with a "mailinator.com" domain, and street name of the form "534 Warry Lane" with differing numbers in the place of "534" were joined and assigned to the same cluster at Step 9 after the final clustering of the remaining accounts was computed.

5. Step 5: Rubric Application to Calculate Gammas

The next step was to determine the similarity of values in each field (the gamma value) for each pair of accounts groupings in the blocking set. For some fields, the corresponding gamma value is binary (0 or 1), indicating complete agreement or total disagreement. For other fields, the corresponding gamma value indicates the degree of agreement, discretized (grouped) to a small number of values representing, for example, "different," similar," or "nearly identical." In most cases, the gammas in the Freer Aggregation Methodology were determined by discretizing the result of an appropriate fuzzy matching function (e.g., a fuzzy match of less than 83% is "different," while a greater than 92% is "nearly identical"). The output of this step was the set of comparison vectors, one vector for each pair in the blocking set. See Freer Report, Ex. 2, Freer Report, ¶¶ 161-176.

6. Step 6: Calculate Match Probabilities

Next, gamma vectors for each account grouping pair were used to determine the match probability for each pair. Dr. Freer used Splink, a probabilistic record linkage software package, using the gamma vectors derived in Step 5 above to assign a match probability to each of the 117,563,883 pairs of account groupings created in the blocking process. See Ex. 2, Freer Report, ¶¶ 177-179.

7. Step 7: Compute Clustering for Several Fixed Thresholds

Next, 149 different clustering scenarios (aggregations) of account groupings were computed for various fixed thresholds of the match probability in the range of 0.006 to 100. For example, with a threshold of 0.95, only the pairs with 95 percent or higher match became part of a cluster. Because every account grouping must be in exactly one cluster, the cluster is "transitively closed" such that if Cluster A matched Cluster B at 95% or higher, and Cluster B

matched Cluster C at 95% or higher, Cluster A and Cluster C would be deemed in the same cluster regardless of individual match probability. See Ex. 2, Freer Report, ¶¶ 180-183.

8. Step 8: Dynamic Clustering

Clustering necessarily balances two competing tensions: (1) records that do not belong in the same cluster are not joined (i.e., one should avoid "overclustering") and (2) records that do belong in the same cluster are joined in the same cluster (i.e., one should avoid "underclustering"). For the Telex dataset, Dr. Freer clustered thresholds dynamically across the dataset, rather than with a single fixed threshold (as described in step 7). This is because some participants entered their information very consistently across many accounts, while other participants, for a variety of reasons, entered information that differed more widely (e.g., name variants, typos, multiple email addresses or phone numbers, etc.). As a result, Dr. Freer devised a method whose implementation in code chose higher thresholds in that portion of the dataset where there was a "tightly bound" collection of account groupings having high match probability with each other and low match probability with every other account grouping, while choosing lower thresholds in that portion of the dataset where there was a more "loosely bound" collection of account groupings having moderately high match probability with each other but still much lower match probability with every other account grouping in the dataset.

Dynamic clustering was accomplished by: (1) defining measures of overclustering and underclustering for a given cluster, and calculating the overclustering measure for every cluster in each of the 149 fixed-threshold clusterings computed in Step 7; (2) using this data to compute a dynamic clustering for each of the 41 values of a "dynamic clustering parameter", where the overclustering calculations and dynamic clustering parameter guide the selection of a fixed threshold for each account grouping; and (3) choosing the dynamic clustering from among these 41 dynamic clusterings that optimally balances the degree of overclustering and

underclustering[7]. The chosen dynamic clustering was the output of step 8. The cluster assignments were tracked by assigning each account grouping in a given cluster an Account Grouping Cluster ID that corresponded to the lowest *rep_id* in that cluster, i.e., the first account created by any of the Participants assigned to the Cluster[8]. See Ex. 2, Freer Report, ¶¶ 184-204.[9]

9. Step 9: Convert the Clustering of Account Groupings into a Clustering of User Accounts

The last step in determining the final clustering was to "ungroup" the account groupings back into User Accounts. This was accomplished by joining the clustering output from the Step 8 with a table containing the mapping of each User Account to its Account Grouping (generated in Step 3). As a result, each User Account was assigned a Cluster_ID that corresponded to lowest *rep_id* in the cluster. The result of this step was 1,566,383 clusters corresponding to 1,566,383 different Participants.[10] See Ex. 2, Freer Report, ¶¶ 205-207.

10. Step 10: Tabulate Total Net Equity for Each Cluster

Each User Account was assigned its Net Equity by Huron in accordance with the court-approved net equity formula[11], and the total Net Equity of each cluster was defined to be the sum of the Net Equity values of each of its User Accounts. Dr. Freer tabulated this quantity for each cluster. See Ex. 2, Freer Report, ¶ 208.

---

[7] This process is described in great detail in ¶¶ 189-204 in the Freer Report, Ex. 2.

[8] See *Affidavit of Jean Louis Sorondo*, ("Sorondo Aff."), ¶ 3.d.

[9] The full rational for using low *rep_id* as the cluster owner is described in detail in the Trustee's Opposition to the Class Representatives' Motion for Summary Judgment and supporting affidavits.

[10] The "John Williams" cluster was added back in at this step.

[11] *Supplemental Order Respecting Motion by Chapter 11 Trustee for Entry of Order Finding That Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief* [A.P. No. 14-40987, Doc. 687], ¶ 5.ii.

11. Results

The Freer Aggregation Methodology determined that the 10,987,617 User Accounts attributable to TelexFree belonged to 1,566,383 different Participants, with an average size of 7.01 User Accounts per Participant. Of the 1,566,383 Participants, 1,484,702 are Net Losers and 81,681 are Net Winners. The Net Winners have a total net gain of $1,555,981,482.42, with an average net gain of $19,049.49 per Net Winner. The Net Losers have a total net loss of $1,782,171,730.09, with an average net loss of $1,200.36 per Net Loser. See Ex. 2, Freer Report, ¶ 209.

Finally, Dr. Freer checked his cluster output by calculating the error rates for the Aggregation Methodology. Taking into consideration two types of errors commonly considered in the field, "false discovery rate" and "false negative rate," he calculated the standard statistical measure known as accuracy (called F1 or the f-measure) to be 0.988 (98.8%). This was computed in accordance with a standard peer-reviewed statistical method based on information about pairwise match probabilities contained in his probabilistic record linkage model. See Ex. 2, Freer Report, ¶¶ 212-213.

II. DR. FREER'S AGGREGATION METHODOLOGY MEETS ALL ADMISSIBILITY REQUIREMENTS UNDER FEDERAL RULE OF EVIDENCE 702

Directly relevant to the Court's consideration of the Class Representatives' motions are the 2023 amendments to Federal Rule of Evidence ("FRE") 702.[12]

[12] While the FRE 702 Amendments do not formally go into effect until December 1, 2023, the Advisory Notes accompanying the Amendments confirm that the Amendment is not intended to change Rule 702, but rather is a clarification brought about by courts shirking their gatekeeper role. See https://www.uscourts.gov/sites/default/files/evidence_rules_report_-_may_2022_0.pdf, attached as Exhibit 3. Thus, commentators and courts are now citing to the Amendment prior to its official adoption date.

> A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if **the proponent demonstrates to the court that it is more likely than not that:**
>
> (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
>
> (b) the testimony is based on sufficient facts or data;
>
> (c) the testimony is the product of reliable principles and methods; and
>
> (d) the ~~expert has reliably applied~~ **expert's opinion reflects a reliable application of** the principles and methods to the facts of the case.

The 2023 Amendments, effective December 1, 2023, are intended to address two issues. First, the Amendments clarify the existing FRE 702 rule: first, a proponent of expert testimony bears the burden of demonstrating to the Court on a "*more likely than not*" standard that the proffered testimony meets the admissibility requirements set forth in the rule[13]; and second, each proffered opinion must stay within the bounds of what can be concluded from a reliable application of an accepted methodology. Dr. Freer meets each of FRE 702's elements of admissibility because:

(a) The peer-reviewed FS-EM probabilistic data linkage approach employed by Dr. Freer is a reliable method for linking large set data;

[13] The 2023 Amendment Notes also make clear that while FRE 104(a) require the Court to decide "any preliminary question about whether the witness is qualified", FRE 702 ultimately controls. ("The amendment clarifies that the preponderance standard applies to the three reliability-based requirements added in 2000—requirements that many courts have incorrectly determined to be governed by the more permissive Rule 104(b) standard. But it remains the case that other admissibility requirements in the rule (such as that the expert must be qualified and the expert's testimony must help the trier of fact) are governed by the Rule 104(a) standard as well."). https://www.uscourts.gov/sites/default/files/2023_congressional_package_april_24_2023_0.pdf at p. 211, attached as Exhibit 4.

(b) Dr. Freer has requisite education, training and experience to apply the FS-EM data linkage methodology to the TelexFree database;

(c) Dr. Freer's Aggregation Methodology determines a fact in issue in the case, and is therefore helpful to the finder of fact;

(d) Dr. Freer's Aggregation Methodology is based on sufficient facts and data because he selected the appropriate fields in the TelexFree database by testing data quality, and then selecting those fields that had the highest linkage power based on well-accepted and statistically significant methods of evaluating data quality; and

(e) As evidenced by statistical error rate analysis Dr. Freer has achieved a high-quality aggregation of individual accounts into clusters that permit the Trustee to identify net winners.

A. Dr. Freer's Proposed Testimony is the Product of Reliable Principles and Methods

There is no dispute that the primarily probabilistic data linkage approach employed by Dr. Freer[14] is the most appropriate method for aggregating the User Accounts in the TelexFree Dataset to Participants. Indeed, Class Representatives and their purported expert, Mr. Dennis, have been calling for a probabilistic approach from the beginning.[15]

[14] While it is mainly probabilistic in its approach, the Freer Aggregation Methodology employs many steps to prepare the data for the probabilistic analysis, including a deterministic step that directly precedes the probabilistic step, and which matched over 7 million accounts across eight fields. Ex. 2, Freer Report, ¶ 111.

[15] "The source literature cited in the Martin Report suggests probabilistic record linkage may be a more appropriate methodology when direct identifiers do not exist . . ." *Domestic & International Class Representatives' Motion To Exclude Testimony Of Timothy Martin As Inadmissible Under Daubert* (the "2021 Daubert Motion") [A.P. No. 16-04006, Doc. 377], pg. 6; Ex. 5, July 31, 2020 Rebuttal Expert Report of Joshua W. Dennis, ¶ 52; see also November 24, 2020 Transcript of Evidentiary Hearing ("Daubert Hearing Tr."), Exhibit 6, 367:22-25 (Dennis Cross).

> So, so a probabilistic record linkage uses math and statistics to compare instances where there aren't high quality unique identifiers and does that by creating weights that suggest whether a single variable is, is, is indicative or not indicative in the population of a match. *So what it allows you to do is precisely what we have here where there is, are small, there can be small details in the data like misspellings and other things like that. It can account for those kind of discrepancies and also account for issues when there, some combination of things isn't inherently distinct. . .*

Ex. 6, Daubert Hearing Tr. 260:18-261:3 (Dennis Direct).

Additionally, the FS-EM (Fellegi-Sunter with Expectation Maximization) Model of probabilistic record linkage utilized by Dr. Freer is a peer-reviewed method for performing probabilistic record linkage in large datasets. Ex. 2, Freer Report, ¶ 50. The *Fellegi-Sunter* model describes a way to track agreement on multiple fields (e.g., name, email, phone number, address) simultaneously, to turn that information into an overall match probability for each pair of records under consideration, and *Expectation Maximization* is a ubiquitous, unsupervised learning algorithm for iteratively estimating parameters (e.g., *m*-variable, *u*-variable) for the statistical model. Ex. 2, Freer Report, ¶¶ 49-50. In 1988, William Winkler of the US Census Bureau described how to combine Expectation Maximization with the Fellegi-Sunter model to more accurately fix the scoring rubric tracking agreement on multiple fields directly from the dataset.[16] Ex. 2, Freer Report, ¶ 50. This combined approach (FS-EM) has been used to perform record linkage on the 1990 US Census[17], and continues to be a standard approach used by the U. S. Census Bureau today. Ex. 2., Freer Report, ¶ 50.

[16] Winkler, W. E. Using the EM Algorithm for Weight Computation in the Fellegi-Sunter Model of Record Linkage, American Statistical Association, *Proceedings of the Section on Survey Research Methods,* 667-671 (1988), attached as Exhibit 7.

[17] William E. Winkler and Yves Thibaudeau. *An application of the Fellegi-Sunter model of record linkage to the 1990 US decennial census.* US Bureau of the Census, 1991, attached as Exhibit 8.

Dr. Freer also used the Splink Python software libraries, which incorporate the FS-EM Model. This Python library set is maintained and published by the UK Ministry of Justice, and the science behind Splink has been extensively tested and peer reviewed. Ex. 2, Freer Report, ¶ 116. Splink's statistical model has been reviewed and validated in a paper authored by researchers from Harvard and Princeton and published in a peer-reviewed journal[18]; that paper has been cited over 100 times in just the past 3 years. Ex. 2, Freer Report, ¶ 116. Additionally, the technique behind the Splink output has been extensively tested against "gold standard" data, and has been found to be quite accurate and with well-calibrated error rates.[19] May 22, 2023 Reply Report of Dr. Cameron E. Freer ("Freer Reply"), Exhibit 10, ¶ 87. As a result, Splink is a reliable probabilistic software record linkage package for use with the TelexFree dataset. Ex. 10, Freer Reply, ¶ 87.

B. Dr. Freer's Scientific, Technical, or Other Specialized Knowledge Will Help the Trier of Fact to Determine a Fact in Issue

1. Dr. Freer's statistical and computer science background make him more than qualified to opine on a probabilistic approach to linking User Account belonging to the same Participants.

With a doctorate in Mathematics, and an extensive research background in statistical and computer science, Dr. Freer undisputedly[20] possesses the scientific and technical knowledge necessary to create and opine on a probabilistic record linkage model for large and unlabeled[21] datasets. Dr. Freer holds a Ph.D. in Mathematics from Harvard, is a Research Scientist at MIT's

[18] Enamorado, T., Fifield, B., and Imai, K. (2019). Using a probabilistic model to assist merging of large-scale administrative records. *American Political Science Review*, 113(2), 353-371, attached as Exhibit 9.

[19] Ex. 9, Using a probabilistic model to assist merging of large-scale administrative records.

[20] Defendants do not challenge Dr. Freer's qualification in their Daubert Motion.

[21] An unlabeled dataset is one for which no tested and verified "ground truth" data exists.

Probabilistic Computing Project, and has authored 28 research papers in mathematics, statistics, and computer science, including probabilistic inference, statistical testing, and clustering. Ex. 2, Freer Report, ¶¶ 9-10. He is the Founder and Chief Scientist at Borelian Corporation, and was previously Chief Scientist at Remine, Inc., performing statistical modeling on large consumer and property record datasets. Ex. 2, Freer Report, ¶¶ 9, 11. Thus, Dr. Freer possesses a reliable basis in the knowledge and experience in probabilistic record linkage.

2. The output of Dr. Freer's aggregation methodology identifies each Net Winner.

Dr. Freer's aggregation methodology determines a fact in issue, namely identifying the TelexFree Participants who are Net Winners based on the account clusters generated by the Freer Methodology. His approach reveals that the 10,987,617 User Accounts analyzed belong to 1,566,383 Participants, of which 81,681 are Net Winners. Ex. 2, Freer Report, ¶ 209.

Dr. Freer's approach determined the Participants by clustering (i.e. grouping) User Accounts based on patterns in the information in the TelexFree database. Each resulting cluster corresponds to a different Participant, and each cluster with a Net Equity greater than zero corresponds to a different Net Winner. Thus, each Net Winner is identified by the information (e.g., name, email, phone number, address) contained in the User Accounts in a given cluster. Ex. 10, Freer Reply, ¶ 118.

Nonetheless, the Class Representatives now claim that, because determining whom to sue is not within Dr. Freer's scope, his method somehow does not actually identify Participants. See *Daubert* Motion, pgs. 43-44. However, this simply is not true. Dr. Freer provided all of the necessary information to identify Participants in order for *the Trustee* to determine whom to sue. With his report, Dr. Freer produced an output file. The file includes 10,987,617 rows of data, each row corresponding to a User Account record, along with a Cluster ID (*Cluster_ID*) in each

row. Ex. 2, Freer Report, output.csv[22]. This Cluster ID was assigned as a result of the Aggregation Methodology and it identifies the cluster to which the User Account in each row was ultimately determined to belong. Ex. 2, Freer Report, output.csv. Importantly, the assigned Cluster ID also identifies the first account opened in each cluster.[23] This is important because the Trustee has logically concluded that the most complete personal identifying information for each Participant would in the first account opened in a given cluster, and is thus the best information to identify the Participant to which the cluster belongs.[24]

For these reasons, the output of Dr. Freer's Aggregation Methodology provides a reliable clustering of Participant accounts from which Net Winner can be determined.

C. Dr. Freer's Aggregation Methodology is Based on Sufficient Facts and Data

Dr. Freer's Aggregation Methodology, which analyzes and aggregates data from the TelexFree database, is based on sufficient facts and data because he carefully selected only the highest quality fields for the TelexFree database for consideration. Moreover, Dr. Freer effectively addressed Mr. Dennis' inaccurate critique of the underlying data quality for these fields in his reply report, a reply that Class Representatives have completely ignored in their motion papers.

[22] This exhibit to the Freer Report was provided natively. It has not been provided with this file due to size constraints but can be provided to the Court upon request.

[23] The Cluster ID for each cluster correspond to the lowest rep_id in that cluster, where rep_id is the TelexFree internal unique account tracking numbers created in increasing order by the TelexFree system each time an account was opened. As such, the lowest rep_id in a cluster corresponds to the first account opened in that cluster. Sorondo Aff., ¶ 3.d.

[24] The Trustee's Summary Judgment Opposition details the support for assigning the low rep_id account as the owner of a given cluster. See also Affidavit of Stephen B. Darr, ¶ 6.

1. Dr. Freer specifically addresses the sufficiency of data quality in the TelexFree database.

Based on a comprehensive analysis of all 44 data fields included in the User Account records, Dr. Freer selected 11 high-quality fields for use in his aggregation methodology. Ex. 2, Freer Report, ¶¶ 85-110. Specifically, he assessed completeness (how often the field values are present vs. missing), consistency (how consistent the values used for matching are), accessibility (inclusion of different kinds of non-overlapping fields such as name, address, phone, email, etc.), and believability (how credible or plausible the actual values are). Ex. 2, Freer Report, ¶¶ 85-110, ¶¶ 22-27.

By way of example, Dr. Freer validated that the name field (rep_nome) is of high quality *at least* 95% of the time via the following steps. First, he determined that the rep_nome field is only empty 0.002% of the time in the TelexFree database. Ex. 2, Freer Report, ¶¶ 93-94. Second, he determined that 97% of the field values met his criteria for a clean name: i.e., 5 or more characters and which included only alphabetical characters, periods, commas, hyphens, spaces, or apostrophes. Third, he randomly selected 300 populated accounts for manual review; and found that only 7 name values (from the 300 sampled accounts) appeared likely to be something other than actual names, while the remaining 293, approximately 98%, were plausible names (albeit with occasional typos, abbreviations or unusual capitalization). Ex. 2, Freer Report, ¶¶ 93-94; Ex. 10, Freer Reply, n. 135. From this, Dr. Freer estimated that at least 95% of the rep-nome values appear to be names (subtracting 3% for the values that did *not* meet his clean name criteria, and another 2% for the values that did meet his criteria, but did not appear to contain a plausible name). Ex. 2, Freer Report, ¶¶ 93-94.

Moreover, Dr. Freer explained why Mr. Dennis's criticism of his data quality sample size is unsupported. Ex. 10, Freer Reply, ¶ 110. As he explained, and as any trained statistician would

know, to obtain a 5% Margin of Error with 90% Confidence Level on a population size of 11,000,000, the Ideal Sample Size is just 271, less than the 300 sample set Dr. Freer used. Ex. 10, Freer Reply, ¶ 110. Nonetheless, to further demonstrate that 300 was sufficient for a statistically significant conclusion, Dr. Freer used the same methodology to review a sample size of 1067, which yields 3% Margin of Error with 95% Confidence Level on a population size of 11,000,000, and arrived at the same result: 98% of the values manually reviewed were plausible names. Ex. 10, Freer Reply, ¶ 111.

Dr. Freer also separately validated the email address field (*rep_email*), phone number fields (*rep_cel* and *rep_fone*), address fields (*rep_end*, *rep_numero*, and *rep_cep*), and login field (*rep_login*) as high quality, thus suitable for inclusion as linkage fields in his Aggregation Methodology. As with the name field (*rep_nom*), those fields were almost always complete, contain the expected components, and were sufficiently diverse. Ex. 10, Freer Reply, ¶¶ 95-98. Furthermore, these fields were found to be the same for multiple pairs of User Accounts with similar or identical name values, supporting a high linkage value. Ex. 10, Freer Reply, ¶¶95-98.

Dr. Freer also included three additional fields: secondary password (*rep_pwd_secondary*), SSN (*rep_cpf*), and DOB (*rep_datanasc*) that were found to of intermediate quality. While password and DOB values were often null, each had a high linkage value when present because it is not common for different parties to select the same password, and there are many instances where multiple User Accounts include identical or similar names and the same date of birth, respectively. Ex. 10, Freer Reply, ¶¶ 99-100, 103. The SSN equivalent field, on the other hand, was only null or empty for just 0.001% of User Accounts and accounts that appeared to belong to the same Participants (based on manual review) often have

the same value entered for this field giving it sufficient linking power. Ex. 10, Freer Reply, ¶¶ 101-102.

Dr. Freer further tested the data quality for these ten fields by manually reviewing a set of 400 randomly chosen accounts, and then manually reviewing each field separately in the manner described with respect to the name field. See Ex. 10, Freer Reply, ¶¶ 68-82, Report Exhibits 2.1-2.13. His review confirmed that the values in the various fields were in fact of high quality and that they generally cohered with one another. Ex. 10, Freer Reply, ¶¶ 68. For example, portions of the name in the name field often appeared as part of the email address, the two phone numbers usually had a similar format to each other (e.g., are of similar length and shared initial digits), and the three address components typically appeared believable in combination. Ex. 10, Freer Reply, ¶ 68.

Dr. Freer also explained that the quality of the dataset is of sufficient high quality as is evidenced by the exact matches obtained in his preliminary deterministic matching step (Step 3). Ex. 2, Freer Report, ¶ 111. Specifically, in Step 3, User Accounts were placed into the same account grouping if there was an exact match across eight cleaned fields. As a result, over 7.5 million User Accounts out just of under 11 million had at least one "exact" match to another User Account, supporting Dr. Freer's finding of sufficient consistency in data entry by the TelexFree Participants to permit him to proceed with probabilistic matching.

Finally, Dr. Freer separately confirmed that unselected fields were assessed and determined to have low quality, and thus were not suitable for inclusion in his linkage model. See Ex. 2, Freer Report, ¶¶ 104-110.

2. Dr. Freer specifically addressed Mr. Dennis's unfounded critiques of his data quality assessment.

Dr. Freer specifically addressed Dennis' misrepresentation of data science literature, and in particular, Mr. Dennis's suggestion that valid data linkage requires a "golden set." Dr. Freer's use of the FS-EM Model is an unsupervised learning technique that was expressly developed not to require a labeled or "golden set" dataset to use the algorithm. [25]. Ex. 10, Freer Reply, ¶ 86. Indeed that very point was previously argued by the Class Representatives in the last Daubert hearing: "[t]he lack of an answer key is the very reason why Mr. Martin should have performed a probabilistic linkage analysis." 2021 Daubert Motion, pg. 19. Moreover, unsupervised learning techniques are crucial in record linkage because "[m]anual labelling of comparisons is error prone and costly so isn't a great option." Ex. 10, Freer Reply, ¶ 85, quoting R. Linacre, *The Intuition Behind the Use of Expectation Maximization to Train Record Linkage Models,* October 14, 2022 (attached as Exhibit 11).

Dennis offered a quote from Peter Christen's treatise, *Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*, Springer (2012): "Arguably the most important data quality dimensions for data matching and deduplication are accuracy and consistency" in support of his argument. But the complete Christensen quote reads:

> Arguably, the most important data quality dimensions for data matching and deduplication are accuracy and consistency, **because a large portion of efforts in the indexing, comparisons and classification steps (that will be covered in Chaps. 4–6) deal with inaccurate and inconsistent data. If data would be of perfect quality, then data matching could be accomplished through**

[25] Some machine learning algorithms fall under the category of "supervised learning", and require a dataset that is labeled with "ground truth" for both training the model and for evaluating the accuracy of the trained model. Other machine learning algorithms perform "unsupervised learning", and do not rely on labeled data for either training or assessment. Unsupervised learning algorithms identify patterns and structure within the input data itself during training. Ex. 10, Freer Reply, ¶ 84.

> **straightforward database join operations and no sophisticated indexing techniques or approximate comparison functions would be needed. As long as data are of imperfect quality, however, techniques are needed that can deal with inaccurate and inconsistent data while still achieving high matching quality.**

Christen's full quote emphasizes that *because* data is often not perfectly accurate, researchers have developed methods for dealing with imperfect data. Ex. 10, Freer Reply, ¶ 96. Far from demanding perfect accuracy and consistency, Christen's full quote underscores the necessity of employing sophisticated techniques to successfully resolve the very types of data entry issues found in the TelexFree database. Ex. 10, Freer Reply, ¶ 96. In essence, the complexities of data matching arise precisely due to the inherent imperfections in data. Ex. 2, Freer Report, ¶ 96. The Freer Aggregation Methodology using the FS-EM algorithm aligns completely with Christen's recommendations.

3. Dr. Freer did not rely on an excluded Net Equity summation, and, even if the calculation changed it would have no impact on the reliability of the Freer Aggregation Methodology.

The Trustee's Summary Judgment Opposition details why (1) the Net Equity amount was determined through a simple mathematical calculation that is not expert testimony and (2) this calculation was not excluded by Judge Hoffman.

Moreover, the Freer Aggregation is agnostic to the choice of net equity formula. Ex. 10, Freer Reply, ¶¶ 133-137. Steps one through nine of the Freer Aggregation Methodology do not use Net Equity as an input, and these steps are the entire portion of the Freer Aggregation Methodology that determine clusters of User Accounts. Ex. 10, Freer Reply, ¶134. Net Equity is determined in step 10 after the clusters are formed. Total Net Equity is simply the sum of Net Equity for each User Account in that cluster. Ex. 10, Freer Reply, ¶134.

Even if the Court chooses to approve a different Net Equity formula than the one previously approved by Judge Hoffman (which it should not), that choice would have no impact on the reliability of the Freer Aggregation Methodology. The only change required would be to arithmetically total the Net Equity for each cluster using the new Net Equity formula. Ex. 10, Freer Reply, ¶¶ 136-137.

As set forth above, Dr. Freer clarified and corrected Mr. Dennis's erroneous claim that verification against real-world data is necessary for reliable account linkage, and there is nothing "speculative and unreliable" in Freer's extensive data quality analysis or his choice of fields. *Cf. Daubert*, 509 U.S. at 590 ("Proposed testimony must be supported by appropriate validation—i.e., 'good grounds,' based on what is known."); see also *Rothe Dev., Inc. v. DOD*, 107 F. Supp. 3d 183, 198-99 (D.D.C. 2015) (holding that the expert's decision to exclude certain data from his statistical model was valid since it was fully explained and scientifically sound, citing *Daubert*.).

D. Dr. Freer's Opinions Reflect a Reliable Application of The Principles and Methods to The Facts of the Case

1. Dr. Freer's estimated error rates are considered better than acceptable results in the data science community.

Based on Dr. Freer's statistical analysis detailed above, the Freer Aggregation Methodology has error rates that substantially exceed the standard considered "very good" in the data science community. Dr. Freer's statistical analysis took into account two types of errors: false positives and false negatives. Ex. 2, Freer Report, ¶ 212. False positives occur when User Accounts are grouped together even if they don't belong to the same Participant, while false negatives occur when User Accounts that do belong to the same Participant are placed in different clusters by the aggregation method. Ex. 2, Freer Report, ¶ 212. To evaluate these errors, the false discovery rate (FDR) and false negative rate (FNR) are calculated by statisticians to

determine the statistical measures of precision and recall. Ex. 2, Freer Report, ¶ 212. The f-measure referred to as accuracy, which is the harmonic mean of precision and recall, ranges from 0.0 to 1.0, with higher values indicating better results. According to a review of record linkage software quality, an accuracy of at least 0.90 is considered "very good." Ex. 2, Freer Report, ¶ 212.

Dr. Freer utilized equations similar to the peer-reviewed equations provided by the authors of the fastLink package -analogous to the Splink package utilized by Dr. Freer, but in a different computer language- to estimate FDR and FNR based on pairwise match probabilities. Ex. 2, Freer Report, ¶ 213. Based on the fastLink equations, Dr. Freer estimated FDR as 0.0198 and FNR as 0.00377, which implies a precision of 0.980, a recall of 0.996, and an accuracy of 0.988. Ex. 2, Freer Report, ¶ 213. These results validate Dr. Freer's finding that his Aggregation Methodology provided results substantially superior to the "very good" criterion.

2. Dr. Freer's methodology does not rely primarily on name.

Contrary to the claims made by Mr. Dennis, Dr. Freer's methodology does not primarily rely on names. All 11 fields selected by Dr. Freer play a crucial role in the overall aggregations process. Ex. 10, Freer Reply, ¶¶ 48-49. For example, the deterministic step that precedes blocking requires exact matches across eight fields. The FS-EM algorithm, which is at the core of Dr. Freer's probabilistic model, evaluated the contents of *all* fields to determine their relative weights in determining match probabilities. Ex. 10, Freer Reply, ¶ 48. The clustering of User Accounts was *not* based on the name field, as records were partitioned into clusters based on their match probabilities by transitive closure, which was not reliant on name alone [26] Ex. 10,

[26] The dynamic clustering performed in step 8 of the aggregation methodology is employs the recommended approach for overcoming issues of "chaining" in which the opposite is true and

Freer Reply, ¶ 114. While there is necessarily tension between over-clustering and under-clustering, the transitive closure approach mostly ensures that records belonging in the same cluster should end up together, even if their names were not considered a match during the blocking step. This mathematical result is well-known in entity resolution literature, and is based on the fact that each record appears in only one cluster of a partition. Ex. 10, Freer Reply, ¶ 114.

In the context of aggregating accounts to individuals, however, names do play an important role. Ex. 10, Freer Reply, ¶ 40. This importance of names is supported by the record linkage literature, which emphasizes personal identifiers as crucial for identifying individuals. Among these identifiers, names stand out:

> Names and other personal details play a crucial role in daily life because people are using them to identify individuals, ranging from family and friends, to work colleagues, and all the way to politicians and celebrities. For organizations both in the private and public sectors, names are often a primary source of identification of the individuals they are in contact with. **Personal names are a major component of the information used in many data matching or deduplication processes to identify records that refer to the same individuals.** [Christen, 2012, p. 42] (Emphasis added).

Because different individuals can share other personal identifiers besides names (such as addresses or phone numbers), the standard solution involves using names to distinguish between records that share identifiers like addresses but not names. Ex. 10, Freer Reply, ¶ 46. Dr. Freer accomplishes this in the blocking step (step 3) of his methodology. Ex. 10, Freer Reply, ¶ 47. In this step, only accounts with similar names are put through for matching consideration. However, accounts with dissimilar names may still end up grouped together for the reasons described above.

Records A and C should not but joined, the result being the B is only joined with only one of A or C, if any. Ex. 10, Freer Reply, ¶¶ 91-93.

Because Dr. Freer is eminently qualified in the field of probabilistic computing by virtue of his education, training and experience, because he chose a peer reviewed data linkage algorithm that was appropriate for linking accounts in the TelexFree database, because he extensively tested data quality in the TelexFree database, and because he correctly implemented the appropriate algorithmic process to arrive at results that fall within a statistically acceptable margin of error, the Freer Clusters should be deemed sufficiently reliable to be admitted into evidence. *United States v. Gissantaner*, 990 F. 3d, 457 (6th Cir. 2021) (district court erred in excluding results of a probabilistic genotyping software program under FRE 702 and *Daubert* because the software was testable, had been peer-reviewed, had gained general acceptance in the scientific community, and the principals and methods had been reliably applied by the expert in the case.); *Powell v. Tosh*, 942 F. Supp. 2d 678 (W.D. Ky. 2013) (AERMOD model was reliable where model was the preferred air quality and dispersion model, making expert's use in this case to satisfy the reliability requirement under *Daubert.*); *Daubert v. Merrill Dow Pharms., Inc*. 509 U.S.579 (1993) ("scientific" implies grounded in methods and procedures of science, and not subjective belief and unsupported speculation).

III. MR. DENNIS'S OPINIONS ARE NOT ADEQUATELY SUPPORTED

The Class Representatives have attacked the Freer Methodology, relying on the opinion testimony of Joshua W. Dennis. There is no requirement than a party file a *Daubert* challenge, especially where, as here, the Class Representatives' challenge to the Trustee's case is not being heard in anticipation of a later jury trial. Rather, the Court has an independent duty under FRE 702 to determine whether proffered opinion testimony, including Mr. Dennis's, meets the reliability standard under FRE 702. "First, the rule has been amended to clarify and emphasize that expert testimony may not be admitted unless the proponent demonstrates to the court that it

is *more likely so than* not that the proffered testimony meets the admissibility requirements set forth in the rule." Ex. 4, Advisory Committee Note to 2023 Amendments to FRE 702, ¶ 1 (Emphasis added).[27] Stated differently, "[t]he trial judge in *all* cases of proffered expert testimony *must find* that it is properly grounded, well-reasoned, and not speculative *before* it can be admitted." Ex. 3, Advisory Committee Note to 2000 Amendment to FRE 702 (Emphasis added); *Newell Rubbermaid, Inc v. Raymond Corp*., 676 F.3d 521 (6th Cir. 2012) (District court identified at least four red flags – anecdotal evidence, improper extrapolation, failure to consider other causes, and lack of testing, sufficient to warrant exclusion); *King v. Enter. Rent-A-Car Co*., 231 F.R.D 255 (E.D. Mich. 2004) (Expert failed to meet *Daubert* standard where he had no educational or professional background or experience in employment statistics).

A. Class Representative's Motions Lack a Scientific Foundation Because Mr. Dennis Lacks the Requisite Education, Training and Experience, and His Opinions are not the Product of Reliable Scientific Principles and Methods.

As set forth in more detail below, Mr. Joshua Dennis fails on each of FRE 702's elements of admissibility: (a) Mr. Dennis lacks the requisite education, training and experience to opine on an unsupervised data linkage methodology generally, and the Freer Aggregation Methodology specifically. (b) Mr. Dennis's testimony is not based on sufficient facts or data because Mr. Dennis performed only a limited number of non-random *ad hoc* table queries relative to the size of the TelexFree database; (c) Mr. Dennis's testimony is not the product of reliable principles and methods because Mr. Dennis never performed any statistical analysis to allow him to opine that the handful of account issues he identifies in his Report were: (1) more likely than not to be actual account over or under aggregations; or (2) were statistically indicative of aggregation

[27] Class Representatives' *Daubert* Motion (pg. 36, FN 32) correctly state that the pending "clarifying" amendments to FRE 702 can be relied upon in anticipation of the Amendment's official adoption date of December 1, 2023.

issues across Dr. Freer's entire aggregation output table; and (d) Mr. Dennis offers no peer-reviewed support for the notion that limited non-random *ad hoc* searching against a 10 million account database is an accepted method of determining the efficacy of the FS-EM data linkage output of the Freer Aggregation Methodology.

1. Dennis Lacks the Requisite Education, Training and Experience to Opine on Dr. Freer's Report, or Probabilistic Data Linkage Methodologies Generally.

Mr. Dennis is not a data scientist by education, nor does he have the necessary training and experience to opine on the scientific foundations of Dr. Freer's Report. The following was confirmed during Mr. Dennis's deposition on May 9, 2023.

- Mr. Dennis graduated from Skidmore College in 2004 with an undergraduate degree in Management and Business, which included courses in accounting, economics, finance and marketing. May 9, 2023 Deposition of Joshua W. Dennis ("Dennis Dep."), Exhibit 12, 33:17-34:34.
- Mr. Dennis remembered taking only one semester-long statistics course as part of his undergraduate studies. *Id.*, 35:20-36:13.
- Mr. Dennis minored in computer science but could remember only one data science course studying how to create relational databases using Microsoft Access. *Id.*, 34:16–35:17.
- After college, Mr. Dennis attended a single "couple of week" SQL programming course, but remembers no other data science study after college. *Id.*, 36:19–38:12.
- Mr. Dennis uses SQL in the majority of his work. *Id.*, 45:15-46:19. SQL is a basic programming language used to perform table queries against a dataset. SQL has a limited number of libraries available to users.

- Python is an open source data science tool used extensively by data scientists to visualize and analyze data using a broad range of available open source libraries. Mr. Dennis has only "some base knowledge" of Python, having used it only "a little bit more than a handful of times." *Id.*, 45:11–45:16.
- Pandas (Python Data Analysis Libraries) is a set of libraries built using Python, which are used for sophisticated data manipulation and analysis. Mr. Dennis has used Pandas only "a handful of times in his career. *Id.*, 45:24–25:12.
- Splink is a probabilistic Python library used by data scientists for record linkage in extremely large datasets that lack unique identifiers (exactly the situation with the Telex data set). Splink incorporates the Fellegi-Sunter Expectation Maximization model of probabilistic record linkage. See generally, https://www.robinlinacre.com/introducing_splink/, attached as Exhibit 13. Spink is used extensively by the US Census Bureau and the UK Office for National Statistics. Splink has been rigorously tested as a canonical model of probabilistic record linkage. See Ex. 9, *Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records*. Splink was used extensively by Dr. Freer in his probabilistic linkage model. Ex. 2, Freer Report, ¶¶ 6, 16. Mr. Dennis has no experience using Splink. Ex. 12, Dennis Dep., 46:20–22.

2. **Mr. Dennis has Minimal Prior Experience with Probabilistic Modeling.**

Mr. Dennis's deposition confirms that he has minimal experience in probabilistic modeling, and what experience he has is unrelated to probabilistic account linkage. When asked to describe his probabilistic computing experience, Mr. Dennis identified only fuzzy matching, "whether it's TF-IDF, or Jaccard Distance or Levenschtein or some combination thereof." Ex. 12, Dennis Dep., 61:24–65:3.

- Fuzzy matching is not probabilistic record linkage. It is a simple method for calculating similarity/distance between two strings. It is used regularly in spell check software. By way of example, below is a Levenschtein matrix calculating the distance between the words Levenschtein and Meilenstein:

		m	e	i	l	e	n	s	t	e	i	n
	0	1	2	3	4	5	6	7	8	9	10	11
l	1	1	2	3	3	4	5	6	7	8	9	10
e	2	2	1	2	3	3	4	5	6	7	8	9
v	3	3	2	2	3	4	4	5	6	7	8	9
e	4	4	3	3	3	3	4	5	6	6	7	8
n	5	5	4	4	4	4	3	4	5	6	7	7
s	6	6	5	5	5	5	4	3	4	5	6	7
h	7	7	6	6	6	6	5	4	4	5	6	7
t	8	8	7	7	7	7	6	5	4	5	6	7
e	9	9	8	8	8	7	7	6	5	4	5	6
i	10	10	9	8	9	8	8	7	6	5	4	5
n	11	11	10	9	9	9	8	8	7	6	5	4

  A simple matrix calculates Levenschtein distance, in this sample, 4 transformations are required to reach equivalence between the words Levenschtein and Meilenstein.

- Probabilistic record linkage, on the other hand, is a far more complex exercise, looking across *multiple* data fields to compute the probability that pairs of records refer to the same entity, often based on a weighted scoring system for each individual field. Ex. 10, Freer Reply, ¶ 6.

- TF-IDF, identified by Mr. Dennis as a fuzzy matching tool, is a statistical measure of how important a given single word is to a document, not a fuzzy matching tool. Ex. 10, Freer Reply, ¶ 8. Mr. Dennis clearly had limited familiarity with TF-IDF (Term Frequency-Inverse Document Frequency) or its actual function.

  Q: You've used TF-IDF?

A: Correct.

Q: TF-IDF stands for what?

A: I'd have to go back and look. I don't recall.

Q: Is it a – is it a fuzzy matching model?

A: It is, yes.

Ex. 12, Dennis Dep., 63:9-16.

Transitive closure is a key aspect of any data linking exercise, including the Freer Aggregation Methodology. Mr. Dennis was unable to identify a prior engagement that required transitive closure. *Id.*, 122:14–23:22. Blocking is a critical element in reducing computation run time when working with large data sets. Absent a blocking scheme, a data set of this size would require trillions of computation. Ex. 2, Freer Report, ¶ 64 ("… in the Telex dataset, which has approximately 11 million user accounts, there would be approximately *60 trillion* pairs to consider.") (Emphasis in original). During his deposition, Mr. Dennis was unable to identify a single prior engagement involving blocking a data set. Ex. 12, Dennis Dep., 74:19 –75:24; 76:21 – 77:23.

3. **Mr. Dennis Has No Experience Working with the Fellegi-Sunter Model that Lies at the Heart of the Freer Aggregation Methodology**

Mr. Dennis has no previous experience working with the Fellegi-Sunter Model at issue in this case, and no understanding of how FS-EM works other that what he has read in the Freer Report. Neither Mr. Dennis nor his team has had any prior experience with the Fellegi-Sunter model, with or without Expectation Maximization. Ex. 12, Dennis Dep., 5:19 – 58:3, 62:24 – 63:3. Mr. Dennis has no understanding as to how FS-EM works other than what Mr. Dennis has read in Dr. Freer's reports. *Id.*, 64:6-11. Mr. Dennis has no understanding of how the FS-EM Model calculates relative field weights other than reading Dr. Freer's description. *Id.*, 63:17 –

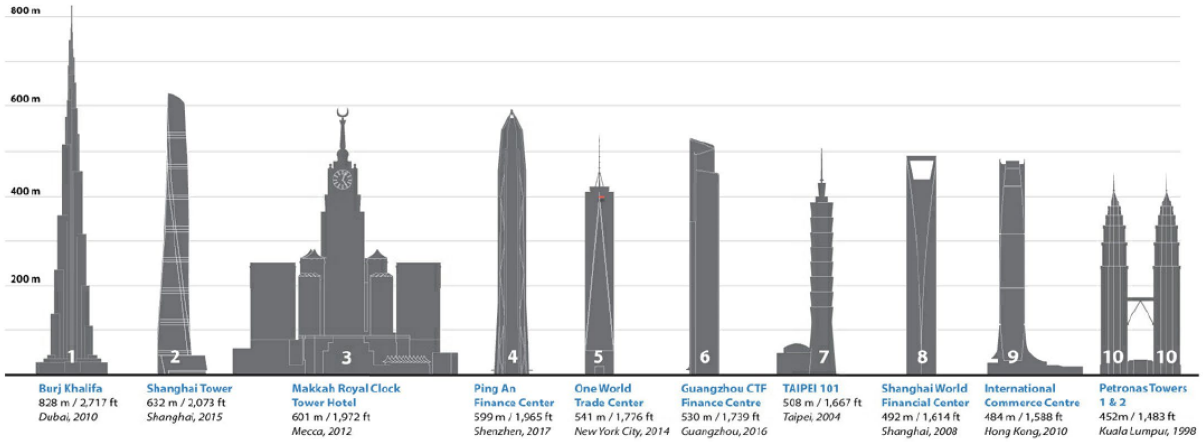
65:8. Mr. Dennis only reviewed minor portions of Dr. Freer's code and offers no admissible opinions on Dr. Freer's relative field weighting choices. Mr. Dennis was provided with detailed instructions on how to load the Freer Aggregation code. Despite having the ability to load and operate the code for all ten steps of the Freer Aggregation Model, Mr. Dennis only loaded code for two of ten steps in the Freer Aggregation Model; blocking and cleaning. *Id.*, 26:3 – 8; 31:18 – 32:12. Mr. Dennis's code review amounts to less than ten percent of the Freer code base. Ex. 10, Freer Reply, ¶ 24. Mr. Dennis can't remember how many steps he actually ran code for. Ex. 12, Dennis Dep., 21:24 – 22:12. Mr. Dennis's choice to limit his code review means that he could not possibly have recreated the Freer Aggregation Methodology steps or output for testing purposes. Of the portions Mr. Dennis did load, Dennis admits that he found no errors in the operation of the code. *Id.*, 72:22 – 73:16.

4. **Mr. Dennis's Critique of Dr. Freer's Aggregation Output Lacks any Statistical Support and Thus Fails the "More Probable than Not" Standard under FRE 702**

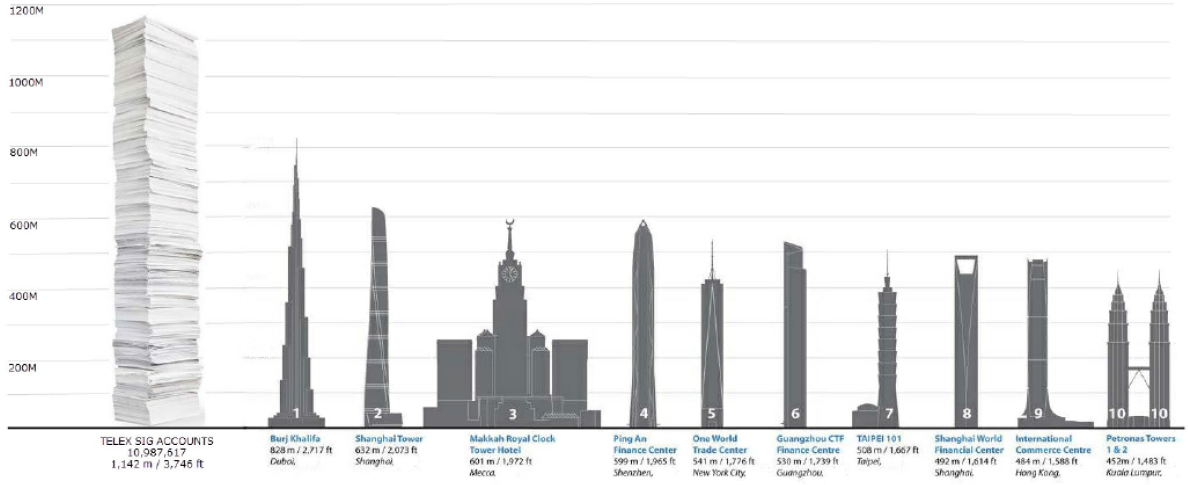
Statistical analysis allows a user to infer set-wide characteristics from a test subset of data. Probability sampling allows users to draw inferences about the characteristics of the entire set population based on the characteristics of the test subset. Dennis's clear bias in selecting a handful of *ad hoc* search results to support his opinion testimony makes Dennis's conclusions inherently unreliable. Modern Scientific Evidence, Vol. I, §4:13, pp. 340-344, attached as Exhibit 14.[28] The TelexFree database contains 10,987,617 accounts exclusive of Ympactus accounts. To understand the magnitude of this database, consider the height of the ten tallest buildings in the

[28] Stated differently, "[E]licitation of expert judgments to produce probability distributions that represent uncertainty about model parameters can be conducted informally, but such judgments are easily affected by unconscious [or conscious] cognitive biases…" The Role of Expert Judgment in Statistical Inference and Evidence-Based Decision Making, The American Statistician, Vol. 73, No. 51, p. 67 (2019) (Text added in bracket), attached as Exhibit 15.

world.



A ream of paper is 500 Sheets. See https://en.wikipedia.org/wiki/Units_of_paper_quantity. Assuming one account per page, a printout of the TelexFree database would require 10,987,617/500 pages equal to 21,975.234 reams of papers. A standard ream of paper is approximately 2.5 inches tall. See https://www.hammermill.com/products/premium-inkjet-laser-printer-paper. So, 21,975 reams times 2.5" equals 54,937.5", 4,578 feet or 1,395 meters, far taller than the tallest building on earth.



Each of those 10,987,617accounts involving 1,566,383 Participants, of which the Freer Methodology has identified 81,681 net winner account clusters.

Against these extraordinary numbers, Mr. Dennis's critique relies on the largely inadmissible affidavit testimony of Franz Balan, whom Mr. Dennis has never spoken to and who purports to testify about what all Participants did based on his own singular experience, together with a handful of tables culled from *ad hoc* queries performed against Freer output tables. Ex. 12, Dennis Dep., 138:10–20. The Balan Affidavit is inadmissible other than to describe his own singular experience.

Offering opinion testimony based on a handful of selected tables to suggest that the Freer Clusters are unreliable violates FRE 702 if Dennis cannot demonstrate that his hand-picked tables are statistically representative of a larger data quality issue. Mr. Dennis, however, lacks the training and experience to demonstrate the statistical significance of his handpicked data tables: Mr. Dennis was unable to describe how to set a sample size relative to the size of a dataset, including what size sampling would be adequate to test the Freer Model. Ex. 12, Dennis Dep., 82:5 – 83:18. Mr. Dennis offers no opinion on whether Dr. Freer's sampling sizes were adequate. *Id*., 83:19 – 84:20. Mr. Dennis was unfamiliar with Cochran's Formula, used by statisticians to calculate sample size within a defined margin of error. *Id.*, 80:9-11. Mr. Dennis could not recall any experience calculating basic statistical notions of precision (how many retrieved items are relevant) and recall (how many relevant items are retrieved). *Id.*, 88:7 – 89:14.

Highlighting his statistical inexperience, Mr. Dennis does not know how to calculate an F measure (a basic statistical measure of accuracy) other than what he has read in the Freer Report. When asked to calculate the false positive rate assuming a precision of 0.980, a recall of 0.996 and an F-measure of 0.988, Mr Dennis was unable to, admitting that he had never calculated false positive rates in an aggregation before. *Id.*, 161:24 – 164:1. The calculation is not difficult;

a precision of 0.980 straightforwardly implies a false discovery rate of 1 – 0.980 = 0.02, meaning 2% of the matches generated from the methodology are expected to be false positives. Ex. 10, Freer Reply, ¶12. Mr. Dennis was also unable to even generally explain the correlation between sample size and population size for purposes of defining the appropriate size of a statistically adequate sample. Ex. 12, Dennis Dep., 82:21–83:18. Finally, Mr. Dennis's selection of which *ad hoc* table queries results to offer in support of his report is classic selection bias, and cannot be suggestive of valid statistical error rates across the Freer cluster output.

Unlike Mr. Dennis, Dr. Freer validated data quality in the TelexFree database via random sampling of 11 fields. For the name field, Dr. Freer's data quality analysis was based on an initial test set of 300 accounts to obtain a 5% Margin of Error with 90% Confidence Level, with a second test set of 1067 accounts confirming the results (with a 3% Margin of Error with 95% Confidence Level). Dr. Freer reviewed the remaining ten fields by manually reviewing a set of 400 randomly accounts chosen accounts. Each test used a fixed-random seed to pick truly random accounts for statistically relevant sampling. Dr. Freer's analysis, detailed in his Reply Report (Ex. 10, Section B., ¶¶ 60-82) is an example of a rigorous statistical analysis that permits scientifically valid inferences to be drawn as to TelexFree data quality.

Mr. Dennis ultimately offers no admissible conclusions about the overall accuracy of the Freer Net Winner Clusters, and instead offers only the *"possibility*" that selected individual clusters *may* include individual over or under account aggregations. Mr. Dennis offers no opinion as to whether Freer's analysis in the *rep_id_summary* table was accurate or reliable for purposes of distinguishing the TelexFree accounts from the Ympactus accounts and hybrid accounts. Ex. 12, Dennis Dep., 157:8–19. Mr. Dennis offers no opinion on Dr. Freer's field cleaning process would have been appropriate. *Id.*, 72:22–73:4. Mr. Dennis offers no opinion

about how Freer's blocking step should have been performed. *Id.*, 197:16-198:12. Mr. Dennis offers no opinion whether Dr. Freer's field weightings were correctly determined. *Id.*, 66:23–67:8. Mr. Dennis offers no opinion on whether transitive closure was a "required step" in Dr. Freer's aggregation process. *Id.*, 134:23–135:10. Mr. Dennis offers no opinion to determine what percentage of the clusters in the aggregation output of the Freer Aggregation Methodology he thinks are erroneous. *Id.*, 94:18–96:2. Mr. Dennis offers no opinion on adequacy of Dr. Freer's expected rates of error. Mr. Dennis formed no opinion on whether a zero overclustering or underclustering error rate is required for a cluster to be acceptable. *Id.*, 159:24–160:5. Mr. Dennis formed no opinion on what error rate would make Freer's clustering methodology acceptable. *Id.*, 160:19–161:19. Mr. Dennis formed no opinion on the number of missed links one would have to examine to validate or invalidate the Freer Aggregation Methodology. *Id.*, 167:7–1. Mr. Dennis performed no analysis of the actual over or underclustering rates in any output cluster generated by the Freer Aggregation Methodology. *Id.*, 96:23–97:16; 114:20–115:8; *see also* Ex. 10, Freer Reply, ¶ 32.

In the end, Mr. Dennis's analysis raises only the "possibility" that the Freer Aggregation Output includes over- or under-aggregations within a given cluster, even in those tables that Mr. Dennis handpicked to include and discuss in his report. Opinions based on "possible" over/under clusters in the Freer Aggregation output is clearly inadmissible under the express *"more likely so than not"* standard required under FRE 702, as expressly called out in FRE 702's 2023 Amendment. By way of example: Mr. Dennis offers no opinion on whether accounts in cluster 132785 (the "Maria Neves" cluster) were in fact over-aggregations, despite referring to them in his Rebuttal as "potential issues". Ex. 12, Dennis Dep., 94:18–96:22. He also did not do any analytical work to specifically arrive at a number of over-aggregations that in his opinion appear

in this cluster, despite discussing this cluster at length in his Rebuttal. Mr. Dennis offers no opinion on whether the "Shal Group" cluster 12650741 and "Dolarex" cluster 10992563, compared in the April 14, 2023 Supplemental Rebuttal Expert Report of Joshua W. Dennis ("2022 Dennis Rebuttal"), Exhibit 16, in paragraph 159 and Exhibit 9.4, should have been aggregated into one cluster or not. Ex. 12, Dennis Dep., 180:2–18, 185:13–22. This is despite specifically referring to these examples as "under-aggregation" in his Rebuttal. Ex. 16, 2022 Dennis Rebuttal, ¶159. Mr. Dennis offers no opinion on whether the "DL1 INC" cluster 11849449, and "Linda Hackett" cluster 707775, compared in his Rebuttal in Exhibit 9.5, should be aggregated into one cluster or not,[48] despite again insinuating in his Rebuttal that it "appears to be under-aggregation". Ex. 16, 2022 Dennis Rebuttal, ¶161.

In the end, Mr. Dennis offers the Court nothing but "maybe" opinions about "possible" errors. His relies on statistically insignificant hand-picked examples that suffer from obvious selection bias, and then concludes nothing based on his sampling. FRE 702 requires both expertise and an accepted scientific process correctly implemented. Mr. Dennis offers the court none of that. His testimony should be afforded little, if any, weight.

IV. CONCLUSION

Dr. Freer's Aggregation Methodology meets all admissibility requirements under FRE 702, and none of Mr. Dennis's unfounded critiques detract from the reliability of Dr. Freer's methods. The Trustee accordingly requests that this Court find that:

i) The 11 TelexFree data fields relied upon by Dr. Freer in his 10-step process have sufficient quality to allow Freer to reliably aggregate Participant accounts into Net Winner clusters;

ii) Dr. Freer made appropriate usage of the selected data fields for arriving at his net winner clusters by using a 10-step process that meets the reliability standard in FRE 702; and

iii) Mr. Dennis's Report provides no useful or reliable critique of the Freer Methodology.

For those reasons, the Court should deny Class Representative's *Daubert* Motion, and Dr. Freer's net winner cluster output opinion testimony should be allowed into evidence.

Respectfully submitted,

STEPHEN B. DARR, LIQUIDATING TRUSTEE,
By his counsel:

Dated: September 11, 2023

/s/ Daniel J. Lyne
Daniel J Lyne (BBO #309290)
Alexandra Papas (BBO #707581)
MURPHY & KING,
Professional Corporation
28 State Street, 31st Floor
Boston, MA 02109
Telephone: (617) 423-0400
Dlyne@murphyking.com
Apapas@muprhyking.com

CERTIFICATE OF SERVICE

I, Daniel J. Lyne, hereby certify that on September 11, 2023, a copy of this document was served by operation of the Court's ECF System upon the registered participants as identified on the Notice of Electronic Filing.

Date: September 11, 2023

/s/ Daniel J. Lyne
Daniel J. Lyne (BBO# 309290)
MURPHY & KING,
Professional Corporation
28 State Street, 31st Floor
Boston, MA 02109
Telephone: 617-423-0400
Dlyne@murphyking.com

EXHIBIT 1

Tell us whether you accept cookies

We would like to use cookies to collect information about how you use **ons.gov.uk**.

We use this information to make the website work as well as possible and improve our services.

Accept all cookies

Set cookie preferences

Office for National Statistics

Data and analysis from Census 2021

Developing standard tools for data linkage: February 2021

Discussing the seven functions that were created to automate part of the data linkage pipeline to improve efficiency and quality.

In this section

1. Overview of data linkage
2. Main approaches to data linkage
3. Expectation maximisation algorithm
4. Data protection
5. Data pre-processing
6. Study objectives and approach
7. Methodology used
8. Main functions
9. Details of functions
10. Conclusions
11. Authors
12. Related links

1. Overview of data linkage

Data linkage is the process of joining together records that pertain to the same entity, such as a person or business. In its most simplistic form, links (or “matches”) can be completed by comparing unique identifier keys specific to each object, for example, a person’s National Insurance number. Connecting data in this way can lead to new insights and discoveries and reduce the need for additional data collection.

As technology advances and computational power grows, it is possible to link larger datasets more efficiently. However, considerable time and effort must be invested in preparing the data prior to linking, to increase the chance of detecting matches while ensuring accuracy is preserved.

This article discusses seven functions that were created to automate part of the data linkage pipeline to improve efficiency and quality. The tools target the preliminary stages of data linkage, however, can also be used for quality evaluation post-linkage.

Each function was coded in PySpark, a Python-based application programming interface (API) appropriate for data hosted on secure distributed computing systems. Five of the functions perform exploratory data analysis and provide linkage-specific metadata, such as extracting common values and providing general descriptive statistics.

The sixth function creates a “candidate links” table from the data; a detailed record of how each data pair compares across every attribute. This is a necessary step in probabilistic data linkage.

The final function employs the Expectation Maximisation (EM) algorithm, an iterative unsupervised machine learning procedure, to estimate the statistical parameters required for probabilistic data linkage.

2. Main approaches to data linkage

Notice regarding the data linkage formula for accuracy (added 27 January 2023)

The Office for National Statistics recommends always reporting on the quality of data linkage using the metrics precision and recall as stated in this section. Originally a formula for accuracy was included in this section but has now been removed as it should not be used. It did not give a good representation of the quality of the linkage and was difficult to interpret. It was never used at the ONS. If you have any questions please contact datalinkage@ons.gov.uk.

Data linkage is a multi-stage process. It is possible to introduce error at each of these stages therefore the potential for compound error is widely acknowledged (see GUILD: Guidance for Information about Linking Data (https://pubmed.ncbi.nlm.nih.gov/28369581/), Matching and record linkage (https://www.census.gov/srd/papers/pdf/rr93-8.pdf), Quantitative data cleaning for large databases (https://dsf.berkeley.edu/jmh/papers/cleaning-unece.pdf), Data quality and record linkage techniques (https://www.researchgate.net/publication/220695391_Data_Quality_and_Record_Linkage) for more information).

Identifying matches

To measure and minimise this error, it is first necessary to quantify the number of missed matches and false matches. Missed matches, also known as false negatives or Type II errors, are when records are not identified as a match when in fact they are. Conversely, false matches (false positives, or Type I errors) are when records are incorrectly identified as a match.

Matches can be identified as follows:

- where a predicted positive is an actual positive, that is a “true positive”
- where a predicted positive is an actual negative, that is a “false positive”
- where a predicted negative is an actual negative, that is a “true negative”
- where a predicted negative is an actual positive, that is a “false negative”

Depending on the context, one type of error may be more significant than another. For example, if data linkage is being used in relation to a medical diagnosis it may be favourable to obtain a false positive over a false negative; the latter being a missed diagnosis.

There are various statistics that incorporate these errors and can be used to assess the performance of data linkage. To calculate these, it is necessary to count the number of true and false negatives and positives, which can be recorded in a table known as a confusion matrix.

We can calculate common performance statistics such as recall, precision and the F score with the following formulas:

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negatives)}$$

$$Precisions = \frac{True\ Positives}{(True\ Positives + False\ Positives)}$$

$$F = 2\left(\frac{Precision\ \times Recall}{Precision + Recall}\right)$$

Main approaches

There are three major approaches to data linkage:

- clerical
- deterministic
- probabilistic

Clerical approach

The clerical approach is a manual process whereby a person physically examines each combination of records and determines whether they match or not. Even with two relatively small datasets, this is a lengthy and arduous process, rendering it unrealistic as a primary method.

For example, given two datasets each containing 100 records, the total possible pairs to evaluate would be 10,000. It is also liable to subjectivity and human error, therefore is mostly used to confirm ambiguous matches (see Matching for automatic record matching and linkage and their use in national statistics (https://s3.amazonaws.com/escoe-website/wp-content/uploads/2019/11/21120237/GSS-Methodology-Series-No.-25-Methods-for-automatic-record-matching-and-linkage-and-their-use-in-national-statistics.pdf)), or assess true positive rates to calculate recall.

Deterministic approach

Deterministic methods employ exact matching techniques, where attributes for a given record pair must match according to some rule. The simplest form of deterministic matching is computationally fast: each attribute in a record pair is compared, and the attributes either match or do not match (see Matching for automatic record matching and linkage and their use in national statistics (https://s3.amazonaws.com/escoe-website/wp-content/uploads/2019/11/21120237/GSS-Methodology-Series-No.-25-Methods-for-automatic-record-matching-and-linkage-and-their-use-in-national-statistics.pdf)). For example, in person-related data the attributes may consist of date of birth, name, and address. If each of these attributes were identical then the record pair would be classed as a match.

Given the speed of deterministic linkage, it is often used as a first pass step to remove records from the pool of possible matches, thereby reducing the load for subsequent probabilistic matching.

Probabilistic approach

Probabilistic methods comprise the focus of this project and involve the calculation of conditional probabilities to determine the likelihood that a record pair is a match. This method was detailed by a seminal paper on probabilistic data linkage from 1969 (https://courses.cs.washington.edu/courses/cse590q/04au/papers/Felligi69.pdf), and formalised the ideas first presented by Newcombe and others in their 1959 article Automatic linkage of vital records (http://www.cs.umd.edu/class/spring2012/cmsc828L/Papers/Newcombe59.pdf).

For a given attribute, where examples of attributes include dates, addresses, names and so on, two conditional probabilities, *m* and *u*, are calculated. The *m* probability is the probability those attributes match given the record pair belongs to the same unit. This allows for some error in the data. For example, if sex was recorded incorrectly in 5% of record pairs, then the *m* probability for sex would be 0.95 (see Challenges in administrative data linkage for research (https://journals.sagepub.com/doi/pdf/10.1177/2053951717745678)).

The *u* probability is the probability the attributes match given the record pair belongs to different units. This can be thought of as the probability of chance agreement. Using sex again, a reasonable estimate would be *u* equals 0.5. Agreement and disagreement weights, W_A and W_D respectively, are then calculated for each attribute using *m* and *u* as follows in equations 1 and 2.

Equation 1

$$W_A = log_2\left(\frac{m}{u}\right)$$

Equation 2

$$W_D = log_2\left(\frac{1-m}{1-u}\right)$$

The calculated weights are summed for each record to produce a score, where the choice of weight for each attribute is determined by the match status for that attribute. For instance, if date of birth matches for a record pair then the agreement weight is used, if it does not match the disagreement weight is used. An example is presented in Tables 1a and 1b.

Table 1a: Example agreement and disagreement weights

Parameter	Forename	Surname	Date of birth	Post code
W_A	7.17	7.11	6.01	4.03
W_D	-1.87	-1.66	-6.06	-1.26

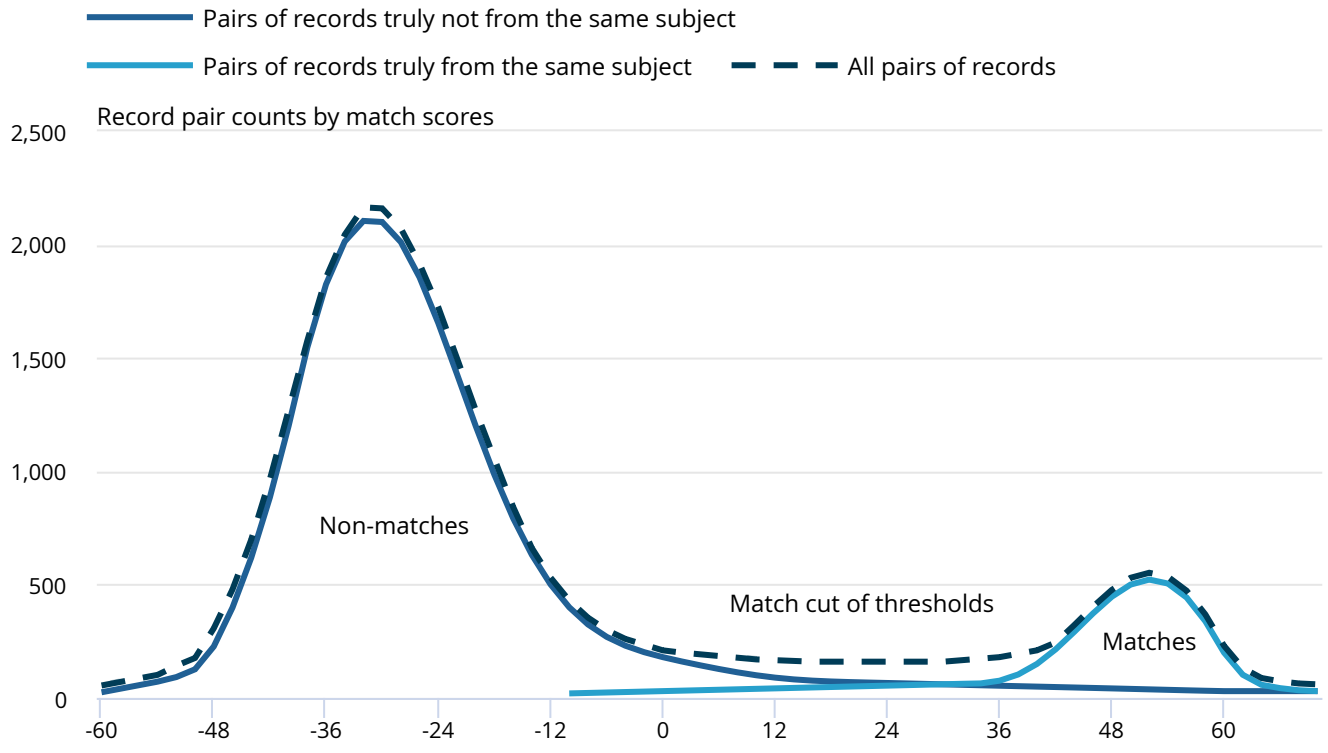
Table 1b: Calculating match scores from agreement and disagreement weights

Record pair	Forename	Surname	Date of birth	Post code	Score
Pair 1	Match W_A = 7.17	Non-match W_D = -1.66	Non-match W_D = -6.06	Non-match W_D = -1.26	-1.81
Pair 2	Non-match W_D = -1.87	Match W_A = 7.11	Non-match W_D = -6.06	Match W_A = 4.03	3.21
Pair 3	Match W_A = 7.17	Match W_A = 7.11	Match W_A = 6.01	Non-match W_D = -1.26	19.03

Match scores are calculated for records where the match status is already known, and the frequency distributions of the scores are examined to determine thresholds. Above the threshold, the records would be classed as a match, and below the threshold the records would be classed as a non-match (Figure 1).

The thresholds can then be applied to record pairs with an unknown match status to determine if they are a match or not. It is possible to use multiple thresholds, and class record pairs between these bounds as ambiguous matches. The match status can then be confirmed by clerical review. Thresholds can also be moved to tune the sensitivity of the linkage depending on the application.

Figure 1: Example of how to evaluate match score thresholds



Source: Office for National Statistics – Developing standard tools for data linkage

Notes:

1. This figure is based on example data and is adapted from a similar figure from John 2019, Cardiff University internal document.

Given the size of many datasets, a technique known as blocking is often employed during probabilistic linkage, where pairs of records that are unlikely to match are removed from consideration to reduce the search space. Grannis and others 2003 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1479910/) liken it to "sorting socks by colour

before pairing them". A block is created on a given variable, where all records in the block must agree on that variable. For example, a block on name and date of birth would only contain records where both name and date of birth match exactly. In this sense, a block is essentially a preliminary pass of low-fidelity deterministic linkage.

Of the three approaches to data linkage described, probabilistic methods are the most widely used and are therefore the focus of this project. They are the most computationally expensive, but are more robust against errors and more effective with large databases (see The effect of data cleaning on record linkage quality (https://www.researchgate.net/publication/237058947_The_effect_of_data_cleaning_on_record_linkage_quality)).

3. Expectation maximisation algorithm

Winkler (2000) (https://courses.cs.washington.edu/courses/cse590q/04au/papers/WinklerEM.pdf) proposes a method for estimating *m* and *u* probabilities for probabilistic linkage using the Expectation Maximisation (EM) algorithm. EM is an unsupervised, iterative machine learning algorithm that is relevant to many applications where data are incomplete. In the context of data linkage, the "incompleteness" is because of the fact the true match status for a record pair is unknown.

The EM algorithm is used to calculate maximum likelihood estimates for certain parameters, in this case the *m* and *u* probabilities, and an additional parameter *p*, which is the total proportion of records that match in the entire dataset.

The EM algorithm iterates back and forth between two steps: expectation and maximisation.

Step 1: Expectation

Given initial guesses of *m*, *u* and *p*, a parameter *g* is estimated, which is a binary indicator variable representing the match status for each record pair (Equation 3). *g* = 1 if the record pair is a match, and 0 if not.

Step 2: Maximisation

Using *g* calculated in the expectation step, the proportion of matches and non-matches are used to update *m*, *u* and *p* (Equations 4, 5 and 6).

m, *u* and *p* are returned to the expectation step to re-calculate *g*.

Formulas

To formulate this mathematically, we define g_j as a binary indicator variable representing the overall match status of each record pair:

$$g_j = \begin{cases} 1 & \text{if record pair } rj \text{ is a match} \\ 0 & \text{otherwise} \end{cases}$$

We also define γ^j_j as a binary indicator variable, representing the match status of each attribute within each record pair:

$$\gamma_i^j = \begin{cases} 1 & \text{if attribute } i \text{ matches for record pair } rj \\ 0 & \text{otherwise} \end{cases}$$

Where $i = 1, 2, \ldots, n$, represents each of n attributes, and $j = 1, 2, \ldots, N$, represents each of N record pairs in the dataset. We define:

- $\hat{m}_i$ as the estimate of the m probability for attribute i
- $\hat{u}_i$ as the estimate of the u probability for attribute i
- $\hat{p}$ as the estimate for the proportion of record pairs j that are matches

The equation for the expectation step to calculate an estimate of g for each record pair j, is as follows (see Data Quality and Record Linkage (https://www.researchgate.net/publication/220695391_Data_Quality_and_Record_Linkage)) in Equation 3.

Equation 3

$$\hat{g}_j = \frac{\hat{p} \prod_{i=1}^{n} \hat{m}_i^{\gamma_i^j} (1 - \hat{m}_i)^{1-\gamma_i^j}}{\hat{p} \prod_{i=1}^{n} \hat{m}_i^{\gamma_i^j} (1 - \hat{m}_i)^{1-\gamma_i^j} + (1 - \hat{p}) \prod_{i=1}^{n} \hat{u}_i^{\gamma_i^j} (1 - \hat{u}_i)^{1-\gamma_i^j}}$$

The maximisation step involves the re-calculation of $\hat{m}_i$, $\hat{u}_i$ and $\hat{p}$ using the following equations (see Data Quality and Record Linkage (https://www.researchgate.net/publication/220695391_Data_Quality_and_Record_Linkage)).

Equation 4

$$\hat{m}_i = \frac{\sum_{j=1}^{N} \hat{g}_j \bullet \gamma_i^j}{\sum_{j=1}^{N} \hat{g}_j}$$

Equation 5

$$\hat{u}_i = \frac{\sum_{j=1}^{N} (1 - \hat{g}_j) \bullet \gamma_i^j}{\sum_{j=1}^{N} (1 - \hat{g}_j)}$$

Equation 6

$$\hat{p} = \frac{\sum_{j=1}^{N} \hat{g}_j}{N}$$

The complete data log likelihood (Equation 7) is calculated at each iteration, and iterations continue until successive values for the log likelihood converge.

Equation 7

$$log\ likelihood = \sum_{j=1}^{N} g_j \log\left(p \bullet \prod_{i=1}^{n} \hat{m}_i^{\gamma_i^j} (1 - \hat{m}_i)^{1-\gamma_i^j}\right) + (1 - g_j)\ log\left((1-p) \bullet \prod_{i=1}^{n} \hat{u}_i^{\gamma_i^j} (1 - \hat{u}_i)\right.$$

There are many advantages of using the EM algorithm to calculate *m* and *u* probabilities for probabilistic matching. Firstly, it is an unsupervised learning method, which means there is no requirement to provide a set of training data to "teach" the algorithm before it can be used. If

training data were required, then a lengthy clerical review process would first be needed to produce a set of true matches and non-matches.

Secondly, it has been shown to perform well on records with minor typographical errors (see Data Quality and Record Linkage (https://www.researchgate.net/publication/220695391_Data_Quality_and_Record_Linkage)). Finally, it is not overly sensitive to the initial parameter estimates of *m*, *u* and *p* (see Data Quality and Record Linkage (https://www.researchgate.net/publication/220695391_Data_Quality_and_Record_Linkage)), which must be provided before running the algorithm.

Regardless of these initial estimates, convergence usually occurs quickly, albeit that there is no guarantee of finding the global optimum (see Using the EM Algorithm for Weight Computation in the Fellegi-Sunter Model of Record Linkage (https://courses.cs.washington.edu/courses/cse590q/04au/papers/WinklerEM.pdf)). The efficiency and accuracy of the EM algorithm highlights it as a valuable tool in the probabilistic linkage pipeline. As such, it formed a target area for this project.

4. Data protection

Data protection and associated ethical obligations are important concepts in record linkage. Some databases may not breach data protection rules in isolation but would post-linkage, for example, after linking a database of medical conditions to name and address details. Confidentiality of personal data is vital and must be strictly observed to maintain trust and comply with legislation. A balance must be sought between protecting privacy while producing statistics at a sufficient resolution to be meaningful. To minimise risk, the tools created as part of this project were developed on synthetic data, which was generated to mimic the statistical properties of real data.

5. Data pre-processing

Data pre-processing, standardisation, or cleaning is the process of forcing data into the same case, format and content, and is an important part of any data analysis. For instance, removal of non-alphanumeric characters, creating consistency in date formats, or parsing (splitting) an attribute into its component parts. This process greatly increases the chances of finding a deterministic match.

For example, comparing the name strings "Smith, Mr Bob", and "Robert Smith" would not result in a match, but if the forenames were standardised, and parsed into forename and surname fields, then the match status would be true (see Data Quality and Record Linkage (https://www.researchgate.net/publication/220695391_Data_Quality_and_Record_Linkage)). There is, however, evidence to suggest that over-standardisation can be detrimental to data linkage (see The effect of data cleaning on record linkage quality (https://www.researchgate.net/publication/237058947_The_effect_of_data_cleaning_on_record_linkage_quality)). This is because variability can be reduced to the extent that items become indistinguishable from one another, resulting in an increased risk of false positive matches.

Pre-processing includes the identification of missing values. This is particularly important in the context of data linkage because if an attribute is missing from one record then its match status with another record cannot be conclusive. Assumptions can lead to false positives or false negatives, which in turn lead to defective *m* and *u* probabilities in probabilistic linkage.

Derived variables, such as string comparators, are also calculated as part of the pre-processing stage. They provide a numerical "distance" between words based on spelling and are used in data linkage to permit a match even when there are minor differences. For example, an exact comparison between the names "Catherine" and "Katherine" would result in a match status of "false", despite the spellings being almost identical. However, a comparison allowing for a string distance of 1 would result in a match status of "true".

There are many string comparators, but the Levenshtein distance is popular in a data linkage context. This quantifies the number of insertions, deletions, or replacement of characters to transform one string into another. It can be expressed mathematically as follows in Equation 8, given two strings, *a* and *b*, where $lev_{a,b}(i, j)$ is the Levenshtein distance between the *i* characters in *a*, and *j* characters in *b*.

Equation 8

$$lev_{a,b}(i,j) = \begin{cases} max(i,j) \text{ if } \min(i,j) = 0 \\ min \begin{cases} lev_{a,b}(i-1,j)+1 \\ lev_{a,b}(i,j-1)+1 \\ lev_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} \text{otherwise} \end{cases}$$

$lev_{a,b}(i\text{-}1, j) + 1$ represents a deletion, $lev_{a,b}(i, j\text{-}1) + 1$ represents an insertion, and $lev_{a,b}(i\text{-}1, j\text{-}1)+ 1_{(a_i \neq b_j)}$ represents a match or mismatch.

$1_{(a_i \neq b_j)}$ is a binary indicator variable equal to 0 when $a_i = b_j$, and 1 otherwise.

As an example, the Levenshtein distance between "Amy" and "Aimee" is three. This is because the transition requires two insertions ("i" and "e") and one replacement ("e" for "y").

Another variable that can be derived from string data is a phonetic code. As with string comparators, phonetic codes provide flexibility when comparing strings by allowing for slight differences in the data. This is achieved by re-coding strings according to their pronunciation. There are many types of phonetic codes, but Soundex is widely used.

A Soundex code consists of the first letter of the string followed by three digits, where each digit represents one of six consonant sounds. The first letter of the string is retained, any vowels dropped, then remaining consonants represented by a number between one and six.

For example, the Soundex code for "Amy" is A500, which is identical to the Soundex code for "Amiee". Comparing the Soundex code would enable a match on these names, which would not be possible with exact matching.

A final part of data pre-processing essential to probabilistic linkage is the generation of a candidate links dataset. This is a table that records the match status of each attribute for a given record pair.

Derived variables can also be compared, such as string distance measures or phonetic codes. In the case of a string distance measure, the output is an integer, which must then be converted to true or false for inclusion in a candidate links dataset. It is therefore necessary to select a threshold, above which the match status should be considered false. The threshold can be changed depending on the requirements of the given linkage project.

An example candidate links dataset is shown in Table 2. It compares records where derived variables have been used in place of forename and surname, but an exact comparison has been used for date of birth. For instance, Pair 3 may comprise:

- forename Soundex codes: A500 and A500 (match status equals true)
- surname Levenshtein distance: 0 (match status equals true)
- date of birth: 02/11/2020 and 10/11/2020 (match status equals false)

Table 2: Example of candidate link dataset

Candidate link pair	Forename (Soundex comparison)	Surname (Levenshtein distance)	Date of birth (exact comparison)
Pair 1	False	False	False
Pair 2	False	True	True
Pair 3	True	True	False

Pair 4	False	False	False

The main purpose of creating a candidate links dataset is to record the match status of each attribute to select the correct weight, either agreement or disagreement, when calculating match scores. However, it is also a requirement of the Expectation Maximisation (EM) algorithm when iteratively calculating *m* and *u* probabilities, as the true or false status is represented by the variable γ^i_j (see Section 3 (https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/developingstandardtoolsfordatalinkagefebruary2021#expectation-maximisation-algorithm)). It is therefore a necessary step in the probabilistic linkage pipeline and so formed another target area for this project.

6. Study objectives and approach

The objectives of this study were to increase the efficiency of the probabilistic data linkage pipeline by creating custom software tools to automate certain tasks. The tools were required for use at the Office for National Statistics (ONS). Commercially available software packages were not appropriate given they are often context-specific and provide limited transparency and control for the user. There may also be concerns over data security.

Three essential steps in the probabilistic linkage pipeline were selected for focus where automation is most widely useful:

- step one: generating linkage-related metadata
- step two: creating candidate link datasets
- step three: implementing the Expectation Maximisation (EM) algorithm to estimate *m* and *u* probabilities and their associated agreement and disagreement weights for use in probabilistic linkage

In relation to the first point, examples of linkage-related metadata include the number of records in each database, the total link space (total number of possible comparisons), quantification of missing data, recognising unique identifiers for use as match keys, and generating general descriptive statistics.

Automating step one delivers consistency and speed in identifying such data features but can also be used to provide quality evaluation post-linkage. Automating the production of a candidate links dataset for step two provides the required input for step three where *m* and *u* probabilities and agreement and disagreement weights are estimated.

7. Methodology used

The Office for National Statistics (ONS) collects and stores a vast quantity of data relating to the UK economy, population and society. This necessitates the use of a sophisticated distributed computing system for storage and data analysis. Because of the sensitive nature of the data, it must also be highly secure.

To achieve this end, the ONS uses Apache Spark technology (https://spark.apache.org/docs/2.1.0/api/python/pyspark.sql.html), which enables efficient handling of large datasets by sharing the workload across multiple networked computers. To uphold a high level of security, the system is only accessible from the ONS offices and users must have security clearance.

Additionally, it is not possible to create a user interface for applications running on the system or run code through the command line. Any functions created as part of this project had to adhere to these conditions, and as such outputs were provided as "raw" code in the form of *.py* scripts. However, this has the dual advantage of full code transparency and adaptability for future use.

The traditional language of communication with Apache Spark is Scala; however, Python and R Application Programming Interfaces (API's) can also be used in the form of PySpark and SparklyR. PySpark was selected for this study because it is widely used within the ONS and has a straightforward syntax. The documentation is thorough, and implementation is widely discussed by the online community. A disadvantage, however, is efficiency: operations in Scala are up to 10 times faster than their PySpark equivalents (see Scala vs. Python for Apache Spark (https://www.dezyre.com/article/scala-vs-python-for-apache-spark/213)). That being said, the syntax of Scala is much more complex than Python, therefore would negate the requirement to provide accessible, transparent code.

All functions created for this project were initially tested on synthetic data, which are fabricated datasets that mimic the structure, content and errors of real data. Such datasets are ideal for testing analytical processes, since they can be "damaged" in a controlled way and then compared with the original version. False positives and false negative rates are known and therefore can be used to accurately calculate precision and recall without the need for clerical matching. Synthetic data also have no risk associated with data protection.

The first step in this study was to undertake a thorough literature review to fully understand the processes involved in data linkage, and to help in selecting which parts of the pipeline to target for automation. Discussions were then had with important stakeholders to further refine requirements, after which pseudocode was developed to map out each function in more detail. The pseudocode was then implemented in PySpark, where each stage was tested to ensure results compared with expectations.

It was then wrapped into functions and further tested on the synthetic datasets, comparing outputs with the relevant non-damaged files where appropriate. Finally, functions were released to the Data Linkage Expert Group at the ONS for user acceptance testing on real data. Any issues were resolved and returned to the user for final testing and checking.

Code was written with speed and efficiency in mind throughout development: for example, all operations were performed on PySpark dataframes, with spurious columns dropped within functions to reduce computational expense. Where possible, function outputs are given in Pandas dataframe format.

Pandas is a flexible Python library that allows manipulation and analysis of data tables. Pandas dataframes offer a wider range of functionality compared with PySpark dataframes, such as extraction of data using row indexing. They can also be rendered to Hypertext Mark-up Language (html), which is more aesthetically pleasing and affords clarity to the output. However, they are unable to deal with extremely large datasets, hence data manipulation is performed on PySpark dataframes and only the output provided in Pandas format.

All functions contain doc strings detailing the purpose, input parameters and outputs to assist the user in running the code. Additionally, a user guide was developed, which explains how to use each function in detail.

8. Main functions

A total of seven functions were developed as part of this project and are summarised in this section.

General summary function

Purpose: Provide general linkage-specific metadata for multiple dataframes.

Output: Pandas dataframe tabulating metadata for given input dataframes.

Missing function

Purpose: Summarise the missingness per column. Enables user to make informed decisions on imputation.

Output: Pandas dataframe tabulating missingness of each attribute.

Unique function

Purpose: Summarise the distinct values per column. Assess suitability of each column for use as identifier keys and blocking variables.

Output: Pandas dataframe tabulating uniqueness of each attribute.

Common records function

Purpose: Summarise the number of distinct records in common between two input dataframes. Assesses potential bias in linkage.

Output: Pandas dataframe tabulating distinct records and associated counts.

Descriptive statistics function

Purpose: Provide descriptive statistics per column for each dataframe. Quality evaluation post-linkage.

Output: Pandas dataframe tabulating descriptive statistics.

Candidate links function

Purpose: Generate a candidate links dataset from random samples of two input dataframes. For deterministic and probabilistic linkage, plus anonymisation.

Output: PySpark dataframe showing Boolean match status per attribute, and Levenshtein distance for strings

m u probability function

Purpose: Iteratively estimate *m* and *u* probabilities and associated agreement and disagreement weights using the EM algorithm.

Output: Pandas dataframe tabulating *m* and *u* probabilities and agreement and disagreement weights for each attribute.

The prior assumptions for running the functions are as follows:

- the PySpark session has been configured (the current version of PySpark, PySpark 3.0.1, requires Python 2.7 and Pandas 0.23.2 as a minimum)
- each dataset for linkage has been loaded as a separate PySpark dataframe
- the PySpark dataframes have a "flat" structure: they are simple two-dimensional tables consisting of single rows of data with no nested information

Most of the diagnostic functions take an arbitrary number of input databases where the upper limit is restricted by computational power. Larger datasets have slower run time; however, the exact run time will depend on the computational resources such as the configuration of the PySpark session. Run time can be minimised by limiting the number of dataframes passed into functions and dropping any spurious attributes.

9. Details of functions

Diagnostic functions

Five of the functions can be classed as diagnostic functions, as they provide metadata about the datasets provided for linkage. These five functions are:

- General summary function
- Missing function
- Unique function
- Common records function
- Descriptive statistics function

These are described in detail in this subsection.

The general summary function gives a broad overview of the datasets passed to it. It can assist in finding compatible variables that could be used as match keys, while identifying areas than may require data cleaning or reformatting. This tool has the added advantage that it can be applied outside data linkage to give a quick summary of the structure and content of any two-dimensional dataframe.

Any comparisons completed by the function, such as searching for common column names, are conducted across all input dataframes. For instance, if four dataframes are passed as inputs, of which only three have identical column names, then no commonality will be detected. The objective of the general summary function is to provide a broad overview of the dataframes of interest.

Subsequent functions can then be used to provide more detail on certain aspects. Two or more dataframes can be passed as arguments, and the output is a single Pandas dataframe containing the following attributes:

- the total link space stating the total possible comparisons between all records across all input dataframes
- **File_Name**: String – the names assigned to the input dataframes when loaded
- **Row_Count**: Integer – the number of records in each dataframe
- **Column_Count**: Integer – the number of columns in each dataframe
- **Common_Columns**: List – the common column names between all dataframes in alphabetical order
- **Common_Columns_All_Data_Types**: List – the common column names across all dataframes and their associated data types; each item in the list is a tuple of (column name, data type) in alphabetical order and this allows identification of columns where names may be shared but data types are not, highlighting where data types require changing such that they are compatible for linkage
- **Common_Columns_Common_Data_Types**: List – the common column names across all dataframes, only where the associated data type is also in common; each item in the list is a tuple of (column_name, data_type) in alphabetical order by column name
- **Unique_Columns**: List – the column names unique to each dataframe in alphabetical order
- **Empty_Cells_Count**: Integer – the number of empty cells per dataframe
- **Empty_Cells_%**: Float – the proportion of empty cells per dataframe as a percentage of all cells in the dataframe

The missing function identifies the quantity and location of missing data and rapidly allows the user to determine imputation requirements. As with the general summary function, it also has applicability outside the field of data linkage.

Imputation is beyond the scope of this project, but this function provides sufficient information for targeting missing fields should this be necessary, in addition to understanding the impact imputation may have on the structure of the data.

One or more dataframes can be passed as arguments in any order, and the output is a single Pandas dataframe, listing the following attributes against each input column name:

- **Count_Missing**: Integer – the count of missing cells per column
- **Percentage_Missing**: Float – the proportion of missing cells per column as a percentage; this is calculated by dividing the count by the total number of rows in the given column and the result is given to six decimal places as this degree of accuracy may be relevant for large datasets with a low number of missing cells
- **File**: Integer – the corresponding file (dataframe) number for that row; this allows differentiation between column names that may be common to multiple dataframes

The unique function identifies if any variables have distinct values. It can be used to evaluate the potential of each variable for use as a match key in deterministic linkage, and to inform decision-making on blocking criteria.

Most datasets are unlikely to contain unique identifiers, therefore the function supplies information on how close each attribute is to being unique. Fields can be concatenated to derive a new match key from multiple non-unique match keys. These derived variables can be passed to the function a second time to test for uniqueness.

The unique function can also inform on blocking criteria. For example, blocking on a variable that has a high distinct count would result in very small blocks and therefore a higher risk of missing potential matches. Equally, a variable with a low number of distinct values would result in very large blocks, which would not substantially reduce the total link space. One or more dataframes can be passed as arguments in any order.

The output is a single Pandas dataframe providing the following information for each input column:

- **File**: Integer – the corresponding input dataframe number for that row; this allows differentiation between dataframes, which share column names

- **Distinct_Values_Count**: Integer – the count of distinct values per column, if duplicates were removed, for example, day of the month could reasonably be expected to have 31 distinct values, and month of the year to have 12; deviation from this would warrant further investigation or indicate that fields may have been swapped (for instance, days recorded in the months column)
- **Distinct_Values_%**: Float – the proportion of distinct values per column as a percentage of non-null values only; the count of the distinct values is divided by the number of non-null entries in the given column then multiplied by 100
- **Distinct_%_Standardised**: Float – The proportion of distinct values per column as a percentage of all values, including nulls; the count of the distinct values is divided by the total number of rows in the dataframe, then multiplied by 100 and this figure is useful in the context of columns with many missing values where Distinct_Values_% would be misleading
- **Count_Per_Distinct**: Float – the count of non-null values divided by the count of distinct values provides a rough indication of the number of times each distinct value occurs; if Count_Per_Distrinct equals 1, all non-null items are distinct, so each value in the column is unique
- **Use_As_Identifier**: Boolean – a unique identifier key requires the number of distinct values to equal the number of non-null values; in this case, Use_As_Identifier equals True and this is unlikely with real data, therefore Distinct_Values_% can be used as an indication of how close the column is to being unique

The common records function returns values that only exist in both dataframes for the given column, automatically removing duplicates. It is an extension of the unique function in that it provides the actual distinct values themselves in addition to quantities. It can help establish linkage quality as values that are not shared between datasets can result in linkage bias.

Only two dataframes can be passed to the function as it uses the PySpark intersect method, which requires two positional arguments. The function will only perform on columns that share both name and data type: the general summary function can be used first to identify these columns.

For each column name common to both input dataframes, the output details:

- **Common_Distinct_Records**: List of strings – this returns values that are in both dataframes for the given attribute, automatically removing duplicates
- **Common_Distinct_Records_Count**: Integer – a count of values that are only in both dataframes; this is the length of the list Common_Distinct_Records

The descriptive statistics function is used to provide general statistics on the input dataframes. It can also be used post-linkage for quality evaluation by comparing outputs between individual and linked datasets. If linkage has been conducted adequately, then distributions should be preserved post-linkage.

It is also a useful tool outside the field of data linkage, as many data analysis projects require the extraction of general descriptive statistics as part of exploratory data analysis. The first argument that must be passed to this function is *error*, a float greater than or equal to 0 and less than or equal to 1. This dictates the acceptable error threshold when calculating quantiles.

As error approaches zero, run time increases, which is why the acceptable error is left to the user's discretion. After *error*, the use must pass one or more dataframes in any order. The output is a single Pandas dataframe, which details the following for each column:

- **Data_Type**: String – the data type for the named column
- **Count**: Integer – the count of non-null values per column
- **Min**: Integer or float – the minimum value for the named column

- **Max**: Integer or float – the maximum value for the named column; useful for range checks, for example, a value greater than 12 in a month column would warrant further investigation
- **Mean**: String, integer or float (the output data type will be the same as the input data type) – the mean value for the named column, not applicable for string data
- **25th_Percentile**: Integer or float – the 25th percentile for the named column, calculated using the acceptable error threshold, *error*, specified by user when calling the function; not applicable for string data
- **Median**: Integer or float – the median value for the named column, calculated using *error*; not applicable for string data
- **75th_Percentile**: Integer or float – the 75th percentile for the named column, calculated using *error*; not applicable for string data
- **Std_Dev**: String, integer or float (the output data type will be the same as the input data type) – the standard deviation for the named column; not applicable for string data
- **Variance**: String, integer or float (the output data type will be the same as the input data type) – the variance for the named column; not applicable for string data
- **Frequency_List**: List of tuples of type (string, integer), where the string is the nominal value and the integer is the associated count; not applicable is returned for numerical data and the list is in descending order of frequency, therefore the first item in the list is the mode
- **StdDev_of_Frequency**: Float – the standard deviation of the frequencies for nominal values; not applicable is returned for numerical data, or data for which there is only one nominal value, for example, if all participants in a survey were male

All functions described are performed directly on the PySpark dataframes to optimise speed, with only the outputs sent to a Pandas dataframe. In Pandas format, row indexing can then be used to extract specific information as required.

Candidate links function

The candidate links function produces a candidate links dataset as a PySpark dataframe. The purpose of the function is twofold: it is a vital step in both deterministic and probabilistic linkage, but also a method of data anonymisation.

Outputs are Boolean such that they can be used in the *m u* probability function for probabilistic matching, given the Boolean outputs directly translate to the binary indicator variable γ^i_j, necessary for use in the Expectation Maximisation algorithm (see Section 3 (https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/developingstandardtoolsfordatalinkagefebruary2021#expectation-maximisation-algorithm)).

The following arguments must be passed to the function.

Sample

Float greater than 0 and less than or equal to 1. The user-defined proportion of records required from each dataframe. Samples are selected by PySpark according to the Bernoulli distribution. Sampling is necessary because the total possible combinations of record pairs for large datasets is beyond the capability even of an advanced distributed computing system.However, if the dataframes for linkage are small then the entire link space can be retained by passing sample equals 1.

Similarity

Float greater than 0 and less than or equal to 1. The acceptable similarity between strings to class them as a match, where 1 indicates the strings must be identical. It is used within the function to calculate a Levenshtein distance threshold according to Equation 9, which also accounts for varying string lengths (see Identifying clones in dynamic websites using similarity thresholds (https://www.scitepress.org/Papers/2004/25973/25973.pdf)).

Equation 9

$$threshold = 2 \times max(length\,(string_1,\ string_2)) \times (1 - similarity)$$

If the Levenshtein distance is less than the threshold, then the agreement status is true. A similarity of one reduces the Levenshtein threshold to zero, which means the strings must be identical to be a match.

df1 and df2

PySpark dataframes intended for linkage. Only two must be passed.

Identifier

String. Name of column that contains unique identifier (optional argument). If the input dataframes have unique identifiers, then the names of these columns can be passed such that they are retained in the output.

The function runs by extracting a sample from each dataframe according to the proportion specified in the sample argument. This accounts for differences in size between the datasets, however, does mean that the total link space will be a function of the size of the smallest dataset. The results are randomised because of the non-indexed nature of PySpark dataframes.

The function then appends a suffix of either _1 or _2 to the column names of each sample, such that they are distinguishable when comparing the values between each dataframe. Comparisons are recorded in a new column before the original columns are dropped. In this way the output ensures data protection.

Fields for which candidate links are desired must have the same column name and data type in both input dataframes. If not, the function assumes these columns are not required and they are dropped. For example, if *df1* contains a column called "Date_of_Birth" and *df2* contains a

column called "DoB" then one name should be selected, and the other column renamed for consistency. The requirement to rename columns and change data types can be efficiently identified using the diagnostics functions. This is not a process that can be automated as column names and data types are context specific.

The output of the candidate links function is a single PySpark dataframe. The column headers depend on the names of the columns of the input dataframes; however, consistent suffixes are applied as follows:

- **identifier_1 and identifier_2**: String – if the optional argument *identifier* is passed, then the first two columns of the output will be the unique identifiers from the original dataframes, with the suffixes "_1" and "_2" respectively
- **Column_bool**: Boolean – the original column name with the suffix "_bool" to indicate whether these columns are an exact match or not
- **Column_soundex_bool**: Boolean – the original column name with the suffix "soundex_bool" to indicate if there is a match on the Soundex phonetic code; Soundex codes are only calculated for strings and are dropped within the function to retain privacy
- **Column_lev**: Integer – the original column name with the suffix "_lev", which represents the Levenshtein distance between two strings, which is widely used in data linkage; this is only calculated for string data types
- **Column_lev_bool**: Boolean – the original column name with the suffix "_lev_bool", which converts the Levenshtein distance to a Boolean, specifying whether the string matches (true) or does not match (false); the threshold Levenshtein distance is controlled by the *similarity* argument when calling the function

Each output row represents a record pair comparison. If either record has a null value, the resulting comparison will also return null to avoid generating false negatives from missing data (it is not safe to assume a false match simply because a value is missing).

m u probability function

The *m u* probability function uses the Expectation Maximisation (EM) algorithm to calculate *m* and *u* probabilities and their associated agreement and disagreement weights, which are essential parameters in probabilistic linkage. This method is complex to implement manually, therefore the function affords significantly improves efficiency for the user. The arguments passed to the function are as follows:

Candidates

PySpark dataframe. It is recommended to use the output from the candidate links function from a formatting point of view, and because the candidate links are generated from random samples so are assumed to be representative of the entire link space. Any columns that are not in Boolean format will be dropped by the function. Any rows containing null values are dropped to avoid false negatives. The user is notified with a warning of any data that have been dropped.

iterations

Integer. The maximum number of iterations to be performed by the EM algorithm.

tol

Float between zero and one. The convergence tolerance required. This is the minimum acceptable difference between the complete data log likelihood at each iteration.

The EM algorithm requires initial parameter estimates for *m*, *u*, and *p* to initiate the expectation step (see Section 3 (https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/developingstandardtoolsfordatalinkagefebruary2021#expectation-maximisation-algorithm)). These are built into the function rather than selected by the user, as

the algorithm is not overly sensitive to initial estimates (see Data Quality and Record Linkage (https://www.researchgate.net/publication/220695391_Data_Quality_and_Record_Linkage)). Jaro 1989 (https://www.tandfonline.com/doi/abs/10.1080/01621459.1989.10478785) recommends that the initial value of $m > u$ and uses $m = 0.9$ therefore this has been chosen going forward. The initial values of u and p are 0.2 and 0.5 respectively as per Analysis of a Probabilistic Record Linkage Technique without Human Review (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1479910/).

The EM algorithm is an unsupervised machine learning algorithm, therefore requires every column to contain some combination of both true and false values so that there is variety for the algorithm to "learn" from. For example, if all comparisons in the Forename column are false then $m = u = 0$, as the probability of having a match given any conditions will always be 0. This will result in an agreement weight of negative infinity and disagreement weight of 0.

Similarly, if all values are true then $m = u = 1$, the corresponding agreement weight would be 0, and disagreement weight would be negative infinity. It is possible that the values for m and u probabilities legitimately approach 0 and 1 as the algorithm runs. Because of limits on precision in any computational system, and subsequent rounding, this can result in 0 or infinite weights as described previously. The m u probability function handles this prior to calculating agreement and disagreement weights by replacing zero with 0.0000000000000001 and one with 0.9999999999999999.

Convergence is based on the complete data log likelihood (equation 7), which is calculated and stored for each iteration such that successive iterations can be compared. The use of complete data log likelihood as a convergence is well recognised (see Probabilistic record linkage with the Fellegi and Sunter framework (https://repository.tudelft.nl/islandora/object/uuid%3A4ebdb50d-c02f-430b-8037-319b45a3e5db), Data Quality and Record Linkage (https://www.researchgate.net/publication/220695391_Data_Quality_and_Record_Linkage), Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida (https://www.tandfonline.com/doi/abs/10.1080/01621459.1989.10478785), and

Maximum Likelihood from Incomplete Data via the EM Algorithm (https://www.ece.iastate.edu/~namrata/EE527_Spring08/Dempster77.pdf) for more information).

The *m u* probability function allows the user to control the precision of the output by providing both an acceptable convergence tolerance and a maximum number of iterations. The function terminates at whichever is reached first.

The output of the function is a single Pandas dataframe providing the following detail for each attribute of the record pair:

- **m**: Float – the *m* probability for the given attribute; the probability the attributes match given they belong to the same unit
- **u**: Float – the u probability for the given attribute; the probability the attributes match given they relate to different units, or the probability of chance agreement
- **W_a**: Float – the agreement weight for the given attribute, $\log_2\left(\frac{m}{u}\right)$
- **W_d**: Float – the disagreement weight for the given attribute, $\log_2\left(\frac{1-m}{1-u}\right)$

There are two main points that should be taken into consideration when using this function: firstly, the assumption of conditional independence in the Fellegi-Sunter method of probabilistic linkage. This is potentially violated given the user is relied upon to select the columns for use in the function. It is necessary for this to be user controlled, however, given content will vary for different datasets and applications.

Secondly, it is vital that the same attribute is not represented more than once in the input candidate links dataset, or convergence will not occur. This is best illustrated by way of an example: if the candidate links dataset contains surname_boolean, surname_soundex_boolean,

and surname_lev_boolean columns, then only one of these columns should be retained as they are all represent the same attribute; in this case surname.

10. Conclusions

Data linkage allows existing data to be re-purposed, which can lead to new discoveries in many fields of research. It is an effective, low-cost approach that can reduce the requirement to collect new data and therefore be less intrusive than traditional survey methods. Robust functions that improve the efficiency and consistency of the process have strong implications for best practise: personnel are no longer required to formulate context-specific code on a case-by-case basis, increasing overall productivity and quality of output. It also makes complicated methods more accessible to non-experts. Additionally, the provision of commented code delivers transparency in approach.

This project successfully achieved the study objective to create functions that automate and improve the preliminary stages of the data linkage pipeline. Seven functions were produced, which generate linkage-related metadata, candidate links and weights for use in the Fellegi-Sunter method of probabilistic record linkage. The latter function involves the practical application of linkage theory widely accepted in the literature (see Data Quality and Record Linkage (https://www.researchgate.net/publication/220695391_Data_Quality_and_Record_Linkage) and Using the EM Algorithm for Weight Computation in the Fellegi-Sunter Model of Record Linkage (https://courses.cs.washington.edu/courses/cse590q/04au/papers/WinklerEM.pdf)) by invoking the Expectation Maximisation (EM) algorithm to estimate *m* and *u* probabilities. This function has potential for expansion into automated machine learning methods of determining matches and non-matches. For instance, the use of clustering techniques to automatically determine match status, rather than manually defining score thresholds.

One limitation of the project functions is that it was not possible to automate certain processes because of a lack of standard formats and naming conventions in different databases. For example, the candidate links function will only execute on columns that share both name and data type. This may necessitate manual re-naming of columns and conversion of data types prior to running the function. Yet this does have the benefit of wider application to a range of

different database types, as there is no dependency on context. Another limitation is that the functions assume data are structured and have been cleaned to a certain extent. However, the diagnostic functions can be used to highlight the need for cleaning and rule-based formatting.

Feedback from users of the functions has been positive in respect of expediting the data linkage process. There has also been wider interest in the application of the diagnostic functions for general exploratory data analysis outside the field of data linkage.

11. Authors

Amy Mayer and Max Stockdale

12. Related links

Government Statistical Service data linking guidance (https://gss.civilservice.gov.uk/policy-store/data-linking/#data-linkage-symposium)
Webpage | Released 16 January 2019
Defines data linking, describes the need for linked data, highlights a review of government data linking methods and summarises how to access the resulting data.

Joined up data in government: the future of data linking methods (https://www.gov.uk/government/publications/joined-up-data-in-government-the-future-of-data-linking-methods)
Webpage | Released 25 August 2020
A cross-government review containing contributed articles on state-of-the-art data linking methods and recommendations for government data linkage.

EXHIBIT 2

UNITED STATES BANKRUPTCY COURT

DISTRICT OF MASSACHUSETTS

In re: TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Debtors.	Chapter 11 Cases 14-40987-MSH 14-40988-MSH 14-40989-MSH Jointly Administered
STEPHEN B. DARR, TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Plaintiff, v. BENJAMIN ARGUETA, A REPRESENTATIVE OF A CLASS OF DEFENDANT NET WINNERS Defendants.	Adversary Proceeding No. 16-4006
STEPHEN B. DARR AS TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Plaintiffs, v. MARCO PUZZARINI, A REPRESENTATIVE OF A CLASS OF DEFENDANT NET WINNERS Defendants.	Adversary Proceeding No. 16-4007

EXPERT REPORT OF DR. CAMERON E. FREER

Respectfully submitted this 29th day of April 2022

Cameron E. Freer, Ph.D.

TABLE OF CONTENTS

I. Overview 1
A. Introduction 1
B. Scope of Engagement 2
C. Qualifications 3
D. Fees and Invoices 4
E. Materials Considered 4
F. Qualifying Language 4
G. Summary of My Opinion 5
II. Introduction to Record Linkage 5
A. Assessing Pairwise Matches 7
1. Data Quality Assessment and Field Selection 7
2. Data Cleaning 8
3. Fuzzy Matching 10
4. Pairwise Match Assessment in Deterministic Record Linkage 11
5. Pairwise Match Assessment in Probabilistic Record Linkage 12
6. Pairwise Match Assessment in Hybrid Record Linkage 18
B. Converting Pairwise Match Assessments into Clusters 19
1. Fixed Thresholding and Transitive Closure 19
2. Dynamic Clustering 21
III. The Telexfree DataSet 22
A. Data Source 22
B. Data Fields 23
1. Special Role of the Name Field in the TelexFree Dataset 24
2. High Quality 26
3. Intermediate Quality 29
4. Low Quality 31
IV. Aggregation Method 37
A. Hybrid Approach 37
B. String Nearness Measures 38
C. Software and Computing Infrastructure Utilized 40
D. Steps in the Aggregation Methodology 43
1. Step 1: Selecting and Transforming the Fields 44

2. Step 2: Cleaning the Fields 45
3. Step 3: Deterministic Step 48
4. Step 4: Blocking 50
5. Step 5: Rubric Application to Calculate Gammas 56
6. Step 6: Calculate Match Probabilities 62
7. Step 7: Compute Clusterings for Several Fixed Thresholds 62
8. Step 8: Dynamic Clustering Process 63
9. Step 9. Convert the Clustering of Account Groupings into a Clustering of User Accounts 72
10. Step 10. Tabulate Total Net Equity for Each Cluster 72

V. Results 73
VI. Discussion of Example Cluster 75
VII. Conclusion 78

EXHIBITS

Exhibit Number	Description
1	Curriculum Vitae of Cameron E. Freer
2	Materials Considered
3	Cluster 132785 Example
4	Gamma Cutoff Selection Example
5	Summary of TelexFree User Accounts and Participants by Tier
6	Aggregation Output, provided natively as *aggregation_output.csv*
7	Summary of the Aggregation Output, provided natively as *aggregation_summary.csv*

I. OVERVIEW

A. Introduction

1. Based upon my discussions with counsel to Stephen Darr, the Trustee of TelexFree, and Huron Consulting Group, LLC ("Huron"), the Trustee's financial advisors, I understand the following general background facts:

2. TelexFree was a Ponzi and pyramid scheme that purportedly sold Voice over Internet Protocol (VoIP) phone plans and operated from 2012 to 2014. Scheme participants ("Participants") were encouraged to create multiple accounts ("User Accounts") to increase their involvement in the scheme. Some Participants created one User Account; others created thousands.

3. An organization known as Ympactus operated a related scheme in Brazil, using the same business model and also maintaining its records in the TelexFree database. In total, there are 17,016,780 User Accounts in the TelexFree database. The Trustee determined that of that 17,016,780 User Accounts, 10,987,617 are attributable to TelexFree, 4,006,422 are attributable to Ympactus, and 138 are a "hybrid". The remaining 2,022,603 User Accounts are those for which no invoice was ever paid.

4. The TelexFree database treats each of the approximately 17 million User Accounts as a standalone record that does not internally link to other User Accounts, even when such User Accounts were opened by the same Participant. While many Participants created multiple User Accounts, some Participants at times entered information that was inconsistent across their various User Accounts. For example, two User Accounts opened by the same Participant may have different email addresses.

5. Because of the pyramid nature of TelexFree's scheme, Participants often recruited friends and relatives into the scheme. My understanding of the TelexFree scheme is that, in the course of User Account creation, it was common practice for recruits to use their own name while using the contact information (email address, phone number, street address, and other fields) of their recruiters, or for family members to share the same contact information, or for a business and an individual to share the same contact information.

6. When the TelexFree scheme collapsed, some Participants received more money than they had paid into the scheme ("Net Winners"), whereas the majority of Participants had contributed more into the scheme than they were paid out ("Net Losers").[1] As previously mentioned, many Participants maintained multiple User Accounts. In order to determine the Net Winners and Net Losers, the Trustee needed to establish a mechanism to aggregate the User Accounts attributable to each Participant.

B. **Scope of Engagement**

7. Borelian Corporation ("Borelian") has been engaged by Stephen Darr, the Trustee of TelexFree, to develop and implement an aggregation methodology for grouping the User Accounts associated with each Participant. Specifically, Borelian was engaged to: (1) develop an aggregation methodology that creates one "cluster" of User Accounts for each Participant such that each User Account appears in precisely one cluster; and (2) provide the set of clusters that results from implementing this methodology on the 10,987,617 User Accounts attributable to TelexFree.

8. My scope of engagement also includes participating in depositions or other court proceedings as necessary.

[1] For purposes of this report, I consider the small number of Participants with zero net gains to be Net Losers.

C. **Qualifications**

9. I am a Principal at Borelian, a data analytics firm with a specific expertise in analyzing, cleaning, and aggregating large commercial data sets. I am also a Research Scientist at Massachusetts Institute of Technology's Probabilistic Computing Project. As part of my work at MIT, I have published a number of papers in the fields of probabilistic inference[2], statistical testing[3], and clustering in graphs[4]. I received a Ph.D. in Mathematics from Harvard University. My CV is attached as Exhibit 1.

10. I have published 28 mathematics, statistics, and computer science research papers; all of these publications, including all publications in the last 10 years, are listed in Exhibit 1. I have been an invited speaker at many international conferences, and have given several dozen invited lectures (see Exhibit 1, pp. 4-7). I have been granted a U.S. Patent (Id. at p. 4), and my research has been featured in the Wall Street Journal, Forbes, the BBC, Wired, The Atlantic, and Fast Company (Id. at p. 11).

11. In addition to my academic research, I have extensive work experience working with complex datasets. I served as Chief Scientist of the real estate data analytics firm Remine, Inc., where I evaluated, cleaned, and merged large datasets of consumer and property records (which came from different vendors and had no canonical unique identifiers), built statistical models based on this data, and helped build their data science team and design their data architecture. I have also worked on data analysis of financial datasets. In total, I have over ten

[2] N. Ackerman, C. Freer, and D. Roy. On the computability of conditional probability, *Journal of the ACM* 66, no. 3, 23:1–23:40, 2019.

[3] F. Saad, N. Ackerman, C. Freer, and V. Mansinghka. A family of exact goodness-of-fit tests for high-dimensional discrete distributions, *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, PMLR 89, 1640–1649, 2019.

[4] D. Cai, N. Ackerman, and C. Freer. Priors on exchangeable directed graphs, *Electronic Journal of Statistics*, 10, no. 2, 3490–3515, 2016.

years of experience with extraction, cleaning, manipulation, and analysis of complex structured data.

12. I have not provided expert testimony in any legal proceeding in the four years prior to the date of this report.

D. **Fees and Invoices**

13. Borelian is compensated for my time and the time of Borelian professionals expended on this matter under my direction at the rate of $300 per hour. My hourly rate is $300 per hour. Compensation to Borelian is not contingent upon the outcome of the related litigation or the opinions I express as a result of my review and analysis.

E. **Materials Considered**

14. I reviewed data and documents from a variety of sources during my work and in the course of generating this report. A complete listing of documents that I considered is presented in Exhibit 2.

F. **Qualifying Language**

15. The opinions presented in this report are based on Borelian's analyses, as carried out under my direction and control, and my experience in analyzing complex datasets over the course of my career. The opinions expressed in this report are subject to change based on additional information that I may receive after the date of this report, and I reserve the right to update my report upon learning of and reviewing additional information that may impact the findings or opinions stated herein. Further, I reserve the right to prepare additional exhibits, charts, graphs, tables, demonstratives, and diagrams to summarize or support the opinions and analyses set forth in this report. Finally, I reserve the right to address those conclusions and opinions raised by the Defendants' Expert which are not addressed herein

G. Summary of My Opinion

16. Based on the work described below, my aggregation methodology has determined that the 10,987,617 User Accounts attributable to TelexFree belong to 1,566,383 different Participants, with a mean size of 7.01 User Accounts per Participant. Of the 1,566,383 Participants, 1,484,702 are Net Losers and 81,681 are Net Winners. The Net Winners have a total net gain of $1,555,981,482.42, with an average net gain of $19,049.49 per Net Winner. The Net Losers have a total net loss of $1,782,171,730.09, with an average net loss of $1,200.36 per Net Loser.

17. As set forth in detail below, I aggregated the User Accounts and tabulated the Total Net Equity for each Participant using a 10-step aggregation methodology that I developed in accordance with standard techniques for data linkage and based on the specific properties of the TelexFree dataset.

II. INTRODUCTION TO RECORD LINKAGE

18. Record linkage is the process of matching records in a dataset based on the similarity of the contents of the records.[5] This is typically done by first assessing the similarity of each pair of records to determine the extent to which the pair of records should be linked. [6] This step is then followed by a step that converts the pairwise match assessments into clusters.

19. There are deterministic, probabilistic, and hybrid approaches to assessing pairwise matches, which I describe in Section II.A below. For any of these approaches, it is crucial to

[5] "Record linkage is the process of matching records from several databases that refer to the same entities. When applied on a single database, this process is known as deduplication." Christen, P. (2011). A survey of indexing techniques for scalable record linkage and deduplication. IEEE Transactions on Knowledge and Data Engineering, 24(9), 1537-1555.

[6] "Traditionally, the de-duplication problem has been solved by making an independent match decision for each candidate pair of records. A similarity score is calculated for each pair…" Parag Singla and Pedro Domingos, Multi-Relational Record Linkage. *KDD-2004 Workshop on MultiRelational Data Mining*, pp. 31-48.

assess the quality of the data in each field, to select fields to be used for linkage based on data quality, to clean the data, and to consider whether to employ fuzzy matching.

20. To convert the pairwise matches into clusters, one can use fixed thresholding and transitive closure, or a dynamic clustering approach, as I describe in Section II.B below.

21. Consider the following example in Table 1, which contains public records from Durham County, North Carolina[7], and which I will discuss throughout this section to illustrate the key concepts of Record Linkage.

Table 1: Example of public NCSBE records retrieved from http://www.ncsbe.gov for the county of Durham in North Carolina and corresponding to unique voter registration numbers. Some fields have been omitted for brevity, including ZIP code, phone number and voter registration number. Street addresses have been permuted with other individuals as to preserve some anonymity.

	Name	Street Address	Age	Sex	Race	Birth	Party
1	Domineck Q. AAshad Jr	914 Monmouth Ave #3	26	M	B	–	LIB
2	Domineck Q. AAshad Sr	1408 Auburndale Dr	55	M	B	NY	DEM
3	Xiomara A. Martinez	1715 Cole Mill Rd	31	F	O	HL	REP
4	Xiomara A. Martinez	2923 Forrestal Dr	31	F	O	HL	–
5	Virginia, L. Mullinix	749 Ninth St #480	101	F	W	PA	REP
6	Jacqueline D. Fuller	141 Bagby LN	54	–	–	–	DEM
7	Jacqueline Fuller	905 Cook Rd	56	F	B	NC	DEM

From an initial inspection, it appears that rows 1 and 2, despite having fairly similar names, likely refer to different people (perhaps father and son). Rows 3 and 4 likely refer to the same person (who possibly moved). Rows 6 and 7 might refer the same person, though it is less clear. No other pairs appear to refer to the same person.

[7] Table 1 of Olivier Binette and Rebecca C. Steorts, (Almost) All of Entity Resolution, arXiv:2008.04443, 2022, augmented by the row numbers 1-7.

A. Assessing Pairwise Matches

1. Data Quality Assessment and Field Selection

22. The goal of data quality assessment is to identify fields that are of sufficiently high quality for reliably assessing pairwise matches. The dimensions relevant to data quality include completeness (how often the field values are present vs. missing), consistency (how consistent the values used for matching are), accessibility (inclusion of different kinds of non-overlapping fields such as name, address, phone, email, etc.), and believability (how credible or plausible the actual values are).[8]

23. Completeness can be quantified in terms of the frequency of missing data in each row. For example, in Table 1, rows 2, 3, 5, and 7 have data present in every field, while rows 1 and 4 are missing data in a single field and row 6 is missing data in three fields. Completeness can also be quantified in terms of the frequency of missing data in each field. For example, the fields Name, Street Address, and Age are present for all records, while each of the fields Sex, Race, Birth, and Party have missing data for 1 or 2 out of 7 records.

24. The consistency of the data in the example varies by field. For example, Age is somewhat consistent among pairs of records that appear to match (as Age is the same for rows 3 and 4, but is not the same for rows 6 and 7). Sex, Race, and Birth are more consistent, since they agree exactly for rows 3 and 4 (and the data is missing for row 6, making it harder to compare to row 7). Party is similarly consistent, as rows 6 and 7 agree (while the data is missing for row 4, making it harder to compare to row 3). The consistency of the Street Address field is less clear (e.g., it could be that rows 6 and 7 refer to the same person who moved between addresses, but there is no direct evidence for this).

[8] Section 3.1 of P. Christen, *Data Matching*, Data-Centric Systems and Applications, Springer-Verlag, 2012.

25. The data in the example is moderately accessible, due to the presence of multiple non-overlapping types of data in the fields Name, Street Address, Age, and Birth. The fields Sex, Race, and Party do provide additional kinds of data, but their usefulness is limited by the relatively small number of possible values for each field (e.g., "M" vs. "F" for Sex). The accessibility would be increased if other kinds of data were included, such as email address or phone number.

26. The data appears believable. One possible exception is the presence of "101" as the Age in row 5; this value is suspicious though not decisively an error (as the age is technically possible).

27. Having assessed the data quality, one then selects which fields to use for linkage. In the case of this example in Table 1, on the basis of the above analysis, all 7 fields seem of high enough quality to use. For a small dataset, as in this example, this assessment can be done visually, but for a large dataset (such as the TelexFree dataset), analogous semi-automated techniques are necessary (such as counting non-empty field values, looking at random samples of rows, etc.).

2. Data Cleaning

28. Data cleaning involves slightly modifying field values of the dataset to improve their linkage power. By appropriately cleaning the dataset, the quality of potential matches can be improved without losing much information. Studies have shown that by cleaning datasets, substantially better linkage results can be achieved.[9]

[9] "There is evidence that improvements in the quality of the underlying data lead to improvements in the quality of the linkage process. For example, early studies of probabilistic linkage in health research demonstrated that greater amounts of personal identifying data greatly improved the accuracy of linkage results … In the absence of strongly identifying personal information, data cleaning has been recognised as one of the key ways to improve the quality of linkage" (p. 2 of Sean M Randall, Anna M Ferrante, James H Boyd, and James B Semmens. The effect of data cleaning on record linkage quality. *BMC Medical Informatics and Decision Making* 2013, 13:64.)

29. One common cleaning technique is to replace multiple consecutive spaces with a single space. For example, “Alice Adams” would be cleaned to “Alice Adams”. This technique is useful, because people do often enter data containing extraneous spaces, but these extra spaces are seldom informative, and can be removed without losing much, if any, information. By removing the extra spaces, more potential matches can be considered.

30. Another common cleaning technique is to convert a field value entirely to lowercase characters (leaving alone the non-alphabetical characters). For example, “Alice Adams” would be cleaned to “alice adams”. This technique is useful, because people are often inconsistent about the capitalization pattern that they use, but such capitalization patterns are seldom informative for linkage. By regularizing the format to all-lowercase, more potential matches can be considered without losing much information.

31. As an example, consider Table 1 augmented by the following new row:

Jacqueline D. Fuller	141 Bagby Ln	54	-	-	-	DEM

This new row is the same as row 6 except that the Name value of the new row has extra spaces in it (between “D.” and “Fuller”), and the Street Address has different capitalization in the two rows (“Ln” in the new row vs. “LN” in row 6). If the cleaning technique used involves replacing multiple consecutive spaces with a single space, then the Name field would exactly agree between row 6 and the new row (as the cleaned value of both would be “Jacqueline D. Fuller”). If the cleaning technique used involves making all characters lowercase, then the Street Address field would exactly agree between row 6 and the new row (as the cleaned value of both would be “141 bagby ln”).

3. Fuzzy Matching

32. Fuzzy matching is a technique often used in assessing how likely it is that field values agree. Fuzzy matching involves computing a measure of nearness between two values, typically with 100 indicating exact agreement and 0 indicating total disagreement. Fuzzy matching is useful in record linkage because it allows for the consideration of pairs of values that are merely similar, rather than identical.[10]

33. For example "Alice Adams" and "Alice Addams" differ in just one character, a difference which might have been caused by a typo. Most nearness measures would return a score for this pair of values that is high, but still below 100.

34. There are a number of nearness measures in the literature, each with its own strengths. For the purpose of illustration, I will use the *ratio* nearness measure, which is one of the three nearness measures used in the aggregation model. (See Section IV.B for additional information about the *ratio* nearness measure.)

35. Without fuzzy matching, values in a dataset may have to be cleaned more thoroughly – and even then, many potential matches might be missed. With fuzzy matching, the entire linkage process is more forgiving of formatting errors (e.g., typos) which allows for more of the value of the original data to be preserved.

36. Consider the example in Table 1. The values for the field Name in row 6 and row 7 are "Jacqueline D. Fuller" and "Jacqueline Fuller", respectively. These values are highly similar, but they are not identical – the *ratio* nearness measure of this pair is 92. In contrast, the *ratio* nearness measure of "Xiomara A. Martinez" and "Virginia, L. Mullinix" (Name values in

[10] "Rather than comparing attribute values between two records using only an exact comparison function (that returns a binary 'same' or 'different' value), it is vital for data matching to employ comparison functions that return some indication of how similar two attribute values are." Section 5.1 of P. Christen, *Data Matching*, Data-Centric Systems and Applications, Springer-Verlag, 2012.

rows 4 and 5) is 45, which is consistent with how dissimilar these Name values are. A record linkage approach that can take the degree of similarity into account is better than one that requires exact agreement, as the latter would miss a potential match between row 6 and row 7.

4. Pairwise Match Assessment in Deterministic Record Linkage

37. Deterministic record linkage involves assessing whether a given pair of records matches on the basis of whether a particular set of fields (or, say, some minimum number of fields) exactly agree on the cleaned values for that pair.[11]

38. A deterministic approach is often appropriate when the dataset is of uniformly high quality in a standardized format, or when there is an accurate, ubiquitous field that serves as a unique identifier (such as Social Security Number).

39. For example, a rule for a deterministic approach to the example in Table 1 could be "match any two records that agree exactly on Name, Street Address, and Sex, treating a field with a missing value as not agreeing with any value in that field". Such a rule would not lead to any pairs of records being matched.

40. As another example, again consider the dataset in Table 1, and suppose the deterministic rule is "match any two records that exactly agree in at least five of the seven fields, treating a field with a missing value as not agreeing with any value in that field". In that case, rows 1 and 2 would not match (since they agree only on Sex and Race), rows 3 and 4 would match (since even though they disagree on Street Address and Party, they agree on the remaining five fields), and rows 6 and 7 would not match (since they agree only on Party).

[11] Dusetzina SB, Tyree S, Meyer AM, Meyer A, Green L, and Carpenter WR. Linking Data for Health Services Research: A Framework and Instructional Guide. *US Agency for Healthcare Research and Quality*, Rockville, MD, 2014.

5. Pairwise Match Assessment in Probabilistic Record Linkage

41. Probabilistic record linkage involves assessing how closely pairs of records match based on how well the fields in the records agree. Unlike deterministic record linkage, this closeness is not all-or-nothing, but rather a value between 0 and 1 (or between 0% and 100%).

42. Although probabilistic record linkage methods can be computationally more complex to implement than deterministic methods, they can perform well even when the dataset involves substantial variation in field values (such as name variants, typos, multiple email addresses or phone numbers, etc.).[12]

43. Probabilistic record linkage involves the use of a "scoring rubric" that defines how much weight to give each field.

44. As an example, consider the following simple rubric:

> Agreement on each field is worth the following amount: Name = 30%, Street Address = 20%, Age = 10%, Sex = 10%, Race = 10%, Birth = 10%, Party = 10%. For the fields Name and Street Address, consider values that have a ratio nearness of 80 or more to be in agreement, while other fields require exact agreement; treat a field with a missing value as not agreeing with any value in that field.[13]

45. Applying this rubric to the dataset from Table 1, we have the following agreement scores: Rows 1 and 2 (which agree on Name, Sex, and Race) have an agreement score of 30% + 10% + 10% = 50%. Rows 3 and 4 (which agree on Name, Age, Sex, Race, Birth) have an agreement score of 30% + 10% + 10% + 10% + 10% = 70%. Rows 6 and 7 (which agree on Name and Party) have an agreement score of 30% + 10% = 40%.

46. This pairwise match assessment step does not itself specify how an agreement score translates into a match or non-match. One possible rule for translating agreement scores

[12] Ong TC, Duca LM, Kahn MG, and Crume TL. A hybrid approach to record linkage using a combination of deterministic and probabilistic methodology. *J Am Med Inform Assoc.* 2020;27(4):505-513.
[13] Note that the maximum possible agreement score is 100%.

into matches is the following: "an agreement score above 70% is a match". Continuing the example above, this rule would deem rows 3 and 4 to be a pairwise match, while neither rows 1 and 2 nor rows 6 and 7 would be deemed pairwise matches. I describe in more detail how an agreement score translates into a match or non-match in Section II.B, "Converting Pairwise Match Assessments into Clusters".

47. While a manually-constructed scoring rubric like this one is illustrative, such rubrics are not a standard approach in modern probabilistic record linkage. Instead, a variant of the Fellegi-Sunter model is typically used to construct the scoring rubric systematically from the dataset.

i. Fellegi-Sunter Model

48. The Fellegi-Sunter model of probabilistic record linkage[14] was described in a seminal 1969 paper by its two authors.[15] Even today, the Fellegi-Sunter model underlies most of the standard approaches to probabilistic record linkage.[16]

49. The Fellegi-Sunter model describes a way to track agreement on multiple fields simultaneously and turn that information into an overall match probability for each pair of records under consideration.

50. A particular extension of the Fellegi-Sunter model, Fellegi-Sunter with Expectation Maximization ("FS-EM"), has become a standard method for performing probabilistic record linkage. Expectation Maximization is a ubiquitous algorithm for iteratively

[14] Fellegi, I. P. and Sunter, A. B. A theory for record linkage. *Journal of the American Statistical Associatio*n, 64(328), 1183-1210 (1969).
[15] There are over 3000 citations of the original Fellegi-Sunter paper.
[16] "... the most commonly used probabilistic model of record linkage (Fellegi and Sunter 1969)." (Enamorado, T., Fifield, B., and Imai, K. (2019). Using a probabilistic model to assist merging of large-scale administrative records. *American Political Science Review*, 113(2), 353-371.)

estimating parameters of a statistical model.[17] In 1988, William Winkler of the US Bureau of the Census described how to combine Expectation Maximization with the Fellegi-Sunter model to more robustly determine the scoring rubric directly from the dataset.[18] This combined approach was then used to perform record linkage on the 1990 US Census[19], and continues to be a standard approach.

51. As noted in Section II.A.5, the Fellegi-Sunter model provides a scoring rubric for transforming patterns of agreement on the fields into agreement scores – one which is informed by the dataset itself.

52. The form of the scoring rubric under Fellegi-Sunter is a set of so-called *m*-probabilities and *u*-probabilities, which characterize, for each field, the frequency of agreement patterns for pairs of records that match, and that do not match, respectively. For example, if pairs that match have agreement on a given field 75% of the time, while pairs that do not match have agreement on that field only 20% of the time, the *m*-probabilities for that field are (0.25, 0.75) and the *u*-probabilities for that field are (0.8, 0.2).[20]

53. The input to the Fellegi-Sunter model is typically specified via the pattern of agreement on the various fields for each pair of records under consideration, and is called a "comparison vector". For example, if a pair of records agrees on the first and third field and

[17] "The Expectation-Maximization (EM) algorithm finds maximum likelihood parameter estimates in problems where some variables were unobserved. Special cases of the algorithm date back several decades, and its use has grown even more since its generality and widespread applicability were discussed by Dempster, Laird, and Rubin (1977)." (Radford M. Neal and Geoffrey E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. *Learning in Graphical Models.* Springer, 1998. 355-368.)

[18] Winkler, W. E. Using the EM Algorithm for Weight Computation in the Fellegi-Sunter Model of Record Linkage, American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 667-671 (1988).

[19] William E. Winkler and Yves Thibaudeau. *An application of the Fellegi-Sunter model of record linkage to the 1990 US decennial census.* US Bureau of the Census, 1991.

[20] This example uses the convention of representing the two possible levels of agreement as (non-agreement, agreement).

disagrees on the second field (and there are only three fields), the comparison vector for the pair will be (1, 0, 1).

54. The individual terms in a comparison vector are often referred to as "gamma" values. For example, if there are two fields, Name and Age, the comparison vector can be written as (gamma_Name, gamma_Age), where gamma_Name and gamma_Age are the respective gamma values for the pair.

55. The output of the Fellegi-Sunter model is the agreement score for each pair of records under consideration, typically formulated as a "match probability". For example, if the Fellegi-Sunter model determines that agreement on just the first and third fields is especially indicative of a match, then a pair of records with comparison vector (1, 0, 1) might have match probability that is relatively high (i.e., close to 100%).

ii. Partial Degrees of Agreement

56. In the discussion above, a comparison vector indicates which fields a given pair of records agrees on and which fields it disagrees on. Agreement need not mean exact agreement (e.g., it could mean a *ratio* value above 80), but still this formulation involves just two possibilities (agreement vs. disagreement).

57. However, it is often possible to distinguish partial degrees of agreement (e.g., via the nearness value itself, which might be any number between 0 and 100). These partial degrees of agreement can also be incorporated into the comparison vector.

58. When incorporating partial degrees of agreement into the comparison vector, the *m*-probabilities and *u*-probabilities must then specify the frequency of each partial degree of agreement among pairs or records that match and pairs of records that do not match, respectively.

59. These partial degrees of agreement are typically discretized into a small number of different levels (or "bins") of agreement. Consider the following example with 3 levels (0, 1, and 2) of agreement for a given field. Level 0 represents substantial disagreement (pair ratio < 80), level 1 represents partial agreement (80 <= pair ratio < 100), and level 2 represents exact agreement (pair ratio = 100), where "pair ratio" is the *ratio* nearness measure of the pair's values on that field. For every pair, its gamma value for this field takes one of the three values 0, 1, or 2.

60. In this example, the *m*-probabilities and *u*-probabilities would each be specified as triples rather than pairs. Suppose that pairs that match have exact agreement on that field 60% of the time, partial agreement 30% of the time, and substantial disagreement 10% of the time, while pairs that do not match have exact agreement on that field 25% of the time, partial agreement 25% of the time, and substantial disagreement 50% of the time. Then the *m*-probabilities for that field are (0.1, 0.3, 0.6) and the *u*-probabilities for that field are (0.25, 0.25, 0.5).[21]

61. Missing values can also be represented naturally using multiple levels of agreement. Typically a gamma value of -1 represents that one or both values of the given field are missing for that pair. This gamma value is treated like any other gamma value in the Fellegi-Sunter machinery, and pairs with this value are assigned partial credit accordingly. This is important, since it allows for datasets that have frequently missing values to receive an appropriate amount of partial credit depending on how that pattern of missingness interacts with the rest of the dataset.

62. For example, consider the records in rows 5, 6, and 7 of Table 1. The Birth field is "PA" for row 5, missing for row 6, and "NC" for row 7. Rows 5 and 7 definitely disagree on Birth value and would likely be assigned gamma value 0. On the other hand, rows 6 and 7 would

[21] This example uses the convention of representing the three levels of agreement in the order (0, 1, 2).

be assigned gamma value -1 because one of the values is missing. The intuition here is that missing values are not definite disagreements, and should perhaps get some partial credit, depending on the details of the dataset.

iii. Blocking

63. In the descriptions above of probabilistic record linkage, I have spoken of how such an approach assesses matches for a set of pairs of records under consideration. In the case of fairly small datasets, this set of pairs might well consist of all pairs of records. For example, in Table 1, there are 7 x (7 - 1) / 2 = 21 pairs of records (ignoring the order and pairs consisting of the same record twice).

64. However, when the dataset is large, it may be computationally unreasonable to run the FS-EM algorithm on the set of all pairs directly. For example, in the TelexFree dataset, which has approximately 11 million User Accounts, there would be approximately 60 *trillion* pairs to consider. In such a circumstance, it is often advantageous to use a technique known as "blocking" to reduce the number of pairs under consideration to a much smaller number.

65. The output of the blocking step is a "blocking set", which contains those pairs of records that will remain under consideration in subsequent steps.

66. The restriction to a smaller subset of pairs is justified only if every pair that is excluded from the blocking set[22] would have been very unlikely to match, had it been considered. For example, if a pair exactly agrees on every field, it is important to keep it under consideration (by including it in the blocking set), because it is likely to have very high match probability. On the other hand, if a pair substantially disagrees on every single field, then it is

[22] "Record pairs are discarded unless they agree on the blocking key." (J.S. Murray, "Probabilistic Record Linkage and Deduplication after Indexing, Blocking, and Filtering". Journal of Privacy and Confidentiality (2015–2016) 7:1, pp. 3–24.)

likely to have very low match probability, and it is permissible to exclude such a pair from consideration (by excluding it from the blocking set).

67. The blocking step can involve one or more rules, any of which suffice for entry into the blocking set. For example, suppose the blocking rules are:

- pair ratio on the Name field > 80
- pair ratio on the Street Address > 80

Under these blocking rules, the blocking set for Table 1 would consist of (row 1, row 2), (row 3, row 4), and (row 6, row 7). All of these are due to the Name rule, and none due to the Street Address rule. As a result, after blocking there are merely 3 pairs (rather than all 21 pairs) for further consideration by the FS-EM algorithm.

68. In this example, each of the two blocking rules involved just a single field. A blocking rule may involve more than one field.[23] For example, a blocking rule might state: "pair ratio on the Name field > 80 and the pair's values of Age exactly agree".

69. Importantly, the presence of a pair of records in the blocking set does *not* signal that the pair is necessarily likely to be a match; rather, it simply means that the pair meets the minimum requirement to be further considered.

6. Pairwise Match Assessment in Hybrid Record Linkage

70. Probabilistic record linkage is computationally complex, and benefits substantially from a smaller input dataset. When dealing with large datasets, it may be advantageous to use a hybrid record linkage approach – specifically, performing a deterministic record linkage step followed by a probabilistic record linkage step. The deterministic record linkage step would match those pairs of records that are unquestionably the same (effectively

[23] "Blocking keys may be constructed by conjunctions of multiple keys (e.g., agreement on last initial and postal code)." (J.S. Murray, 2016)

establishing a 100% agreement score), while leaving a smaller set of pairs for the probabilistic record linkage to consider.

71. A hybrid approach can be quite helpful computationally, as it can substantially reduce the number of pairs that need to be considered by the subsequent probabilistic machinery. A hybrid record linkage approach is standard, and can often lead to better matching results than either deterministic or probabilistic record linkage techniques on their own.[24]

72. Consider the example dataset in Table 1 with the following new row added:

Jacqueline Fuller	905 Cook Rd	56	F	B	NC	DEM

Observe that the new row exactly agrees on every field with row 7. If the rule for the deterministic step (as part of a hybrid approach) is "match any two records that exactly agree in all fields, treating a field with a missing value as not agreeing with any value in that field", then row 7 will be matched with the new row (but no other pairs of rows), and so the subsequent probabilistic step need only consider 7 (groupings of) rows, rather than all 8 rows. Later, once clusters have been formed on this reduced dataset, these clusters can be expanded into the corresponding clusters on the full dataset.

B. Converting Pairwise Match Assessments into Clusters

1. Fixed Thresholding and Transitive Closure

73. After the pairwise match assessment is complete, the remaining task is to convert the agreement scores (e.g., match probabilities in the case of FS-EM) into a clustering (a set of

[24] Ong TC, Duca LM, Kahn MG, and Crume TL. A hybrid approach to record linkage using a combination of deterministic and probabilistic methodology. *J Am Med Inform Assoc.* 2020;27(4):505-513. See also Kabudula, C.W., Clark, B.D., Gómez-Olivé, F.X., Tollman, S., Menken, J. and Reniers, G. The promise of record linkage for assessing the uptake of health services in resource constrained settings: a pilot study from South Africa. *BMC Med Res Methodol* 2014; 14, 71.

clusters). This is typically achieved via two steps: fixed thresholding, followed by transitive closure.[25]

74. Fixed thresholding consists of linking all pairs whose agreement score exceeds a particular fixed threshold.

75. Transitive closure is a standard technique for creating clusters (from a set of linked pairs of records) such that every record is in exactly one cluster. The transitive closure of a set of linked pairs is formed by repeating the following process until no additional links need to be added: if record A is linked to record B, and record B is linked to record C, then a link is added between record A and record C (regardless of the agreement score between records A and C).

76. Any particular fixed threshold involves a tradeoff between two competing tensions: overclustering (when clusters include many pairs with low agreement scores) and underclustering (when clusters fail to include many pairs with high agreement scores).

77. As an example, consider Table 1 with the following new row added:

Jacqueline D. Fuller	141 Bagby LN	54	F	B	-	DEM

Suppose the deterministic rule being used is "match any two records that agree in at least four of the seven fields, where agreement in the Name or Street Address fields requires a *ratio* of at least 80%, and where other fields must agree exactly, treating a field with a missing value as not agreeing with any value in that field". This rule does not match rows 6 and 7 directly, though it does match row 6 to the new row and row 7 to the new row. Taking the transitive closure of

[25] "[T]he pairs whose similarity score is above some pre-determined threshold are merged. This is followed by taking a transitive closure over matching pairs." (Parag Singla and Pedro Domingos, Multi-Relational Record Linkage. KDD-2004 Workshop on MultiRelational Data Mining, pp. 31-48.)

these matches means that rows 6 and 7 do end up matched, by virtue of the linkages provided via the new row.

2. Dynamic Clustering

78. Some datasets exhibit heterogeneity in the sense that some portions of the dataset are "tightly bound" with little variation (e.g., only minor typos) and very high agreement scores among their pairs, while other portions are "loosely bound" with more substantial variation (e.g., name variants, multiple addresses, small age differences, change in party affiliation) and lower agreement scores among their pairs.

79. For heterogeneous datasets, it is often the case that performing fixed thresholding and transitive closure, for any single choice of fixed threshold, is suboptimal, as any such choice will lead to substantially overclustering one portion of the dataset while simultaneously substantially underclustering another portion.

80. The idea of dynamic clustering (also known as heterogeneous clustering) is to better navigate this tradeoff between overclustering and underclustering by choosing higher thresholds in the tightly bound portion of the dataset while choosing lower thresholds in the loosely bound portion of the dataset.

81. As an example, consider again the dataset in Table 1. If one chooses a high fixed threshold (i.e., being highly selective about which pairs of rows to group together), then rows 3 and 4 will likely be linked into a cluster (because they are so similar), but rows 6 and 7 will likely not be linked into a cluster (because they are substantially less similar). If one instead chooses a low fixed threshold (i.e. being less selective about which pairs of rows to group together), then rows 3 and 4 will still likely be linked together, and now also rows 6 and 7 will likely be. However, in that low fixed-threshold situation, there is also a risk that rows 1 and 2 will be linked together as well (even though they probably should not be). Neither the high nor

the low threshold is ideal; it would be better if one could choose a high threshold for rows 3 and 4 (linking them into a cluster), a lower threshold for rows 6 and 7 (linking them into a cluster), and a high enough threshold for rows 1 and 2 so that these rows are not linked into a cluster. This is the essence of the dynamic clustering approach.

III. THE TELEXFREE DATASET

A. Data Source

82. Borelian was provided with four tables that I understand are the full and complete exports for the original TelexFree database. These tables are: *representante* (the "Account Table"), *fatura* (the "Invoice Table"), *transferencia* (the "Transfer Table"), and *bonus* (the "Bonus Table"). (The names of the tables, as well as of the fields in these tables, appear to be in Portuguese.)

83. For purposes of input into the aggregation methodology that is the focus of this report, only information from the Account Table was used. The Invoice Table, Transfer Table, and Bonus Table all contain information about transfers between various Participants or between TelexFree and Participants, rather than information about the Participants themselves, and these tables are not used in the aggregation methodology. My understanding is that Huron used the information in these three tables to calculate Net Equity for each User Account.

84. The Account Table includes 17,016,780 User Account records and 44 fields. Based on information provided to Borelian from Huron in the form of a *rep_id_summary* table (the "Summary Table")[26], I filtered (via the *entity* field) the User Account records considered for aggregation down to the 10,987,617 User Accounts attributable to TelexFree and associated with at least one paid invoice. The Summary Table also contains (in the *net_equity* field) the Net

[26] The Account Table, Invoice Table, Transfer Table, Bonus Table, and Summary Table were sent by Huron to Borelian as MySQL table dumps.

Equity for each User Account as calculated by Huron, which I used in Step 10 to tabulate net equity after clustering (see Section IV.D.10 below).

B. **Data Fields**

85. The Account Table (representante table) includes 44 fields, whose contents I now describe. The 44 fields are: *rep_id, rep_login, rep_id_direto, rep_id_pai, lado, rep_id_disk, ord_id_disk, rep_data_cad, rep_datanasc, rep_sexo, rep_nome, rep_pwd, rep_pwd_secondary, rep_razao_social, rep_cnpj, rep_cpf, rep_rg, rep_ssn, rep_email, rep_end, rep_comp, rep_numero, rep_bairro, rep_cidade, cod_rep_cidade, rep_cep, rep_uf, cod_rep_uf, rep_pais, cod_rep_pais, rep_fone, rep_cel, rep_obs, token, token_secondary, chargeback, blocked, transaction, rep_ticket_id, unsubs_mark, unsubs_trans, valida_email, rep_primeiro_nome, rep_sobrenome*. I relied on both the "Reference" tab of spreadsheet "10ksample.xlsx", provided by Huron to Borelian, and my own subsequent analysis in order to ascertain the meaning of these fields.

86. Most database tables contain a unique identifier field that lets a user of the database identify a particular row. In the Account Table, this role is played by the *rep_id* field. Accordingly, it contains unique values and is never null or empty. The aggregation methodology uses it to uniquely identify User Accounts but it is not itself considered for linkage determinations.

87. I next describe each of the other 43 fields, and discuss the quality of each. Of those 43 fields, 11 fields are of sufficient quality to warrant inclusion in the aggregation methodology. These fields are: *rep_nome, rep_email, rep_cel, rep_fone, rep_end, rep_numero, rep_cep, rep_login, rep_pwd_secondary, rep_cpf,* and *rep_datanasc*.

88. There are several relevant dimensions to data quality (completeness, consistency, accessibility, and believability), as I have discussed in Section II.A.1. The completeness of the TelexFree dataset varies by field, but is high for the 8 high-quality fields described below in Sections III.B.1 and III.B.2 (namely, *rep_nome*, *rep_email*, *rep_cel*, *rep_fone*, *rep_end*, *rep_numero*, *rep_cep*, and *rep_login*). Consistency is also generally high for these 8 fields, albeit with occasional variation. The dataset has high accessibility, due to the presence of many different kinds of non-overlapping fields, representing name, email, address, phone number, etc. The dataset is also generally rather believable, despite a small portion of exceptions (due to garbage like "11111" or values like "Mickey Mouse"). For a visual illustration of the general quality of the dataset, see the cluster shown in Exhibit 3.

1. Special Role of the Name Field in the TelexFree Dataset

89. In a typical user-account dataset, email addresses and phone numbers are unique to an individual. However, this is often not the case in the TelexFree dataset. As described in Paragraph 5, a major source of complexity with the TelexFree dataset is the prevalence of User Accounts that contain different names, but the same email address, phone number, address, etc.[27]

90. As a result, the name field (*rep_nome*) in the TelexFree dataset gains a special meaning that it might not have in another dataset. In particular, it is the *only true differentiator* of Participants – if User Accounts A and B match in all other fields (email address, cell phone number, etc.) but A's name field is quite different from B's name field (say, "Alice Adams" and

[27] For multiple examples of distinct Participants (with different names) having the same contact information as each other in several fields, see the following Exhibits of the July 31, 2020 *Rebuttal Expert Report of Joshua W. Dennis*: Exhibits 9.1 and 9.5 (individual and business), Exhibits 9.2 and 9.3 (same email or address), Exhibit 9.4 (different businesses with the same address and email), Exhibits 9.6 and 9.11 (same last name and phone, but different first names), Exhibit 10 (same email, address, and last name, but verified by affidavits to be spouses – see July 29, 2020 affidavits of Frantz Balan and Lusette Balan).

"Bob Baker"), then these User Accounts should be treated as belonging to different Participants and should not be aggregated.

91. Only when the names in the TelexFree dataset are near enough (in the sense of string nearness; see Section II.A.3), say "Alice Adams" and "Alice Addams", and there are other markers of agreement (such as a matching email address, cell phone number, etc.), should these two User Accounts be treated as belonging to the same Participant and should be aggregated.[28] Accordingly, in the aggregation methodology, we require that any two User Accounts under consideration for possible linkage have at least fairly similar name (see Section IV.D.4).

92. It is worth noting that while it may turn out that a name is made up (such as "Mickey Mouse"), this itself is not a critical issue for data aggregation. The relevant issue for aggregation is whether the values in the field support the matches, not whether or not the values in the field reflect the true name. Even when a Participant uses a pseudonym that is totally unrelated to their real name, it is possible to successfully aggregate their User Accounts if (a) they use the pseudonym consistently, and (b) not too many other Participants use that same pseudonym without including different information in other fields. In the best case scenario, the true identity of the Participant can, at a later time, be ascertained from the other indicia; in the worst case scenario, the Participant will never be identified, but their records do not contaminate the rest of the aggregation.

93. Indeed, the name field (*rep_nome*) in the TelexFree dataset is of very high quality (despite variations in case, punctuation, spaces, and other typical issues that data cleaning and fuzzy matching can account for). First, the name field is only empty in 0.002% of User

[28] What counts as "near enough" is a parameter of the model, and needs to be appropriately balanced: too restrictive (such as requiring exact name matching), and one misses some potential linkages; too expansive, and one risks incorrectly aggregating the User Accounts of different Participants into a single cluster.

Accounts and is never null. Second, nearly all of the name values contain strings consisting entirely of alphabetical characters or the symbols commonly found in names (period, comma, hyphen, space, or apostrophe) – as opposed to numbers or other symbols – which is an indication that the value is a legitimate name. Specifically, 97% of the name field values are both five characters or longer and contain only the characters described above. Furthermore, a manual review reveals that in a random sample of 300 name field values meeting this criterion, 98% look like a name. Here are the first 10 values taken from that random sample:

```
jack zheng
alcebiades de souza neto souza
olger hancco condori
JESUS PEREZ
jose correia da silva
Francisca Orlanda de Lima Ferreira
long fang zhang
lizardo juarez apaza
Gustavo Maroso Gessi
alex malo
```

94. Based on these facts, I estimate that in at least 95% of the User Accounts, the value of the *rep_nome* field appears to be a name. Furthermore, *rep_nome* values in the remaining User Accounts are sufficiently diverse in value (i.e., no one value predominates, nor does a small set of values) that they are unlikely to substantially contaminate the rest of the matches – and even less so after cleaning (see Section IV.D.2). As a result, I use *rep_nome* in the aggregation methodology.

2. High Quality

95. I understand that the field *rep_email* is meant to contain the email address of the User Account. Indeed, the field values appear to be email addresses. First, the email field is only empty in 0.84% of User Accounts and is never null. Second, 98.2% of User Accounts contain

rep_email values that passed a validation test for determining valid email addresses: after light cleaning (e.g., removing spaces), I tested the values against the standard format for email addresses as defined in an Internet specification concerning email address, RFC 2822.[29] Third, 71.9% of User Accounts had an email address (after light cleaning) at the domain "gmail.com" or "hotmail.com", which is also indicative of these *rep_email* values being e-mail addresses. The values of *rep_email* were found to be the same for many pairs of User Accounts with similar or identical name (*rep_nome*) values, which is indicative of high linkage value for these two fields. As a result, I use *rep_email* in the aggregation methodology.

96. I understand that the fields *rep_cel* and *rep_fone* are meant to contain the "mobile phone" number and "telephone" number, respectively. Indeed, these fields appear to contain phone numbers. First, *rep_cel* and *rep_fone* are null or empty for only 0.001% and 0.03% of User Accounts, respectively. Second, the values in these fields correspond to what would be expected in a phone number field. Specifically, they nearly always contain numeric data or characters typically associated with phone numbers (space, left and right parentheses, hyphen/dash, and plus). Third, in 95.5% and 95.6% of User Accounts, respectively, *rep_cel* and *rep_fone* values match the validation criterion of being 6 digits or more of numeric or these special characters. Furthermore, the values of *rep_cel* and/or *rep_fone* were found to be the same for many pairs of User Accounts with similar or identical name (*rep_nome*) values, which is indicative of high linkage value for these two fields. As a result, I use *rep_cel* and *rep_fone* in the aggregation methodology.

97. It is my understanding that the fields *rep_end*, *rep_numero*, and *rep_cep* are meant to contain various address components. Indeed, *rep_end* appears to contain a street name

[29] RFCs are specifications for Internet services, and RFCs 2821 and 2822 concern email addresses. For RFC 2822 specifically, see "RFC 2822", retrieved from https://datatracker.ietf.org/doc/html/rfc2822

(sometimes including a house number as well), *rep_numero* appears to contain a house number, and *rep_cep* appears to contain a postal code (CEP stands for "Código de Endereçamento Postal" in Brazil, and is analogous to ZIP Code in the United States)[30]. The fields *rep_end*, *rep_numero*, and *rep_cep* are null or empty in only 0.005%, 22%, and 0.001% of User Accounts, respectively. Manual inspection of representative samples of these three fields confirms that they contain what would be expected of the respective address components (when present). These fields, taken together, provide the core of the address, and appear to have substantial linkage value, and so I use *rep_end*, *rep_numero*, and *rep_cep* in the aggregation methodology.[31]

98. I understand that the field *rep_login* contains the login name chosen by the user. This field is never null or empty and contains unique values. While this is the only user-entered (as opposed to system-generated, like *rep_id*) field that the TelexFree system required to be unique (as is typical of login names), it appears that many Participants who created multiple User Accounts nevertheless attempted to choose *rep_login* values that were similar across their User Accounts (e.g., "ttmilagres", "ttmilagres3", and "ttmilagres6" all appear as *rep_login* for accounts with *rep_nome* "Maria Teresa Milagres Neves"; see page 1 of Exhibit 3). This similarity in value gives the *rep_login* field substantial linkage value despite the fact that the values are literally different for every record. As a result, I use *rep_login* in the aggregation methodology, subject to a truncation transformation described below in Section IV.D.1.

[30] "Código de Endereçamento Postal", retrieved from https://en.wikipedia.org/wiki/C%C3%B3digo_de_Endere%C3%A7amento_Postal

[31] The high linkage value of these components is supported by the fact that at least in the United States, credit card processors verify card transactions by using house number and postal code (the analogues of *rep_numero* and *rep_cep*). See "Address verification service", retrieved from https://en.wikipedia.org/wiki/Address_verification_service

3. Intermediate Quality

99. I understand that the field *rep_pwd_secondary* contains a transformation of the user's password (specifically, its md5sum hash) that allows a database to store the password more securely than storing the actual password itself. This is further supported by the fact that md5sum hashes are always 32-character hexadecimal strings, and so are the values of the *rep_pwd_secondary* field (after mild cleaning). md5sum is a one-way function, which means that it is easy to compute but hard to reverse. For example, if the user's password is "123456", the md5sum function converts that password to "e10adc3949ba59abbe56e057f20f883e". It is easy to convert a password into its md5sum hash, but it is computationally difficult in general to figure out the password from a particular md5sum hash. Two User Accounts having identical passwords will have identical md5sum hashes. However, two User Accounts having similar but not identical passwords will have very different md5sum hashes. For example, the md5sum hash of "123457" is "f1887d3f9e6ee7a32fe5e76f4ab80d63". As a result, comparisons on the value of the *rep_pwd_secondary* field can only identify passwords that are identical.

100. The *rep_pwd_secondary* field is null for 76.95% of User Accounts. However, when present, this field is nevertheless useful, because (1) most of the values are not particularly frequent, (2) the data requires essentially no cleaning when present, and (3) it does appear that there are many instances where multiple User Accounts that have identical or similar name values (*rep_nome*) also have the same password. In other words, it looks like at least some Participants re-used the exact same password for multiple different User Accounts. While it is possible for different parties to choose the same password, this happens relatively infrequently. In total, this field has high linkage value when present, and so I use *rep_pwd_secondary* in the aggregation methodology.

101. I understand that the field *rep_cpf* is meant to contain a national ID value analogous to a social security number ("CPF" stands for Cadastro de Pessoas Físicas, Portuguese for "Natural Persons Register").[32] On the face of it, *rep_cpf* would therefore seem to represent information that could, on its own, uniquely identify an individual. However, upon close examination of its values, *rep_cpf* appears to be of merely intermediate quality.

102. The field *rep_cpf* is null or empty for just 0.001% of User Accounts, and is therefore ubiquitous in this dataset. Manual inspection of representative samples of this field confirms that its values are generally consistent with being a national ID value. Furthermore, a number of User Accounts that appear to be connected to the same Participant do often share the same value of *rep_cpf*. However, even though national identifiers have the potential to be unique to a Participant, the contents of the *rep_cpf* field do not actually appear to be unique to a Participant. For many sets of records having identical or near-identical values of *rep_nome* (e.g., *rep_cpf* values of [redacted], [redacted], [redacted]", and "[redacted], among others, all appear in User Accounts with *rep_nome* value "Maria Neves"), the value for *rep_cpf* is not consistent. Nor is it the case that a given *rep_cpf* value occurs only for records having similar *rep_nome* values (e.g., *rep_cpf* value [redacted] appears in User Accounts with *rep_nome* values of "ting bu", "yuanhua liu", and "hao zhang", among others). It therefore does not meet the standard for a unique identifier, that of being both ubiquitous and unique (to a Participant). Nevertheless, because of the linkage value that it adds, I include *rep_cpf* in the aggregation methodology.

103. I understand that the field *rep_datanasc* is meant to contain the date of birth ("data de nascimento" is Portuguese for "date of birth"). The *rep_datanasc* field is null for

[32] "CPF number", retrieved from https://en.wikipedia.org/wiki/CPF_number

78.2% of User Accounts. However, when present, this field is nevertheless useful, because (1) most of the values are not particularly frequent (though, naturally, some Participants in a dataset of this size will share birth dates), (2) the data is always a proper date and requires no cleaning when present, and (3) it does appear that there are many instances where multiple User Accounts that have identical or similar name (*rep_nome*) values also have the same *rep_datanasc* value. I therefore include *rep_datanasc* in the aggregation methodology.

4. Low Quality

104. The remaining fields in the TelexFree dataset were assessed and determined to have low quality at best. Most of these fields are of low quality because they (1) contain information not related to Participants' identity, (2) contain information whose linkage value is substantially subsumed by higher quality fields, or (3) are mostly empty. However, I will first describe the fields that did not fall neatly into one of these three categories.

105. The field *rep_data_cad*, which I understand to mean "date account created," is never null or empty, and always contains a date in 2012, 2013, or 2014. It provides little linkage value because it typically stands in a many-to-many relationship with the Participants, meaning that many Participants apparently created accounts on a number of different dates, and on most days, many accounts from apparently different Participants were created. As a result, *rep_data_cad* is not included in the aggregation methodology.

106. The field *valida_email*, which I understand is meant to contain information about the "validation [of] the email address," is never null or empty, and it always contains the value "0," "1," or "2". Examination of the *valida_email* values in conjunction with *rep_email* values reveals that any of the three possible values occurs in conjunction with many valid-looking values of *rep_email*. Furthermore, all three possible values often occur with the same address (e.g., "telexfreeben@gmail.com" is the value of *rep_email* for 253 accounts in which

valida_email is "0," for 960 accounts in which *valida_email* is "1," and for 735 accounts in which *valida_email* is "2"). Hence, *valida_email* does not describe a property of the email address itself, and is therefore not used in the aggregation methodology.

107. The fields *rep_primeiro_nome* and *rep_sobrenome*, which I understand represent some notion of "first name" and "last name", are empty for 0.009% and 0.97% of User Accounts, respectively, and are never null. Examination of their contents in detail, in tandem with the *rep_id* value, reveals that there were at least 3 distinct regimes for how these two fields relate to *rep_nome*. In the first regime, *rep_primeiro_nome* and *rep_sobrenome* contain the first and last tokens of *rep_nome*. (A token is a sequence of non-space characters separated by a space; in "Quick brown fox", the tokens are "Quick", "brown", and "fox".) In the second regime, *rep_primeiro_nome* and *rep_sobrenome* were mostly empty. In the third regime, *rep_primeiro_nome* contained the first token of *rep_nome* and *rep_sobrenome* contained the remaining tokens. To avoid these inconsistencies, I extracted the first and last token directly from the *rep_nome* field when using it as part of the evaluating matches between names; for more details, see Sections IV.D.4 and IV.D.5. Therefore, *rep_primeiro_nome* and *rep_sobrenome* are effectively redundant and are not used in the aggregation methodology.

108. When two fields contain data that is highly correlated, it is common practice in probabilistic record linkage to combine their information into a single variable.[33] In the case where fields are highly correlated and the contents of one field is largely subsumed by that of the other field (as typically occurs with fields describing smaller vs. larger geographic regions such as ZIP Code vs. state), the combined variable would provide little additional linkage value over the more specific of the two fields. As a result, the aggregation methodology includes the more

[33] Supplement S4.1 of Enamorado, T., Fifield, B., and Imai, K. (2019). Using a probabilistic model to assist merging of large-scale administrative records. *American Political Science Review*, 113(2), 353-371.

specific field *rep_cep* but not the fields *rep_bairro, rep_cidade, rep_uf,* or *rep_pais.* The linkage value of the following fields are substantially subsumed by that of the higher quality fields that I do include:

- The field *rep_bairro*, which I understand is meant to contain the user's address component consisting of the district or neighborhood within a city, is null or empty for 37.44% of User Accounts. The value of this field often appears to be such a district or neighborhood (e.g., "centro" or "pudong"), although it sometimes appears to instead be a city name itself (e.g., "Funchal" or "Lisboa") or occasionally a larger region (e.g., "Madeira") or garbage (e.g., "," or ".."). The information *rep_barrio* provides towards potential linkage is substantially subsumed by the *rep_cep* (ZIP Code analogue) field, which is of high quality. As a result, *rep_barrio* is not included in the aggregation methodology.
- The field *rep_cidade,* which I understand is meant to contain the city component of the user's address, is never null or empty. The values in this field appear to predominantly include a city name (e.g., "Phnom Penh", "framingham"), but there are occasional garbage values (e.g., "." or "1"). The information it provides towards potential linkage is substantially subsumed by the *rep_cep* (ZIP Code analogue) field, which is of high quality and more specific. As a result, *rep_cidade* is not included in the aggregation methodology.
- The field *rep_uf,* which I understand is meant to contain the state component of the user's address, is null for 2.00% of User Accounts. It appears to predominantly be a short alphabetical string (e.g., "MA", "LIM") or short numerical string (e.g., "30", "11"). The information it provides towards potential

linkage is substantially subsumed by the *rep_cep* (ZIP Code analogue) field, which is of high quality and much more specific. As a result, *rep_uf* is not included in the aggregation methodology.

- The field *rep_pais*, which I understand is meant to contain a 2-letter code for the country component of the user's address, is never null or empty. It provides little linkage value because of the small number of typical values (e.g., over 55% of User Accounts have one of the top 5 codes), and also because the information it provides is substantially subsumed by other address components such as *rep_cep* (ZIP Code analogue) that are used. As a result, *rep_pais* is not included in the aggregation methodology.
- The field *transaction* is null on 0.11% of User Accounts, and agrees in its value with the field rep_login on 99.85% of User Account records. As such, its linkage value is almost entirely subsumed by that of *rep_login* (which is used for linkage). As a result, *transaction* is not included in the aggregation methodology.

109. The following fields contain information not related to a Participant's identity and do not provide much, if any, linkage value:

- The field *rep_id_pai* is null for 81.56% of User Accounts. The field *rep_id_direto* is always non-null. I understand that the meaning of these fields is related to a User Account's "placement in [the] pyramid", rather than any information about the Participant. Therefore *rep_id_pai* and *rep_id_direto* are not included in the aggregation methodology.
- The field *rep_lado* is null or empty for 81.57% of User Accounts. I understand that the meaning of this field is related to a User Account's "placement on [the]

left side or right side of [the] pyramid", rather than any information about the Participant. Therefore, *rep_lado* is not included in the aggregation methodology.

- The fields *ord_id_disk* and *rep_id_disk* are each null for 6.35% of User Accounts. I understand that the meaning of these fields contains a "phone plan related" datum of the account (that is to say, related to the VoIP product TelexFree sold), rather than any information about the Participant. Therefore, *ord_id_disk* and *rep_id_disk* are not included in the aggregation methodology.
- The field *rep_pwd* is never null or empty, is always 32 characters, and appears to be a hexadecimal hash (e.g., as produced by md5sum). I understand that the meaning of this field is "system generated password". As such, *rep_pwd* does not seem to provide information about the Participant. Therefore, *rep_pwd* is not included in the aggregation methodology.

110. Finally, the following fields in the dataset are mostly zero, empty, or null for each record and do not provide much, if any, linkage value:

- The field *rep_comp* is null in 97.33% of User Accounts. In the remaining 2.67% of records, it appears to contain an address component of some sort, although the nature of the address component varies across the data, e.g., apartment number ("Ste 56"), city name ("beijing"), street name ("calle principal"), neighborhood ("barrio el carmen 6155"), or garbage ("..."). Due to a high presence of null values, and the inconsistent meaning of values, this field is not included in the aggregation methodology.

- The field *rep_ticket_id* is empty on 97.90% of User Accounts. Additionally, I understand that it is one of several "Administrative fields", and this field is therefore not included in the aggregation methodology.
- The field *token* is null or empty for 99.378% of User Accounts, and this field is therefore not included in the aggregation methodology.
- The field *rep_obs* is null or empty for 99.900% of User Accounts, and this field is therefore not included in the aggregation methodology.
- The field *token_secondary* is null or empty for 99.944% of User Accounts, and this field is therefore not included in the aggregation methodology.
- The field *chargeback* has value "0" on 99.954% of User Accounts, and this field is therefore not included in the aggregation methodology.
- The field *blocked* has value "0" on 99.991% of User Accounts, and this field is therefore not included in the aggregation methodology.
- The field *unsubs_trans* has value "0" on 99.997% of User Accounts, and this field is therefore not included in the aggregation methodology.
- The field *unsubs_mark* has value "0" on 99.997% of User Accounts, and this field is therefore not included in the aggregation methodology.
- The field *rep_cnpj* is null or empty for 99.998% of User Accounts, and this field is therefore not included in the aggregation methodology.
- The field *rep_rg* is null or empty for 99.999% of User Accounts, and this field is therefore not included in the aggregation methodology.
- The field *rep_razao_social* is null or empty for 99.999% of User Accounts, and this field is therefore not included in the aggregation methodology.

- The field *rep_sexo* is null for 99.999% of User Accounts, and this field is therefore not included in the aggregation methodology.
- The field *rep_ssn* consists entirely of null or empty values, and this field is therefore not included in the aggregation methodology.
- The fields *cod_rep_cidade*, *cod_rep_uf*, and *cod_rep_pais* consist entirely of null values, and are therefore not included in the aggregation methodology.

IV. AGGREGATION METHOD

A. Hybrid Approach

111. The aggregation methodology developed by Borelian, under my direction, is mainly probabilistic in its approach because it is necessary to have a probabilistic step in the aggregation methodology to account for near-misses in data field matching, to assign partial credit for various fields based on these near-misses, and to account for all of this information when forming clusters.[34] However, a deterministic step that precedes the probabilistic step is helpful for several reasons. First, there is a huge number of pairs of User Account records in the TelexFree dataset that match exactly (after cleaning) on all 8 of the highest quality fields; those pairs can be declared a match without resorting to any further probabilistic analysis.[35] In fact, the dataset is of sufficiently high quality that these exact matches alone constitute a substantial portion of the eventual output. Second, reducing the size of the input is useful for effective implementation of a probabilistic method: adding the deterministic step reduces the number of

[34] "Whereas deterministic linkage is most suitable when unique identifiers are available and the quality of the data are high, probabilistic linkage yields better results when unique identifiers are lacking or in situations where there is variation in reporting or transcription of personal identifiers…" (Kabudula, C.W., Clark, B.D., Gómez-Olivé, F.X., Tollman, S., Menken, J. and Reniers, G. The promise of record linkage for assessing the uptake of health services in resource constrained settings: a pilot study from South Africa. *BMC Med Res Methodol* 2014; 14, 71.)

[35] "Typical Assumptions Made … If two records/mentions are identical, then they are true matches" (L. Getoor and A. Machanavajjhala, Entity Resolution: Tutorial, *Advances in Social Network Analysis and Mining*, 2012.

records from approximately 11 million User Accounts to approximately 4 million groups of User Accounts, which results in a large reduction in the number of pairs under consideration for probabilistic analysis. Therefore, the aggregation methodology consists of a hybrid approach, in which a deterministic step is followed by a probabilistic step. Such a combination is a standard approach, often leading to better matching results.[36]

B. **String Nearness Measures**

112. String nearness match ("fuzzy matching") plays an important role in probabilistic record linkage.[37] There is a wide range of techniques for string nearness matching. Two Python packages that implement a number of these techniques are FuzzyWuzzy and RapidFuzz.[38] (Python is a programming language commonly used in data analysis; many packages exist on top of Python to extend its functionality into specific domains.) I used RapidFuzz to perform string nearness matching between cleaned values of the fields I used in the aggregation methodology (*rep_nome, rep_email*, etc.).

113. Specifically, for blocking in Step 4 and in the rubric application in Step 5, I use three different string nearness methods from RapidFuzz — *ratio*, *WRatio*, and *jwsim* — each with its own strengths.[39] All three functions are used with the *default_process* processor and produce values between 0 (non-match) and 100 (perfect match).

[36] Ong TC, Duca LM, Kahn MG, and Crume TL. A hybrid approach to record linkage using a combination of deterministic and probabilistic methodology. *J Am Med Inform Assoc.* 2020;27(4):505-513. See also Kabudula et al. (2014) above.

[37] Scott L DuVall, Richard A Kerber, and Alun Thomas. Extending the Fellegi-Sunter probabilistic record linkage method for approximate field comparators, *J Biomed Inform.* 2010; 43(1):24-30.

[38] FuzzyWuzzy is a Python package developed and maintained by SeatGeek, the largest event ticket search engine (see https://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-in-python/ for more details). RapidFuzz is another popular Python package that implements FuzzyWuzzy's algorithms in native C++ for faster performance (see https://github.com/maxbachmann/RapidFuzz for more details).

[39] "Such comparison functions need to be tailored to the type of data that are being compared." Section 5.1 of P. Christen, *Data Matching*, Data-Centric Systems and Applications, Springer-Verlag, 2012.

- *ratio* is a standard string nearness measure that is part of the *difflib* Python package; *difflib* itself is a built-in Python library, which speaks to its ubiquity. *ratio* is used by both RapidFuzz and FuzzyWuzzy. The *ratio* nearness measure involves "find[ing] the longest contiguous matching subsequence that contains no "junk" elements; these "junk" elements are ones that are uninteresting in some sense, such as blank lines or whitespace … This does not yield minimal edit sequences, but does tend to yield matches that "look right" to people."[40] *ratio* is good at avoiding typos, treats various parts of the string the same (unlike *jwsim*), and is good at grouping together edits that occur in a similar part of the string. As a result, I use it for nearness matching in the cleaned versions of *rep_nome*, *rep_email*, *rep_cel*, *rep_fone*, *rep_cep*, *rep_numero*, *rep_login*, and *rep_cpf* fields, as well as the derived quantities of "first and last token of name" and "the part of the email address before the @ sign" that are defined using *firstLast* and *beforeAt* functions in Step 5 of the aggregation methodology.
- *WRatio* is an "attempt to weight (the name stands for 'Weighted Ratio') results from different algorithms to calculate the 'best' score".[41] It uses *ratio* as well as variants of *ratio* that are good at finding whole tokens in any order. (Again, a token is a sequence of non-space characters separated by a space; in "Quick brown fox", the tokens are "Quick", "brown", and "fox".) As a result, I use it for nearness matching on the cleaned version of the *rep_nome* field, which has multiple names in various orders, sometimes with typos (e.g., "maria teresa

[40] https://docs.python.org/3/library/difflib.html
[41] https://stackoverflow.com/questions/31806695/when-to-use-which-fuzz-function-to-compare-2-strings

milagres neves", "maria teresa neves", "marai teresa neves", maria milagres", and "maria t m neves" may all refer to the same name).

- *jwsim* is an implementation of the Jaro-Winkler similarity measure that is a staple of probabilistic record linkage.[42] Jaro-Winkler similarity gives a higher score to strings that match in the earlier part of the string (unlike other measures that do not consider the location of the edits). As a result, I use it on the cleaned version of the *rep_end* field, which contains addresses whose endings often vary more than their beginnings (e.g., "13 Maple St. Apt 17", "13 Maple", and "13 Maple Street" may all refer to the same address).

C. Software and Computing Infrastructure Utilized

114. Many software tools have been created that implement the algorithms, methods, and processes necessary for the probabilistic record linkage that I have described above. Of these, I make key use of Splink, a Python package for probabilistic record linkage that is maintained and published by the UK Ministry of Justice and peer-reviewed by the UK Office for National Statistics[43]. The Splink package implements the Fellegi-Sunter model (described above in Section II.A.5.i) of probabilistic record linkage, including the EM algorithm (also described above in II.A.5.i) to estimate parameters of the model (collectively, "FS-EM").

115. Most FS-EM implementations are not up to the task of handling very large datasets (like the TelexFree dataset): "although there exists a well-known statistics literature on probabilistic record linkage … the current open-source implementation does not scale to large data sets commonly used in today's social science research."[44] The fastLink package for R

[42] Winkler, W. E. String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. *Proceedings of the Section on Survey Research.* American Statistical Association, pp. 354–359, 1990.
[43] https://github.com/moj-analytical-services/splink
[44] Enamorado, T., Fifield, B., and Imai, K. (2019). Using a probabilistic model to assist merging of large-scale

(another programming language) solves this problem by providing "a fast and scalable implementation of the canonical probabilistic record linkage model originally proposed by Fellegi and Sunter"[45] and adding parallelization. The Splink package implements the same statistical model as fastLink, but in Python: "The [Splink] package closely follows the approach described in fastLink. In particular it implements the same mathematical model and likelihood functions described in the fastLink paper … with a comprehensive suite of tests to ensure correctness of the implementation".[46]

116. The science behind Splink and fastLink has been extensively tested and peer-reviewed. The statistical model is described in a paper authored by researchers from Harvard and Princeton and published in a peer-reviewed journal[47]; that paper has been cited over 100 times in just the past 3 years. The continued existence and active public maintenance, by a number of contributors, of the Splink and fastLink packages speaks to the standards controlling their operation and to the wide acceptance in the scientific community of the underlying science.

administrative records. *American Political Science Review*, 113(2), 353-371.

[45] Id.

[46] UK Office of National Statistics. (2021). Splink: MoJ's open source library for probabilistic record linkage at scale. Retrieved from https://www.gov.uk/government/publications/joined-up-data-in-government-the-future-of-data-linking-methods/splink-mojs-open-source-library-for-probabilistic-record-linkage-at-scale

[47] Enamorado, T., Fifield, B., and Imai, K. (2019). Using a probabilistic model to assist merging of large-scale administrative records. *American Political Science Review*, 113(2), 353-371.

Additionally, fastLink is commonly used in academic record linkage[48] [49] and Splink is used at scale in the UK Ministry of Justice.[50]

117. I also use the Pandas[51] package for Python, a ubiquitous data analysis framework[52] that is convenient for transforming and cleaning selected fields, performing deterministic record linkage, assisting in performing the blocking step, computing gamma values for the comparison vector, and in the dynamic clustering step. I use Pandas in combination with Splink to perform the probabilistic record linkage step and fixed-threshold clustering.

118. I used string nearness measures provided by the RapidFuzz package, as described in Section IV.B. These are used in the blocking step and in computing gamma values for the comparison vector.

119. I used a massively parallel cloud computing infrastructure (on Microsoft Azure) with hundreds of CPU cores and terabytes of RAM to perform the steps of the aggregation methodology.

[48] "To shed light on this debate, we apply a probabilistic record linkage model, implemented via the open-source software package fastLink, to merge two major election studies—the American National Election Studies and the Cooperative Congressional Election Survey—with a national voter file of over 180 million records." (T. Enamorado and K. Imai. Validating Self-Reported Turnout by Linking Public Opinion Surveys with Administrative Records. *Public Opinion Quarterly*, Volume 83, Issue 4, Winter 2019, pp. 723–748.

[49] "The highly computationally efficient open-source probabilistic record linkage package fastLink (v. 0.5.1) (7) was adapted to utilize the name-matching classifiers described here during record linkage. The package follows an extended Fellegi-Sunter framework (24) for probabilistic record linkage." (P. A. Collender et al. Machine-learning classifiers for logographic name matching in public health applications: approaches for incorporating phonetic, visual, and keystroke similarity in large-scale probabilistic record linkage. *arXiv:2001.01895*, 2020.)

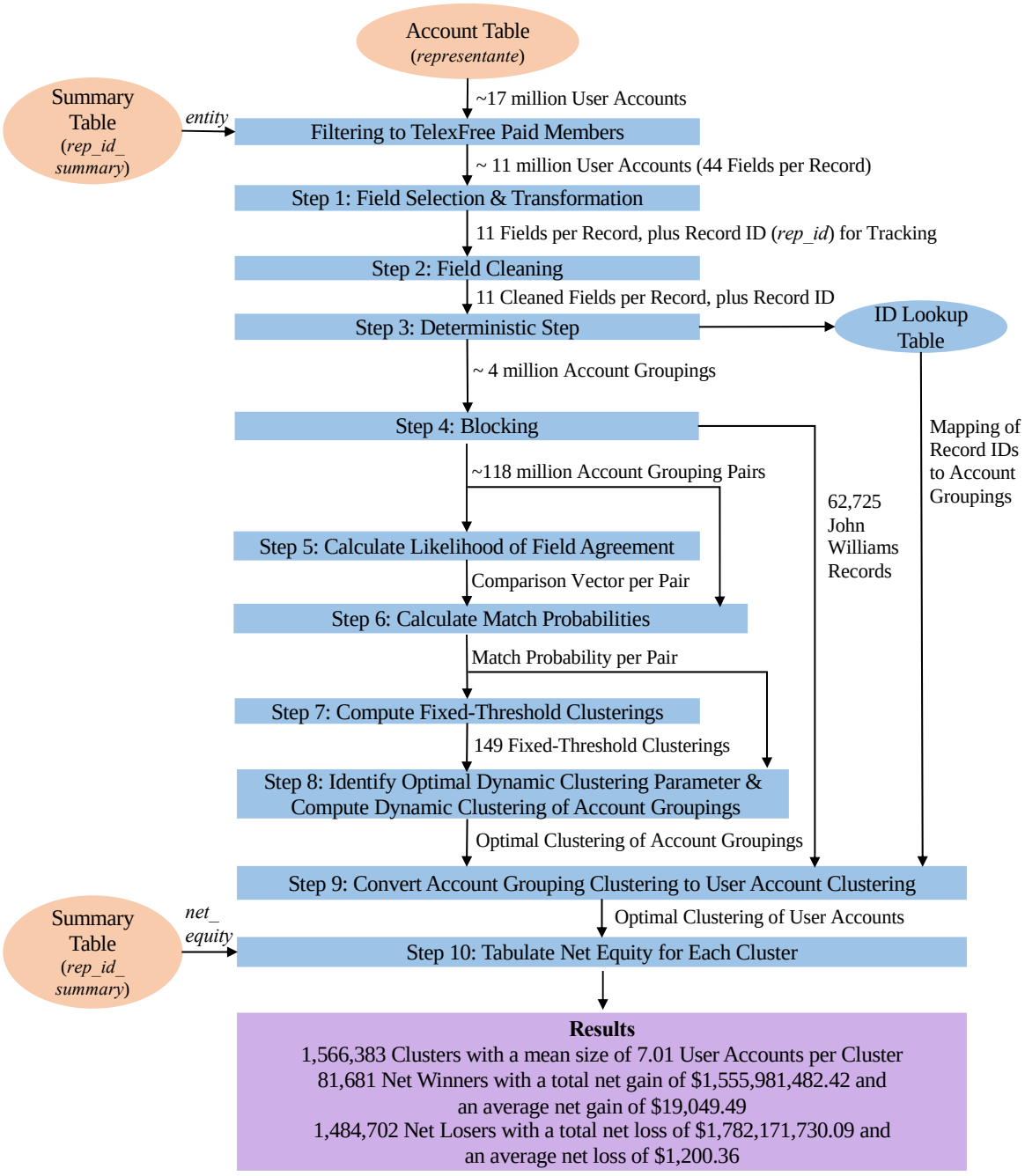
[50] Tom Jackson, Caris Greyson, Ian Rickard and Andromachi Tseloni. Data First: Criminal Courts Linked Data. An exploratory analysis of returning defendants and the potential of linked criminal courts data from 2011 to 2019 in England and Wales. Ministry of Justice Analytical Series, UK Ministry of Justice, 2022.

[51] https://pandas.pydata.org/

[52] Wes McKinney, *Python for Data Analysis*, 2nd Edition, 2017.

D. **Steps in the Aggregation Methodology**

120. The aggregation methodology consists of 10 steps. The following flowchart illustrates the movement of data through the aggregation methodology, from inputs to output.



1. Step 1: Selecting and Transforming the Fields

121. Based on the analysis of data field quality in Section III.B above, I selected the following fields in the *representante* table (Account Table) to use in the aggregation methodology: *rep_nome* (name), *rep_email* (email), *rep_cel* (cell phone), *rep_fone* (phone), *rep_end* (street name), *rep_numero* (house number), *rep_cep* (ZIP code analogue), *rep_pwd_secondary* (hashed password), *rep_cpf* (SSN analogue), *rep_datanasc* (date of birth), and a truncated version of *rep_login* (login name), as I now describe.

122. The *rep_login* field is of high quality, as described in Section III.B.2. However, the values of the *rep_login* field are unique to each User Account by design, as every User Account needed to have a different login. However, using the field without transforming its values would render the deterministic step moot, as no User Account agrees exactly with any other on the *rep_login* field. I address this issue by using a truncated version of the *rep_login* values. (Truncation here means using the first N characters or the last N characters, for an appropriate choice of number N.) A truncated version of the *rep_login* field provides high linkage value: an examination of the values of the *rep_login* field in conjunction with *rep_nome* suggests that many Participants used a variation on the same login pattern across multiple accounts. Specifically, it appears that Participants often used a fixed alphabetical prefix followed by a numeric number suffix (e.g., “ttmilagres3”, “ttmilagres6”, “ttmilagres9”). In many cases, these numeric suffixes were even sequential (e.g., “callandsave2”, “callandsave3”, “callandsave4”). In other cases, an incrementing numeric prefix preceded the alphabetical suffix (e.g., “1000despiste”, “1001despiste”, “1002despiste”), but this was less common.

123. A natural transformation for the *rep_login* field was to truncate it to preserve the pattern Participants tended to follow, that of using the same alphabetical pattern followed by a different number for each of their various login names. By using truncated values, in the case of

a Participant using an alphabetical prefix scheme for *rep_login*, many of their account records will match exactly (and others by fuzzy matching techniques). For the less common case of a numerical prefix with an alphabetical suffix, this truncation scheme may lead to fewer exact matches (post-truncation), but even then, the truncated *rep_login* values may be matched by fuzzy matching techniques.

124. Based on examination of the average length of the alphabetical and numeric components of *rep_login* values and a manual review of a representative sample of the field's values, a prefix length of 4 was determined to be the optimal choice for truncation. Therefore, the first four characters of *rep_login* values for each User Account are used in the aggregation methodology.

125. In addition to these 11 fields, the field *rep_id* is used to identify each User Account record (as described above in Section III.B), but it is not itself used for any linkage determinations.

2. Step 2: Cleaning the Fields

126. Once the fields were selected for use in the aggregation methodology, I applied standard cleaning techniques to the values of each field. The field *rep_nome* was transformed to ***clean***_*rep_nome*, the field *rep_email* to ***clean***_*rep_email*, etc. As previously described, the goal of data cleaning is to ensure that the data in each selected field across all records is consistent and uniform so as to maximize the linkage value between the values of the fields. The data cleaning techniques used for each field varied, but were generally designed to apply industry-accepted transformations[53], such as transforming to lowercase and eliminating spaces and

[53] Randall, S.M., Ferrante, A.M., Boyd, J.H., and Semmens J.B. The effect of data cleaning on record linkage quality. *BMC Med Inform Decis Mak* 13, 64 (2013).

garbage characters, while retaining the informational content of the field to the maximum extent possible.

127. For all 11 selected fields, I first performed a basic cleaning step: leading and trailing spaces were removed, all characters were converted to their lowercase equivalent, and multiple consecutive internal spaces within the field were converted to a single space. Depending on the field, additional cleaning steps were also performed as described below. Finally, strings for which this process resulted in an empty string were replaced with null.

128. For *rep_nome* (name), the additional cleaning steps consist of: (a) replacing, by the empty string, any values of length 1; (b) replacing, by the empty string, any values of length 3 where the second character is a space and the first and third characters are standard Latin alphabetical (a-z) or digit (0-9) characters. For example, "a a" or "1 1" are both replaced by the empty string. This pattern occurs in just 1.37% of the User Accounts, but replacing it with the empty string avoids a substantial increase in blocking size.

129. For *rep_email* (email), all spaces were removed. (Participants sometimes entered values for *rep_email* that appear to be valid email addresses except for extra spaces.)

130. For *rep_cel* (cell phone) and *rep_fone* (phone), all non-digit (0-9) characters were removed, and then the value was truncated to the last 7 characters (e.g., "+1 (202) 867-5309" would be cleaned to "8675309"). The non-digit characters were removed because phone numbers are numeric in nature. Only the last 7 digits[54] were used in order to make it easier for phone numbers to agree (whether or not they include full country code or area code). For

[54] The parameter of 7 digits was chosen because it is a reasonable lower bound on global phone number length. Specifically, North American phone numbers are typically 10 digits long or 7 if the area code is not specified (see "North American Numbering Plan", retrieved from https://en.wikipedia.org/wiki/North_American_Numbering_Plan), and Brazilian phone numbers are typically 10 or 11 digits long, or 8 or 9 without area code (see "Telephone numbers in Brazil", retrieved from https://en.wikipedia.org/wiki/Telephone_numbers_in_Brazil).

example, "2028675309" and "8675309" are very likely to refer to the same phone number, even as there is a small possibility of multiple phone numbers being collapsed to the same one by this cleaning step.

131. The values of *rep_end* (street name), *rep_numero* (house number), *rep_cep* (ZIP code analogue) were not cleaned beyond the basic cleaning step described above.

132. The truncated values of *rep_login* (login name), were not cleaned beyond the basic cleaning step.

133. The values of *rep_pwd_secondary* (hashed password) were not further cleaned, because they contain the automatically-generated output of a hash function[55].

134. The values of *rep_cpf* (SSN analogue) were further cleaned by removing all characters other than the standard Latin alphabetical (a-z) and digit (0-9) characters, followed by another round of the basic cleaning step (to collapse multiple internal spaces).[56] While this may remove some structure (e.g., hyphens or spaces between components of the SSN analogue), this was outweighed by the additional linkages generated as a result of this cleaning.

135. The values of *rep_datanasc* (date of birth) were not cleaned beyond the basic cleaning step described above, because these values already follow a standardized format for date (e.g., "1985-08-21") when present.

[55] For 19 of the 10,987,617 values of *rep_pwd_secondary*, there was a "-T" suffix after the md5sum hash, making the value 34 characters long. It is possible that these values were manually modified by a database administrator to disable the login of those users. In any case, these were the only records for which the value of *rep_pwd_secondary* did not follow the standard 32-character hexadecimal md5sum hash convention. As part of the cleaning, I removed the "-T" suffix in these 19 cases.

[56] This is consistent with standard schemes for national ID numbers, which tend to use Latin and digit characters, even in countries where non-Latin alphabets are in common use (see "National identification number", retrieved from https://en.wikipedia.org/wiki/National_identification_number).

3. Step 3: Deterministic Step

136. The aggregation methodology performs a deterministic step by aggregating User Accounts into account groupings. In order for User Accounts to be placed into the same account grouping, exact agreement is required on these 8 fields: *clean_rep_nome*, *clean_rep_email*, *clean_rep_cel*, *clean_rep_fone*, *clean_rep_end*, *clean_rep_numero*, *clean_rep_cep*, and *clean_rep_login*. The (non-null) values of the remaining 3 fields, *clean_rep_pwd_secondary*, *clean_rep_cpf*, and *clean_rep_datanasc*, are combined into sets of values in the account grouping.

137. For example, consider a dataset containing the following two User Accounts:

Field	**User Account 1**	**User Account 2**
clean_rep_nome	maria teresa neves	maria teresa neves
clean_rep_email	ttmilagres@gmail.com	ttmilagres@gmail.com
clean_rep_cel	9043810	9043810
clean_rep_fone	9043810	9043810
clean_rep_end	2b brackett rd	2b brackett rd
clean_rep_numero	2b	2b
clean_rep_cep	01702	01703
clean_rep_login	call	call
clean_rep_pwd_secondary	c99fb769d8bcee3287936bf71335976e	c99fb769d8bcee3287936bf71335976e
clean_rep_cpf	[redacted]	[redacted]
clean_rep_datanasc	1953-10-15	1953-10-15

In the deterministic step, they will become the following account grouping:

Field	**Account Grouping 1**
clean_rep_nome	maria teresa neves
clean_rep_email	ttmilagres@gmail.com

clean_rep_cel	9043810
clean_rep_fone	9043810
clean_rep_end	2b brackett rd
clean_rep_numero	2b
clean_rep_cep	01702
clean_rep_login	call
clean_rep_pwd_secondary_set	[c99fb769d8bcee3287936bf71335976e]
clean_rep_cpf_set	[950890438, 012901903]
clean_rep_datanasc_set	[1953-10-15]

138. On the other hand, a pair of User Accounts that differ on any of the first 8 cleaned fields are placed in different account groupings.

139. When considering agreement on the first 8 cleaned fields for a pair of User Accounts, a pair of values for a given field is deemed to exactly agree when they have the same non-null value. If either value is null, the User Accounts are deemed to disagree on that field.

140. This deterministic step has minimal effect on the output of the aggregation methodology, because any User Accounts that are part of the same account grouping were already extremely likely to be matched during the probabilistic step. However, the deterministic step is extremely useful in practice, as it makes it computationally reasonable to perform probabilistic linkage methods on a dataset as large as the TelexFree dataset.

141. Each account grouping is assigned a new unique id, called *ag_rep_id*, to keep track of the newly grouped User Accounts. The *ag_rep_id* value is defined to be the lowest *rep_id* value occurring among its User Accounts.

142. The deterministic step results in 4,367,728 account groupings. Each account grouping is a grouping of one or more User Accounts. For analysis purposes, each account

grouping is treated as equivalent to a separate record in the blocking and probabilistic linkage steps. (Account groupings will be expanded back into their constituent User Accounts in Step 9.)

143. In addition to forming the set of account groupings as described above, I also generate a lookup table mapping each User Account (identified by the *rep_id* field) to its account grouping (identified by the *ag_rep_id* field). This lookup table will be used in Step 9 (when these account groupings will be "ungrouped" back into User Accounts).

4. Step 4: Blocking

144. As previously described (in Section II.A.5.iii), when a dataset is large, it may be computationally unreasonable to run the FS-EM algorithm on the set of all pairs of records directly. This is the case in the TelexFree dataset: because it has approximately 4 million account groupings, there are approximately 10 *trillion* pairs of account groupings. Due to this large number, it is advantageous to use a technique known as "blocking" to reduce the number of pairs of account groupings under consideration to a much smaller number.

145. The restriction to a smaller subset of pairs is justified only if every pair that is excluded would have been very unlikely to match, had it been considered. In other words, if a pair is at all likely to match, it is important to keep it under consideration. On the other hand, the presence of a pair of account groupings in the blocking set does not signal that the pair is necessarily likely to be a match; rather, it simply means that the pair meets the minimum requirement to be further considered.

146. The blocking step[57] involves specifying a blocking scheme (described in detail below), which is the criterion that each pair of account groupings must pass in order to remain in

[57] I use the word "blocking" to refer to this entire step; some literature would describe this step as technically involving a combination of blocking and "indexing by disjunctions". In fact, "[t]he terminology is not standardized

the blocking set. Once the blocking step is complete, the blocking set is passed along to the next step in the aggregation methodology.

147. Since any pair not in the blocking set will (implicitly) be assigned a match probability 0, blocking should generally be inclusive, balanced by the need to avoid generating a computationally unreasonably large blocking set. The size of a blocking set is the size of the union of its components, so if any single component results in an unreasonably large set, then the entire blocking set will be too large.

148. Because of the special role that the name field plays in the TelexFree dataset (see Section III.B.1 above), it is important that no pairs of account groupings that substantially disagree on the name end up matched together. I use blocking as a mechanism to accomplish this goal, by ensuring that the blocking set contains only pairs of account groupings with substantial agreement on name. In this way, only such pairs (i.e. those with substantial agreement on name) are even considered for potential linkage.

149. The way I accomplish this is for every blocking rule to be conjunctively joined with a rule ensuring name agreement, as I will describe. On the other hand, there is no need to have a blocking rule that involves only the name field. If two account groupings agree only on name, and substantially disagree on all other fields, then the aggregation procedure is likely to deem them a non-match. As a result, I do not have a blocking rule that contains only the name field.

in the literature." (footnote 1 on p. 4 of J.S. Murray, 2016, "Probabilistic Record Linkage and Deduplication after Indexing, Blocking, and Filtering". Journal of Privacy and Confidentiality (2015–2016) 7:1, pp. 3–24.)

150. I choose all 11 cleaned fields from the previous step of the aggregation methodology to be used in the blocking scheme, with the exception of *clean_rep_nome* (for reasons just described), and *clean_rep_numero*.[58]

151. A very natural blocking scheme is the disjunction of one rule for each of the fields chosen to be part of the blocking scheme, where this rule captures all pairs of records that are sufficiently near on the given field. For the TelexFree dataset, the first such rule is:

- *ratio*(accountGrouping1.*clean_rep_email*, accountGrouping2.*clean_rep_email*) >= 75

for a pair consisting of (accountGrouping1, accountGrouping2), where *ratio* is the nearness measure described in Section IV.B.

152. There are additional rules of the same form, for each of the other fields that are part of the blocking scheme:

- *ratio*(accountGrouping1.*clean_rep_cel*, accountGrouping2.*clean_rep_cel*) >= 75
- *ratio*(accountGrouping1.*clean_rep_fone*, accountGrouping2.*clean_rep_fone*) >= 75
- *jwsim*(accountGrouping1.*clean_rep_end*, accountGrouping2.*clean_rep_end*) >= 75
- *ratio*(accountGrouping1.*clean_rep_cep*, accountGrouping2.*clean_rep_cep*) >= 75
- *ratio*(accountGrouping1.*clean_rep_login*, accountGrouping2.*clean_rep_login*) >= 75
- accountGrouping1.*clean_rep_pwd_secondary_set* has nonempty intersection with accountGrouping2.*clean_rep_pwd_secondary_set*
- *ratio*(*clean_rep_cpf1*, *clean_rep_cpf2*) >=88

[58] The field *clean_rep_numero* is not included in the blocking scheme because (1) the *clean_rep_numero* field has many values with high frequency (e.g., "1", "2", etc.) that make the blocking too large if the field were used along with *clean_rep_nome* but no third field, and (2) a pair that matched on only *clean_rep_nome* and *clean_rep_numero* would be of sufficiently poor quality that it need not be included in the blocking set, whereas a pair matching on *clean_rep_nome, clean_rep_numero,* and any third field would already be included in the blocking set.

for some *clean_rep_cpf1* in *accountGrouping1.clean_rep_pwd_secondary_set* and some *clean_rep_cpf2* in *accountGrouping2.clean_rep_pwd_secondary_set*

- accountGrouping1.*clean_rep_datanasc_set* has nonempty intersection with accountGrouping2.*clean_rep_datanasc_set*

where *ratio* and *jwsim* are the nearness measures described in Section IV.B.

153. Each of these blocking rules is conjunctively joined with the following rule ensuring name agreement:

WRatio(accountGrouping1.*clean_rep_nome*, accountGrouping2.*clean_rep_nome*) >= 85.5
AND

(*ratio*(accountGrouping1.*clean_rep_nome*, accountGrouping2.*clean_rep_nome*) >= 70

OR

ratio(*firstLast*(accountGrouping1.*clean_rep_nome*), *firstLast(*accountGrouping2.*clean_rep_nome*)) >= 80)

where *WRatio* and *ratio* are the nearness measures described in Section IV.B and where *firstLast* is a function that returns the first and last token of a string. (For mononyms, the entire string is returned.) For example:

- *firstLast*("Maria Teresa Milagres Neves") = "Maria Neves"
- *firstLast*("Maria Milagres Neves") = "Maria Neves"
- *firstLast*("Maria Neves") = "Maria Neves"

154. This name rule captures the various ways that a pair of names can substantially agree. In particular, the *WRatio* clause allows for tokens in the name to be dropped or added so long as there is substantial agreement on the rest. The *firstLast* clause allows for tokens to be

dropped or added in the middle of the name so long as the first and last tokens agree. For example,

- *WRatio*("Maria Teresa Milagres Neves", "Teresa Milagres Neves") = 95
- *ratio*(*firstLast*("Maria Teresa Milagres Neves"), *firstLast*("Maria Neves")) = 100

155. The choice of string nearness measure for each of the fields above is made for the reasons described in Section IV.B.

156. All of these numerical cutoffs (e.g., 75 for *clean_rep_email*) were chosen carefully. In order to select these cutoffs, I examined random slices of pairs of account groupings from the dataset, stratified by the value of the relevant nearness measure for that field. For each cutoff, one must balance between (a) potentially excluding pairs that are worth considering in the probabilistic record linkage process and (b) potentially letting in so many pairs that the task becomes computationally unreasonable. I chose cutoffs that erred on the side of inclusion in the blocking set, thereby giving as many pairs as possible the opportunity to participate in the probabilistic record linkage process. This choice is justified because these cutoffs are effectively insensitive in the downward direction, so long as the blocking set size remains computationally reasonable; for example, a *clean_rep_end* cutoff of 70 instead of 75 has little effect on the results – the probabilistic record linkage process would simply be considering a few more pairs, the overwhelming majority of which would be deemed non-matches in the end.

157. This blocking scheme, if applied to all 4,367,728 account groupings output from the deterministic step, would result in approximately 2 billion pairs of account groupings in the blocking set. While not necessarily infeasible to generate, this would be quite difficult to handle in the remaining steps. Furthermore, this large number is overwhelmingly caused by an easily identified set of accounts, which I now describe.

158. There are 62,725 User Accounts that include a *rep_nome* value of "John Williams". They all have *rep_email* (email) with a "mailinator.com" domain, and all have *rep_end* (street name) of the form "534 Warry Lane" with differing numbers in the place of "534". Furthermore, they are the only accounts with the latter two properties (email address ending in "mailinator.com" and street name ending in "Warry Lane") at all, and by a semi-manual inspection of records having a fuzzy match on name, it is clear that no other User Account is remotely similar enough to any of these 62,725 User Accounts to be matched with any of them according to the rest of the aggregation methodology. This is more than enough information to justify manually combining these 62,725 User Accounts into a single cluster that contains no other accounts.

159. It is worth noting that because these records differ from each other on the *full* email address and house address, they were not grouped together in the deterministic step (in fact, each of these User Accounts formed an account grouping with just a single User Account). If the account groupings with name "John Williams" were not excluded from blocking, every pair of records with both name values "John Williams", which make up a large portion of the total potential 2 billion pairs output from the blocking scheme, would make it into the blocking set. Hence, for computational reasons, the pairs attributable to John Williams are excluded from the blocking step and subsequent aggregation steps and are all assigned to be in the same cluster as each other (with no other User Accounts) at Step 9 after the final clustering is computed.

160. After excluding the John Williams pairs, the output of the blocking step is a blocking set consisting of 117,563,883 pairs of account groupings.

5. Step 5: Rubric Application to Calculate Gammas

161. This step considers every pair of account groupings in the blocking set, and assigns each such pair a comparison vector, containing "gamma" values for the cleaned fields.[59]

162. The gamma value of -1 is reserved for cases where data is missing in at least one account grouping of the pair; I do not describe this case explicitly in the discussion or code below.

163. Typically, larger values of gamma denote more agreement between a pair and 0 denotes the least agreement between a pair. In general, the gamma value indicates the degree of agreement, discretized (see Paragraph 59) to a small number of values representing, for example, "substantial disagreement", "substantial agreement", or "exact agreement". In most cases, the gamma value is determined by applying cutoffs to the result of an appropriate string nearness measure. I now describe the way the gamma value is generated for each field.

164. The gamma values for the *clean_rep_nome* (cleaned name) comparisons are grouped into six categories. Specifically[60]:

gamma_clean_rep_nome:
- 5 **if** *nome_ratio* == 100
- 4 **if** (*nome_ratio* >= 95) **or** (*firstLast_ratio* == 100)
- 3 **if** (*nome_ratio* >= 88) **or** (*firstLast_ratio* >= 95)
- 2 **if** (*nome_ratio* >= 80) **or** (*firstLast_ratio* >= 88)
- 1 **if** (*nome_ratio* >= 70) **or** (*firstLast_ratio* >= 80)
- 0 otherwise

where

- *firstLast* returns just the first and last token (excluding all middle names), as described above
- *nome_ratio* = *ratio*(accountGrouping1.*clean_rep_nome*,

[59] The fields *clean_rep_cel* and *clean_rep_fone* are combined into a single gamma field, as described below.
[60] Each gamma value is given by the first line whose condition matches.

accountGrouping2.*clean_rep_nome*)

- *firstLast_ratio* = *ratio*(*firstLast*(accountGrouping1.*clean_rep_nome*),

  firstLast(accountGrouping2.*clean_rep_nome*))

165. The gamma values for the *clean_rep_email* (cleaned email) comparisons are grouped into four categories. Specifically:

gamma_clean_rep_email:
- 3 **if** *email_ratio* == 100
- 2 **if** (*email_ratio* >= 92)
- 1 **if** (*email_beforeAt_ratio* >= 90)
- 0 otherwise

where

- *beforeAt* returns the portion of the string before the first @ character
- *email_ratio* = *ratio*(accountGrouping1.*clean_rep_email*,

  accountGrouping2.*clean_rep_email*)
- *email_beforeAt_ratio* = *ratio*(*beforeAt*(accountGrouping1.*clean_rep_email*),

  beforeAt(accountGrouping2.*clean_rep_email*))

166. The *clean_rep_cel* (cleaned cell phone) and *clean_rep_fone* (cleaned phone) comparisons get one combined gamma value grouped into four categories. It is standard to combine closely related fields into a single gamma value, as with *clean_rep_cel* and *clean_rep_fone*, so as to not "double-count" the weight of fields whose values are very often the same. It is also standard to involve quantities derived from the strings, such as minimum length restrictions in phone number, to improve the ability to detect intermediate degrees of agreement. The gamma value for *clean_rep_cel* and *clean_rep_fone* is determined by:

gamma_clean_rep_cel_or_fone:
- 3 **if** (*cel_ratio* == 100) or (*fone_ratio* == 100)
- 2 **if** (*cel_ratio* >= 92) and (*cel_bothLongerThan6*)
- 2 **if** (*fone_ratio* >= 92) and (*fone_bothLongerThan6*)

- 1 **if** (*cel_ratio* >= 83) and (*cel_bothLongerThan6*)
- 1 **if** (*fone_ratio* >= 83) and (*fone_bothLongerThan6*)
- 0 otherwise

where

- *len* returns the length of a string
- *cel_ratio* = *ratio*(accountGrouping1.*clean_rep_cel*,

  accountGrouping2.*clean_rep_cel*)
- *cel_bothLongerThan6* = true if

  len(accountGrouping1.*clean_rep_cel*) >= 6 AND

  len(accountGrouping2.*clean_rep_cel*) >= 6
- *fone_ratio* = *ratio*(accountGrouping1.*clean_rep_fone*,

  accountGrouping2.*clean_rep_fone*)
- *fone_bothLongerThan6* = true if

  len(accountGrouping1.*clean_rep_fone*) >= 6 AND

  len(accountGrouping2.*clean_rep_fone*) >= 6.

167. The gamma values for the *clean_rep_end* (cleaned street name) comparisons are grouped into four categories. Specifically:

gamma_clean_rep_end:
- 3 **if** *end_jwsim* == 100
- 2 **if** *end_jwsim* >= 92
- 1 **if** *end_jwsim* >= 75
- 0 otherwise

where

- *end_jwsim* = *jwsim*(accountGrouping1.*clean_rep_end*,

  accountGrouping2.*clean_rep_end*)

168. The gamma values for the *clean_rep_numero* (cleaned house number) comparisons are grouped into four categories. Specifically:

gamma_clean_rep_numero:
- 3 **if** *numero_ratio* == 100
- 2 **if** numero_ratio >= 82
- 1 **if** numero_ratio >= 75
- 0 otherwise

where

- *numero_ratio* = *ratio*(accountGrouping1.*clean_rep_numero*, accountGrouping2.*clean_rep_numero*)

169. The gamma values for the *clean_rep_cep* (cleaned ZIP code analogue) comparisons are grouped into four categories. Specifically:

gamma_clean_rep_cep:
- 3 **if** *cep_ratio* == 100
- 2 **if** *cep_ratio* >= 82 **else**
- 1 **if** *cep_ratio* >= 75
- 0 otherwise

where

- *cep_ratio* = *ratio*(accountGrouping1.*clean_rep_cep*, accountGrouping2.*clean_rep_cep*)

170. The gamma values for the *clean_rep_login* (cleaned login name) comparisons are grouped into three categories. Specifically:

gamma_clean_rep_login:
- 2 **if** *login_ratio* == 100
- 1 **if** *login_ratio* >= 75
- 0 otherwise

where

- *login_ratio* = *ratio*(accountGrouping1.*clean_rep_login*, accountGrouping2.*clean_rep_login*)

171. The gamma values for the *clean_rep_pwd_secondary_set* (set of cleaned hashed passwords) comparison are binary. Specifically,

gamma_clean_rep_pwd_secondary_set:
- 1 **if** accountGrouping1.*clean_rep_pwd_secondary_set* has a nonempty intersection with accountGrouping2.*clean_rep_pwd_secondary_set*
- 0 otherwise

Set intersection leads to a binary outcome, which is appropriate for *clean_rep_pwd_secondary_set* because string nearness of two password hashes does not indicate nearness of the corresponding passwords themselves.

172. The gamma values for the *clean_rep_cpf_set* (set of cleaned SSN analogues) comparisons are grouped into three categories. Specifically:

gamma_clean_rep_cpf_set:
- 2 **if** *max_cpf_ratio* == 100
- 1 **if** *max_cpf_ratio* >= 88
- 0 otherwise

where

- *max_cpf_ratio* is the largest value achieved by *ratio*(*clean_rep_cpf1*, *clean_rep_cpf2*)
- *clean_rep_cpf1* ranges over all values in accountGrouping1.*clean_rep_cpf_set*
- *clean_rep_cpf2* ranges over all values in accountGrouping2.*clean_rep_cpf_set*

173. The gamma values for the *clean_rep_datanasc_set* (set of cleaned dates of birth) comparison are binary. Specifically,

gamma_clean_rep_datasanc_set:
- 1 **if** accountGrouping1.*clean_rep_datanasc_set* has a nonempty intersection with accountGrouping2.*clean_rep_datanasc_set*
- 0 otherwise

Set intersection again leads to a binary outcome, which is appropriate for *clean_rep_datanasc_set* because string nearness of two dates is not very helpful for determining linkage due to collisions in the nearby values for this field.

174. The choice of string nearness measure for each of the fields above is made for the reasons described in Section IV.B.

175. As an illustration of how the gamma cutoff values were selected, we now describe part of the process for the *rep_end* (street name) field.[61] Exhibit 4 shows 100 randomly-selected pairs from the blocking set for which *end_jwsim* is at least 50 and less than 100. Those pairs of account groupings for which *end_jwsim* is below 50 all clearly disagree on the *clean_rep_end* field, and so should be assigned *gamma_clean_rep_end* value 0. On the other hand, those pairs with *end_jwsim* equal to 100 are exact matches (after cleaning); all such pairs (and only such pairs) should be assigned *gamma_clean_rep_end* value 3 (the highest value). The randomly-selected pairs with intermediate values are used to determine the boundary between *gamma_clean_rep_end* values 0 and 1 and between values 1 and 2. Observe that the extent of agreement among those pairs with *end_jwsim* 92 and higher is generally quite high, even if not perfect. Those pairs with *end_jwsim* below 92 generally have less good agreement, but still are of some value above 75. For example, row 27 largely agrees (mostly differing in punctuation), while row 29 is dubious (having the same city name, but with none of the rest agreeing). Very few, if any, pairs with *end_jwsim* below 75 have much agreement. I performed a similar analysis on a larger subset of blocking to determine the cutoffs of 75, 92, and 100 that appear in the definition of *gamma_clean_rep_end* above.[62] Cutoffs for the other fields were selected similarly.

[61] "The selection of a distance function and a proper threshold is … a knowledge intensive activity that demands experimental evaluation." (p. 15 of Hernández, M.A. and Stolfo, S.J. Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem. *Data Mining and Knowledge Discovery* 2, 9–37 (1998).)

[62] Of course, there is some freedom in the choice of these cutoffs, and other choices of numbers that are fairly similar to these could also potentially be justified: "small changes to some thresholds defining equality over two fields … are expected to have little impact on the contents of the formed clusters." (Id. at p. 31). I have verified that the overall probabilistic record linkage process is not especially sensitive to the details of this choice, so long as the decisions are made reasonably.

176. The output of this step is the set of comparison vectors, one vector for each pair in the blocking set.

6. Step 6: Calculate Match Probabilities

177. Next, the set of comparison vectors for those pairs of account groupings in the blocking set are used to determine match probabilities in accordance with the Fellegi-Sunter model of probabilistic linkage. I accomplished this task by running the Splink probabilistic record linkage package (see Section IV.C) on this set of comparison vectors. Splink uses the EM algorithm to automatically calculate the *m*-probabilities and *u*-probabilities for the fields in the comparison vector, and then assigns a match probability to each pair of account groupings in the blocking set. I used Splink's default settings for the maximum number of EM iterations (25) and the EM convergence tolerance (0.0001).

178. Typically, the input to Splink would also contain the specification on how to carry out the blocking step (step 4) and the rubric-calculation step (step 5). I performed these steps outside of Splink (as described in steps 4 and 5) using the massively-parallel cloud infrastructure previously described, and the results were fed into Splink at the appropriate entry point.

179. The output of this step is an assignment of the match probability to each of the 117,563,883 pairs of account groupings in the blocking set.

7. Step 7: Compute Clusterings for Several Fixed Thresholds

180. I next use Splink to compute clusterings of account groupings for various fixed thresholds on the match probability; these clusterings will be the input to Step 8 (dynamic clustering step). These fixed thresholds are the match probabilities occurring at percentiles [0.006 … 0.009, 0.01 … 0.09, 0.1 … 0.9, 1 ... 99, 99.1 … 99.9, 99.91 … 99.99, 99.991 … 99.999, 100] within the blocking set, for a total of 149 fixed-threshold values. These thresholds were chosen to provide broad coverage of coarse to fine clusterings, especially at the extremes.

181. For any given fixed threshold, the goal is to create a clustering such that each account grouping is in exactly one cluster of the clustering. This clustering is formed by taking the transitive closure of the set of all account grouping pairs whose match probability is at least the given threshold.[63]

182. This collection of 149 clustering threshold percentiles represents the range of (quantized) trade-off decisions for how tightly or how loosely to generate the clusters from the blocking set of account grouping pairs with their respective match probabilities. For example, with threshold percentile 95, only the most closely-related pairs (as measured by match probability) will become part of the same cluster, and there will be a large number of clusters and a small average cluster size. With threshold percentile 5, on the other hand, even very distantly-related pairs (as measured by match probability) will become part of the same cluster, and there will be a smaller number of clusters and a higher average cluster size.

183. The 149 clusterings of account groupings are the output of this step.

8. Step 8: Dynamic Clustering Process

184. Now that there are 149 fixed-threshold clusterings, in principle one could simply select a particular fixed-threshold clustering from among them as the clustering output. The clustering literature describes standard techniques for selecting a fixed threshold based on balancing two competing tensions[64]: (1) records that do not belong in the same cluster should not be joined (i.e., one should avoid "overclustering") and (2) records that do belong in the same cluster should be joined in the same cluster (i.e., one should avoid "underclustering"). A high

[63] Equivalently, this clustering is the finest partition consistent with adding linkages between all account grouping pairs whose match probability is at least the given threshold.

[64] Zhu and coauthors describe the need to balance false positives with false negatives (Y. Zhu, Y. Matsuyama, Y. Ohashi, and S. Setoguchi. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *Journal of Biomedical Informatics* 56 (2015): 80-86.)

fixed threshold will tend to undercluster, while a low fixed threshold will tend to overcluster. However, even choosing an optimal fixed-threshold clustering with respect to these tensions is not the best choice in this circumstance, as I now explain.

185. The TelexFree dataset presents an additional challenge due to the presence of heterogeneity: some Participants entered their information very consistently across many accounts, while other Participants, for a variety of reasons, entered information that differed more widely (e.g., name variants, typos, multiple email addresses or phone numbers, etc.). As a result, any fixed threshold will deal appropriately with at most one of these two situations (and often neither) – both high and low fixed thresholds will typically simultaneously overcluster some clusters while underclustering others.

186. It is possible to better navigate the tradeoff between overclustering and underclustering by choosing higher thresholds in a portion of the dataset where there is a "tightly bound" collection of account groupings having high match probability with each other and low match probability with every other account grouping, while choosing lower thresholds in a portion of the dataset where there is a more "loosely bound" collection of account groupings having moderately high match probability with each other but still much lower match probability with every other account grouping in the dataset.

187. Such a "dynamic clustering" technique (also known as heterogeneous clustering[65]), where higher thresholds are selected for some parts of the dataset, and lower thresholds are selected for other parts, is more appropriate than using a single fixed threshold for

[65] Heterogeneous clustering involves selecting a different threshold for different regions of a graphs so as to achieve clusters that balance overclustering and underclustering. (See Section 4.2.2 of Charpentier, Bertrand. *Multi-scale clustering in graphs using modularity.* Master's thesis, KTH, 2019.)

a dataset like the TelexFree dataset. I therefore chose a dynamic clustering technique in the aggregation methodology.

188. This task of selecting clustering thresholds dynamically across the dataset is accomplished in the aggregation methodology by (1) defining measures of overclustering and underclustering for a given cluster, and calculating the overclustering measure for every cluster in each of the 149 fixed-threshold clusterings computed in the last step, (2) using this data to compute a dynamic clustering of the dataset for each of 41 values of a "dynamic clustering parameter", where the overclustering calculations and dynamic clustering parameter guide the selection of a fixed threshold for each account grouping, and (3) choosing the dynamic clustering from among these 41 dynamic clusterings that optimally balances the degree of overclustering and underclustering.

i. Overclustering and Underclustering Measures

189. The clustering step of probabilistic record linkage, once match probabilities are assigned, can also be framed in terms of graph theory. Namely, the set of match probabilities determines a weighted graph[66], and in the graph clustering literature it is a standard technique to turn a weighted graph into clusters by taking the connected components of the subgraph consisting of those edges whose weight is above the given threshold[67]; this is equivalent to fixed-threshold clustering (and indeed Splink uses the standard graph package GraphFrame to compute fixed-threshold clusterings).

---

[66] The set of match probabilities determines an undirected weighted graph whose vertices (nodes) are records. (In the case of this aggregation methodology, the records are account groupings.) Each edge (connection) between a pair of vertices has weight equal to the match probability of the pair (and weight 0, or non-edge, if the pair is not in the blocking set).

[67] "Weight thresholding removes all edges with weight lower than a threshold value." (X Yan, LGS Jeub, A Flammini, F Radicchi, and S Fortunato. Weight thresholding on complex networks. *Physical Review E* 98.4 (2018): 042304.)

190. Determining a clustering within such a weighted graph (or "network") is sometimes framed in terms of "community detection"; here, the nodes are the account groupings, the weighted edges represent match probabilities between account groupings, and the communities one aims to find are the clusters consisting of those account groupings belonging to a single Participant.

191. Often the communities are determined by optimizing some scoring function, of which there are many possible options:

> The idea is that given a community scoring function, one can then find sets of nodes with high score and consider these sets as communities. All scoring functions build on the intuition that communities are sets of nodes with many connections between the members and few connections from the members to the rest of the network. There are many possible ways to mathematically formalize this intuition … The experiment demonstrates that even though many different structural definitions of communities have been proposed, these definitions are heavily correlated.[68]

192. As previously noted, in record linkage there is a tension between overclustering (the extent to which a cluster includes more records than is justified) and underclustering (the extent to which a cluster includes fewer records than is justified). In any automated technique, it is typically impossible to achieve both perfectly, but one can often obtain high quality results by balancing the two.[69] Towards achieving such a balance, I first define notions of overclustering and underclustering that are appropriate for the TelexFree dataset.

193. One standard way to measure overclustering in an unweighted graph is by computing the fraction of pairs of nodes within a cluster that occur within the graph. Within the community detection literature, this is closely related to the concept of "density", as well as other

[68] Yang, J. and Leskovec, J.. Defining and evaluating network communities based on ground-truth. *Knowledge and Information Systems, 42*(1): 2015, 181-213.

[69] Y. Zhu, Y. Matsuyama, Y. Ohashi, and S. Setoguchi. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *Journal of Biomedical Informatics* 56 (2015): 80-86.

notions of "internal connectivity".[70] A standard way to convert a weighted graph (with weights between 0 and 1) into an unweighted graph is to restrict to those edges having weight above 0.5; in the case of this aggregation methodology, this amounts to selecting those pairs of account groupings deemed by the Fellegi-Sunter model to have match probability greater than 0.5, which is a reasonable default for building a binary classifier (a function that turns a probability into an edge or non-edge, in this case).[71]

194. Accordingly, the aggregation methodology uses a notion of overclustering defined in terms of internal connectivity of this unweighted graph. Namely, define the *overclustering measure* of a given account grouping (with respect to a proposed clustering) to be the fraction of account groupings within its cluster with which it has a match probability above 0.5. In other words, it is the number of account groupings in its cluster with which it has a match probability over 0.5, divided by the total number of account groupings in its cluster. The average overclustering measure of a cluster (e.g., the fixed-threshold cluster of an account grouping for some particular fixed threshold) is then defined to be the average of the overclustering measures of each account grouping in the cluster. Finally, the overclustering measure of an entire proposed clustering (e.g., a candidate clustering for the dynamic clustering process) is defined to be the average overclustering measure of each account grouping's cluster, averaged across all account groupings. In this way, if the overclustering measure of a given clustering is close to 1, this indicates that the clusters do not contain many pairs of account groupings having low match

[70] "Density builds on intuition that good communities are well connected... It measures the fraction of the edges (out of all possible edges) that appear between the nodes in [the node set]." (Yang and Lescovec, 2015)

[71] "Most classifiers assign a probability for class membership to the instance to be classified. The default is to assume that a probability greater than or equal to 0.5 is for one class and below 0.5 for the other in binary classification." Section 3.3.4 of Uday Kamath and John Liu. *Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning*. Springer, 2021.

probability, while if the overclustering measure is close to 0, this indicates that the clusters contain many pairs of account groupings having low match probability.

195. One standard way to measure underclustering is by computing the fraction of pairs of nodes going between the cluster and its complement that occur within the graph. Within the community detection literature, this is closely related to the concepts of "separability" and "expansion", as well as other notions of "external connectivity".[72] As before, in the case of a weighted graph with weights representing probabilities, this can be computed in terms of the unweighted graph given by the restriction to edges with weight above 0.5.

196. Accordingly, the aggregation methodology uses a notion of underclustering defined in terms of external connectivity of this unweighted graph. Namely, define the *underclustering measure* of a given account grouping (with respect to a given clustering) to be the fraction of all account groupings with which it has a match probability above 0.5 that are already within its cluster. In other words, it is the number of account groupings in its cluster with which it has a match probability over 0.5, divided by the total number of account groupings in the dataset with which it has a match probability over 0.5. The underclustering measure of an entire proposed clustering (e.g., a candidate clustering for the dynamic clustering process) is then defined to be this quantity averaged across all account groupings. In this way, if the underclustering measure of a given clustering is close to 1, this indicates that its set of clusters already contains most of the pairs of account groupings having high match probability, while if the underclustering measure is close to 0, this indicates that there are many pairs of account groupings having high match probability which failed to both be included in any cluster.

[72] "Separability captures the intuition that good communities are well-separated from the rest of the network … meaning that they have relatively few edges pointing from [the node set] to the rest of the network. Separability measures the ratio between the internal and the external number of edges of [the node set]." (Yang and Leskovec)

ii. Computing Dynamic Clusterings for a Set of Dynamic Clustering Parameters

197. As previously described, for every fixed threshold, each account grouping is assigned a cluster. As the threshold decreases, each such cluster assignment grows larger (because it is the transitive closure of a set consisting of all pairs from the higher threshold, along with possible additional pairs). Because of how the fixed-threshold clusterings are defined, clusters grow larger by joining with one or more other clusters from a higher threshold, a process known as "agglomeration".[73] The goal of dynamic clustering is to select a different fixed threshold for each account grouping in such a way that (a) the choices are compatible across the entire dataset (i.e., if a particular account grouping chooses a given fixed threshold, then so must every other account grouping in the cluster resulting from that threshold), so that these choices collectively determine a clustering of the entire dataset, and (b) the measures of overclustering and underclustering for this clustering are balanced with each other.

198. This is achieved in the aggregation methodology by choosing, for each account grouping, the fixed-threshold clustering with the smallest threshold that is still consistent with the resulting cluster having average overclustering measure above a set parameter[74]. This parameter, representing the minimum required overclustering average, is referred to as the *dynamic clustering parameter*. This technique achieves condition (a) because the choice is based on a quantity that is averaged across the account grouping's cluster, ensuring that the choices are compatible across the entire dataset. This technique will achieve condition (b) upon optimizing the choice of dynamic clustering parameter, the procedure for which I now describe.

---

[73] See Section 5.4 of Sneath, P.H.A. and Sokal, R.R.. *Numerical Taxonomy: the Principles and Practice of Numerical Classification.* Freeman, New York (1973).

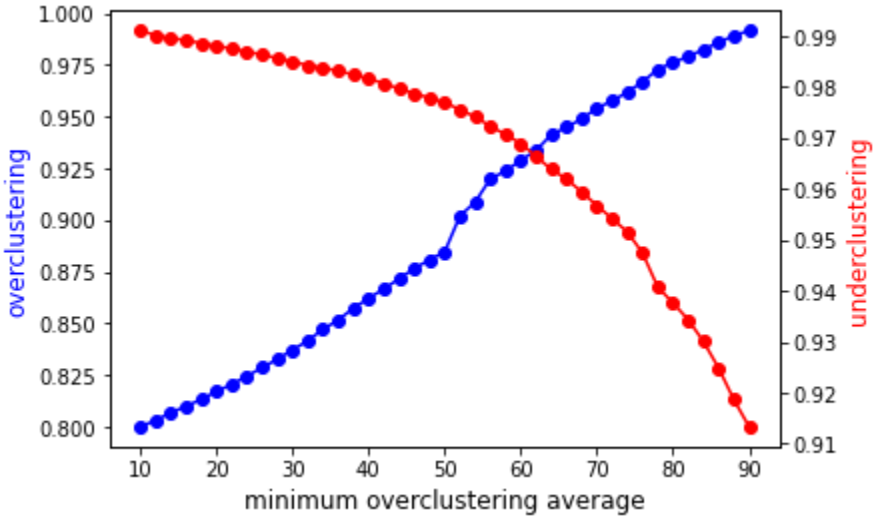
[74] Recall that an overclustering measure closer to 1 indicates that the clusters do not contain many pairs with low match probability.

199. For each of 41 choices of dynamic clustering parameter (0.10, 0.12, 0.14, …, 0.88, 0.90)[75], I computed the dynamic clustering for that parameter.

iii. Choosing the Optimal Dynamic Clustering

200. To choose the optimal dynamic clustering, I compared the overclustering and underclustering measures for each of the clusterings that result from these 41 dynamic clustering parameter values, in order to determine the clustering that achieves the best balance.

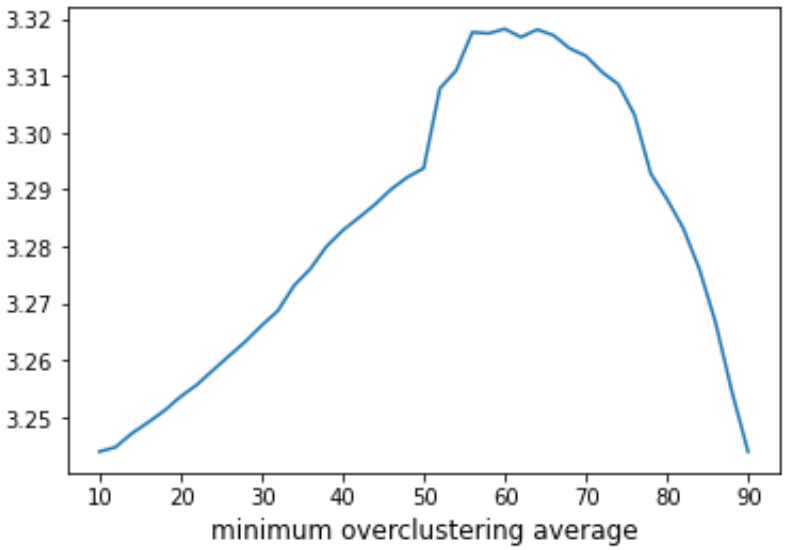
201. The first plot below shows the overclustering measure vs. the underclustering measure achieved for the 41 values of the dynamic clustering parameter (i.e., the minimum required overclustering average): 0.10, 0.12, 0.14, …, 0.88, 0.90. While the degree of underclustering is typically higher in absolute terms than the degree of overclustering, it is still important to make sure both are balanced.[76] I have therefore plotted the overclustering and underclustering averages with respective *y*-axes chosen so that their values have the same dynamic range.



[75] This range extends low and high enough to yield dynamic clusterings to provide a wide range of tradeoffs between overclustering and underclustering, and also has a small enough increment to provide fine-grained resolution that will help me select among these choices.

[76] Y. Zhu, Y. Matsuyama, Y. Ohashi, and S. Setoguchi. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *Journal of Biomedical Informatics* 56 (2015): 80-86.

202. Next, the second plot shows the sum of these two measures for overclustering and underclustering (i.e., the quantities that were shown independently in the first plot), where underclustering is scaled linearly to achieve the same dynamic range as overclustering. Again, the dynamic clustering parameter (i.e., the minimum required overclustering average) ranges from 0.10 to 0.90. The sum of overclustering and underclustering measures achieves a maximum for dynamic clustering parameter 0.60. This setting achieves an optimal tradeoff between overclustering and underclustering, and is therefore the value chosen for the aggregation methodology. This is also consistent with the first plot, which illustrates that for parameter values below 0.60, a small improvement in underclustering measure is exchanged for a larger drop in overclustering measure, while for parameter values above 0.60, a small improvement in overclustering measure is exchanged for a larger drop in underclustering measure.



203. The selected dynamic clustering (with dynamic clustering parameter 0.60) exhibits a better tradeoff between overclustering and underclustering than that of the dynamic clusterings for other dynamic clustering parameter choices. Furthermore, the selected dynamic clustering is much better than any choice of fixed-threshold clustering.

204. The resulting dynamic clustering of the 4,367,728 account groupings is the output of Step 8.

9. Step 9. Convert the Clustering of Account Groupings into a Clustering of User Accounts

205. The last step in determining the final clustering is to "ungroup" the account groupings back into User Accounts. In the deterministic step (Step 3), the 10,987,617 User Accounts were grouped into 4,367,728 account groupings. I accomplish the ungrouping in this step by joining the clustering output from the previous step (Step 8) with the lookup table mapping each User Account (identified by the *rep_id* field) to its account grouping (identified by the *ag_rep_id* field). (As previously mentioned, the lookup table was generated in Step 3, the deterministic step.)

206. The output of this join is the final clustering on all User Accounts other than the 62,725 User Accounts with *rep_nome* "John Williams". I then reassign all of the "John Williams" User Accounts to be in a single cluster of their own, as was previously determined in Step 4.

207. This step results in 1,566,383 clusters of User Accounts with a mean size of 7.01 User Accounts per cluster. For a summary of this clustering, see Section V and Exhibit 5.

10. Step 10. Tabulate Total Net Equity for Each Cluster

208. Each User Account has been assigned its Net Equity by Huron, as described above in Section III.A. The Total Net Equity of each cluster is defined to be the sum of the Net Equity values of each of its User Accounts. I tabulate this quantity for each cluster. For a summary of Net Winners and Net Losers, stratified into tiers by Total Net Equity, see Exhibit 5.

V. RESULTS

209. Based on the work described above, my aggregation methodology has determined that the 10,987,617 User Accounts attributable to TelexFree belong to 1,566,383 different Participants, with a mean size of 7.01 User Accounts per Participant. Of the 1,566,383 Participants, 1,484,702 are Net Losers and 81,681 are Net Winners. The Net Winners have a total net gain of $1,555,981,482.42, with an average net gain of $19,049.49 per Net Winner. The Net Losers have a total net loss of $1,782,171,730.09, with an average net loss of $1,200.36 per Net Loser. For more details and stratification by tier of gains or losses, see Exhibit 5.

210. The results are attached in two exhibits, Exhibit 6 and Exhibit 7; both files are in the CSV (comma-separated values) format. Exhibit 6 contains the main output of the aggregation methodology: there is one row per User Account (identified by *rep_id*), along with a Cluster ID indicating the cluster to which that row is assigned. For illustration purposes, the 11 fields used in the aggregation methodology, as well as *net_equity*, have been included for each row.

211. Exhibit 7 contains a summary of the output: there is one row per cluster (identified by the Cluster ID), along with a tabulated sum of *net_equity* values for all of the User Accounts in that cluster.

212. I also calculated error rates for the aggregation methodology. In evaluating the error rate of record linkage, one typically considers two types of error[77], "false positives" ("FP") and "false negatives" ("FN"). In the context of our aggregation setting, a false positive, also known as a "Type I error", occurs when a pair of User Accounts that do not actually belong to the same Participant are nevertheless placed in the same cluster by the aggregation methodology.

[77] D. Resnick and J. Asher. Measurement of Type I and Type II Record Linkage Error. *Proceedings of the American Statistical Association*, Government Statistics Section; Arlington, VA, USA: American Statistical Association; 2019. pp. 293–311.

A false negative, or "Type II error", occurs when a pair of User Accounts that do actually belong to the same Participant are placed in different clusters by the aggregation methodology. The "false discovery rate" ("FDR") and "false negative rate" ("FNR"), respectively, are standard ways of assessing these two types of errors in clustering problems, and these two error rates contribute respectively to the statistical measures of quality known as *precision* and *recall*. A survey of record linkage methodologies explains: "There are two types of errors that can be generated in record linkage, FN and FP, which respectively contribute to sensitivity (or recall) and PPV (or precision). Ideally both types of errors need to be minimized but there is always a trade-off. The harmonic mean of sensitivity and PPV, or f-measure, finds the best compromise between them … The f-measure attains a high value only if both sensitivity and PPV are high."[78] This final quantity, f-measure (also known as "F_1" or "accuracy"), is a standard statistical measure between 0.0 and 1.0, with larger values indicating a higher quality result. A accuracy of at least 0.90 has been considered "very good" by a review of record linkage software quality[79].

213. The information about pairwise match probabilities that is contained in a probabilistic record linkage model can be used to estimate these quantities. The authors of the fastLink package note this ("One advantage of probabilistic models over deterministic methods is that we can estimate the false discovery rate (FDR) and the false negative rate (FNR)")[80] and provide equations for estimating the two error rates, FDR and FNR.[81] Using the appropriate analogue of those equations to our setting, I estimate an FDR of 0.0198 and an FNR of 0.00377.

[78] Y. Zhu, Y. Matsuyama, Y. Ohashi, and S. Setoguchi. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *Journal of Biomedical Informatics* 56 (2015): 80-86.
[79] A. Ferrante and J. Boyd. A transparent and transportable methodology for evaluating Data Linkage software. *Journal of Biomedical Informatics* 45, no. 1 (2012): 165-172.
[80] T. Enamorado, B. Fifield, and K. Imai. Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records. *American Political Science Review*, 113(2) (2019): 353-371.
[81] Id. at equations (5) and (6)

These quantities imply a precision of 0.980, a recall of 0.996, and an accuracy of 0.988, indicating that the results of my aggregation methodology substantially exceeds the "very good" criterion.

VI. DISCUSSION OF EXAMPLE CLUSTER

214. To illustrate several aspects of the aggregation methodology, I now discuss an example cluster. Exhibit 3 shows the contents of the cluster that has the greatest Total Net Equity (the cluster with Cluster ID 132785).

215. The Cluster ID, Total Net Equity, and Total Number of User Accounts for this cluster are displayed on the top left.

216. This cluster contains 471 User Accounts, each of which is displayed on its own row. For each User Account, the *rep_id* field is displayed along with *ag_rep_id* (the ID number of the account grouping that the User Account is contained in), the 11 fields used in the aggregation process (*rep_nome, rep_email, rep_cel, rep_fone, rep_end, rep_numero, rep_cep, rep_login, rep_pwd_secondary, rep_cpf, rep_datanasc*), and *net_equity* (as provided by Huron; see Section III.A). The rows are sorted by *net_equity*.

217. Looking at the contents of any given column, one can get some sense of both the similarity and the diversity of values for that field. For example, *rep_nome* is often "Maria Teresa Milagres Neves" or "Maria Neves", but many variations (involving punctuation, capitalization, spacing, abbreviations, typos, etc.) appear: "Maria Teresa m Neves", "Maria T Neves", "maria neves", "Mar ives", "Maria T. M. Neves", "maria Teresa Milagres", "Mara ves", "Marai Teresa Neves", "Mara Tes Nev", etc. Similarly, in the fields *rep_email*, *rep_cel,* and *rep_fone*, a small number of values predominate, but there is some variation. Observe that the

value of *rep_numero* sometimes matches the street number from *rep_end* ("2B"), sometimes matches the unit/apartment number from *rep_end* ("626"), and other times is null/empty.

218. In Step 1, the values of *rep_login* are transformed by truncating to the first 4 characters. For this set of records, this results in multiple instances of the prefixes "ttmi", "call", "galo", etc.

219. The cleaned values of the fields that result from Step 2 sometimes exactly agree even when the original (uncleaned) versions do not. For example, the *rep_nome* values "Maria Neves" and "maria neves" are both cleaned to the (lowercase) *clean_rep_nome* value "maria neves". When field values consist entirely of garbage characters (e.g., "." in *rep_numero*), they are cleaned to the empty string and then replaced by null, which means that they will count as missing data when assigning degrees of agreement in Step 5 (rather than, e.g., counting "." as agreeing perfectly with other instances of ".").

220. In Step 3, the 471 User Accounts are collapsed into 129 account groupings based on grouping together those User Accounts having exact agreement on the cleaned versions of the 8 fields *rep_nome, rep_email, rep_cel, rep_fone, rep_end, rep_numero, rep_cep,* and *rep_login*. The account grouping is indicated by the field *ag_rep_id* (which is defined to be the least *rep_id* among User Accounts in that account grouping). These account groupings range in size from 1 User Account to 71 User Accounts.

221. Blocking (Step 4) is not directly illustrated by this cluster output, but it must be the case that for each row in the table, there is some other row in the table that appears together with it in the blocking set (via their respective account groupings).

222. For each pair of records in the blocking set, the comparison vector (indicating degrees of agreement on each field) was computed in Step 5. This too is not directly illustrated

by the cluster output, though one can see this indirectly by visually comparing the field values. For example, rows 1 and 2 have exact agreement on *rep_nome* ("Maria Teresa Milagres Neves"), *rep_email* ("ttmilagres@gmail.com"), *rep_cel* ("5089043810"), *rep_fone* ("5089043810"), *rep_cep* ("01702"), and *rep_pwd_secondary* ("c99fb769d8bcee3287936bf71335976e"), but partial or no agreement on the other fields. Therefore the associated pair of account groupings was assigned the highest level of agreement for the gamma values corresponding to these fields.

223. In Step 6, the comparison vector is used to compute a match probability for each pair of account groupings. Again, this is not directly illustrated by the cluster output, but it is straightforward to infer, on the basis of the patterns of cleaned field agreement, that certain pairs of rows are likely in an account grouping pair that is assigned high match probability because they agree exactly in many fields, while other pairs of rows are likely in an account grouping pair that is assigned lower match probability because they agree only partially or not at all on several fields. For example, the match probability between rows 1 and 2 is likely to be quite high.

224. In the fixed-threshold clusterings computed in Step 7, pairs with high match probability are clustered together for many different options of fixed threshold (both high and low thresholds), while pairs with low match probability are clustered together only for low fixed thresholds (if any). (The clustering resulting from a low fixed threshold is coarse, with a smaller number of bigger clusters, while the clustering resulting from a high fixed threshold is fine, with a larger number of smaller clusters.)

225. The dynamic clustering process (Step 8) determines an optimal tradeoff between overclustering and underclustering, and uses that information to compute the final clustering of account groupings. Although the resulting cluster in the exhibit may not be perfectly accurate (and indeed, no method will achieve perfect accuracy on any real-world dataset), by achieving a

balance between overclustering and underclustering, this cluster appears to exhibit an error rate well within what would be expected from the estimates computed in Section V, making it high quality and suitable for the goals of the aggregation methodology.

226. In Step 9, these 129 account groupings are expanded back into the corresponding set of 471 User Accounts that are shown in the cluster output. The Cluster ID of the cluster is defined to be the least value of *rep_id* occurring for a User Account in the cluster.

227. Finally, in Step 10, the Total Net Equity for the cluster is tabulated by summing the Net Equity for each of these 471 User Accounts. Observe that the gains are highly concentrated in the first 2 User Accounts (rows 1 and 2), and that only 10 of 471 User Accounts (rows 1-10) have net gains. This pattern is fairly typical among the Net Winners having large gains.

VII. CONCLUSION

228. Based on the methodology above, and for the reasons stated, I believe that the results in the output files, Exhibit 6 and Exhibit 7, provide a reasonable and reliable assignment of User Accounts to Participants.

EXHIBIT 1

Cameron E. Freer

955 Massachusetts Ave. #256
Cambridge, MA 02139-3233
USA

email: freer@mit.edu
www: https://cfreer.org/
phone: +1-716-713-8228

Experience (Academia)

Massachusetts Institute of Technology

Research Scientist, Department of Brain and Cognitive Sciences. 2019 – Present
Research scientist in the Probabilistic Computing Project.

Postdoctoral Associate, Department of Brain and Cognitive Sciences. 2013 – 2015
Research postdoc. Postdoctoral mentors: Joshua B. Tenenbaum and Vikash K. Mansinghka

Postdoctoral Fellow, Computer Science and Artificial Intelligence Laboratory. 2011 – 2013
Research postdoc. Postdoctoral mentors: Joshua B. Tenenbaum and Leslie P. Kaelbling

Instructor in Pure Mathematics, Department of Mathematics. 2008 – 2010
Research and teaching postdoc. Postdoctoral mentors: Hartley Rogers, Jr. and Michael Sipser

University of Hawai'i at Mānoa

Junior Researcher, Department of Mathematics. 2010 – 2011
Research and teaching postdoc. Postdoctoral mentor: Bjørn Kjos-Hanssen

Experience (Industry)

Borelian Corporation

Founder and Chief Scientist. 2016 – Present
Scientific consulting. Contractor to Gamalon (2016–2017), Remine (2016–2018), University of Toronto (2017–2018), MIT (2018–2019), Boston University (2020).
Other projects include an algorithmic trading platform (building the live trading system, designing strategies, and aggregating and cleaning financial data from multiple sources).

Remine

Chief Scientist. 2017 – 2018
Real estate data analytics firm. I built their initial machine learning models, and helped design their data architecture and build their data science team. This involved testing performance of several databases, and aggregating large datasets based on fuzzy matches.

Gamalon Labs

Research Scientist. 2013 – 2016
DARPA contractor, Probabilistic Programming for Advancing Machine Learning.
I researched theoretical and applied aspects of probabilistic programming languages.

Analog Devices

Lyric Labs Visiting Fellow. 2013 – 2014

Hibernia Atlantic

Advisory Board Member. 2011 – 2012

EDUCATION

Harvard University

Ph.D. in Mathematics, 2008. Dissertation: *Models with high Scott rank.*
Advisor: Gerald E. Sacks

University of Chicago

S.B. in Mathematics with Honors, 2003.

RESEARCH INTERESTS

Interactions of randomness and computation, including the foundations of probabilistic computing, efficient samplers and testing methods for probabilistic inference, and the mathematics of random structures.

PUBLICATIONS

Note: several of my publications use alphabetical author lists, as is the convention in mathematics and theoretical computer science. These have been indicated by a ⋆.

⋆ N. L. Ackerman, C. E. Freer, and R. Patel, *On computable aspects of algebraic and definable closure,* Journal of Logic and Computation, 31, no. 1, 2–19, 2021.

F. A. Saad, C. E. Freer, M. C. Rinard, and V. K. Mansinghka, *The Fast Loaded Dice Roller: A near-optimal exact sampler for discrete probability distributions,* Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020), Proceedings of Machine Learning Research (PMLR) 108, 1036–1046, 2020.

⋆ N. L. Ackerman, C. E. Freer, and R. S. Lubarsky, *An introduction to feedback Turing computability,* Journal of Logic and Computation 30, no. 1, 27–60, 2020.

F. A. Saad, C. E. Freer, M. C. Rinard, and V. K. Mansinghka, *Optimal approximate sampling from discrete probability distributions,* Proceedings of the ACM on Programming Languages 4, POPL, 36:1–36:31, 2020.

⋆ N. L. Ackerman, C. E. Freer, and R. Patel, *Computability of algebraic and definable closure,* Proceedings of the Symposium on Logical Foundations of Computer Science (LFCS 2020), LNCS Vol. 11972, 1–11, 2020.

⋆ N. L. Ackerman, J. Avigad, C. E. Freer, D. M. Roy, and J. M. Rute, *Algorithmic barriers to representing conditional independence,* Proceedings of the 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2019), 2019.

⋆ N. L. Ackerman, C. E. Freer, and D. M. Roy, *On the computability of conditional probability,* Journal of the ACM 66, no. 3, 23:1–23:40, 2019.

⋆ N. L. Ackerman, C. E. Freer, and R. S. Lubarsky, *Feedback computability on Cantor space,* Selected Papers of Logic in Computer Science (LICS) 2015 and 2016, Logical Methods in Computer Science, 15, no. 2, 7:1–7:18, 2019.

F. A. Saad, C. E. Freer, N. L. Ackerman, and V. K. Mansinghka, *A family of exact goodness-of-fit tests for high-dimensional discrete distributions,* Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS 2019), Proceedings of Machine Learning Research (PMLR) 89, 1640–1649, 2019.

PUBLICATIONS (CONT'D)

⋆ N. L. Ackerman, C. E. Freer, and R. Patel, *The entropy function of an invariant measure*, Proceedings of the 14th and 15th Asian Logic Conferences, World Scientific, 3–34, 2019.

S. Staton, D. Stein, H. Yang, N. L. Ackerman, C. E. Freer, and D. M. Roy, *The Beta-Bernoulli process and algebraic effects,* Proceedings of the 45th International Colloquium on Automata, Languages, and Programming (ICALP 2018), 141:1-141:15, 2018.

⋆ N. L. Ackerman, C. E. Freer, and D. M. Roy, *On computability and disintegration,* Mathematical Structures in Computer Science 27, no. 8, 1287–1314, 2017.

⋆ N. L. Ackerman and C. E. Freer, *Graph Turing machines,* Logic, Language, Information, and Computation, Proceedings of WoLLIC 2017, LNCS Vol. 10388, Springer, 1–13, 2017.

⋆ N. L. Ackerman, C. E. Freer, A. Kwiatkowska, and R. Patel, *A classification of orbits admitting a unique invariant measure,* Annals of Pure and Applied Logic 168, no. 1, 19–36, 2017.

D. Cai, N. L. Ackerman, and C. E. Freer, *Priors on exchangeable directed graphs,* Electronic Journal of Statistics 10, no. 2, 3490–3515, 2016.

⋆ N. L. Ackerman, C. E. Freer, and R. Patel, *Invariant measures concentrated on countable structures,* Forum of Mathematics Sigma 4, e17, 59 pp., 2016.

⋆ N. L. Ackerman, J. Nešetřil, C. E. Freer, and R. Patel, *Invariant measures via inverse limits of finite structures,* European Journal of Combinatorics 52, 248–289, 2016.

⋆ N. L. Ackerman, C. E. Freer, and R. S. Lubarsky, *Feedback Turing computability, and Turing computability as feedback,* Proceedings of the 30th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2015), 523–534, 2015.

⋆ C. E. Freer, B. Kjos-Hanssen, A. Nies, and F. Stephan, *Algorithmic aspects of Lipschitz functions,* Computability 3, 45–61, 2014.

⋆ C. E. Freer, D. M. Roy, J. B. Tenenbaum, *Towards common-sense reasoning via conditional simulation: legacies of Turing in Artificial Intelligence,* in Turing's Legacy: Developments from Turing's Ideas in Logic, ed. Rod Downey, ASL Lecture Notes in Logic, Cambridge University Press, 2014.

⋆ C. E. Freer and B. Kjos-Hanssen, *Randomness extraction and asymptotic Hamming distance,* Selected Papers of the 9th International Conference on Computability and Complexity in Analysis (CCA 2012), Logical Methods in Computer Science, 2013.

A. D. Wissner-Gross and C. E. Freer, *Causal entropic forces,* Physical Review Letters 110, 168702, 2013.

⋆ N. L. Ackerman and C. E. Freer, *A notion of a computational step for Partial Combinatory Algebras,* Proceedings of the 10th Annual Conference on Theory and Applications of Models of Computation (TAMC 2013), LNCS Vol. 7876, Springer, 133–143, 2013.

⋆ C. E. Freer and D. M. Roy, *Computable de Finetti measures,* Annals of Pure and Applied Logic 163, no. 5, 530–546, 2012.

PUBLICATIONS (CONT'D)

⋆ N. L. Ackerman, C. E. Freer, and D. M. Roy, *Noncomputable conditional distributions,* Proceedings of the 26th Annual IEEE Symposium on Logic in Computer Science (LICS 2011), 107–116, 2011.

A. D. Wissner-Gross and C. E. Freer, *Relativistic statistical arbitrage,* Physical Review E 82, 056104, 2010.

⋆ C. E. Freer and D. M. Roy, *Posterior distributions are computable from predictive distributions,* Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS 2010), Journal of Machine Learning Research W&CP 9, 2010.

⋆ C. E. Freer and D. M. Roy, *Computable exchangeable sequences have computable de Finetti measures,* in K. Ambos-Spies, B. Löwe, and W. Merkle (eds.): Mathematical Theory and Computational Practice, Proc. of the 5th Conf. on Computability in Europe (CiE 2009), LNCS Vol. 5635, Springer, 218–231, 2009.

DISSERTATION

C. E. Freer, *Models with high Scott rank*, Ph.D. Dissertation, Harvard University, 2008.

PREPRINTS

S. Spanbauer, C. E. Freer, and V. K. Mansinghka, *Deep involutive generative models for neural MCMC,* arXiv:2006.15167.

⋆ N. L. Ackerman, J. Avigad, C. E. Freer, D. M. Roy, and J. M. Rute, *On the computability of graphons,* arXiv:1801.10387.

⋆ N. L. Ackerman, C. E. Freer, and R. Patel, *Stable regularity for relational structures,* arXiv:1712.09305.

⋆ N. L. Ackerman, C. E. Freer, A. Kruckman, and R. Patel, *Properly ergodic structures,* arXiv:1710.09336.

⋆ N. L. Ackerman, C. E. Freer, and R. Patel, *Countable infinitary theories admitting an invariant measure,* arXiv:1710.06128.

⋆ N. L. Ackerman and C. E. Freer, *On the computability of graph Turing machines,* arXiv:1703.09406.

D. Cai, N. L. Ackerman, and C. E. Freer, *An iterative step-function estimator for graphons,* arXiv:1412.2129.

PATENTS

C. E. Freer and A. D. Wissner-Gross, *System and method for relativistic statistical securities trading,* U.S. Patent 8,635,133 (2014).

INVITED TALKS

- *Representation theorems for exchangeable structures: a computability-theoretic perspective.* Joint Mathematics Meetings, Denver. ASL invited address, January 17, 2020.
- *Computability of algebraic and definable closure.* Logical Foundations of Computer Science (LFCS 2020), Deerfield Beach, FL. Invited to give talk based on refereed conference proceedings, January 6, 2020.
- *Step algebras.* NII Shonan meeting on higher-order complexity and its applications, Shonan Village Center, Japan. Invited workshop talk, October 10, 2019.
- *Computable representations of exchangeable data.* Computability in Europe (CiE 2019), Special Session on Probabilistic Programming and Higher-Order Computation, Durham University, Durham, UK. Invited special session talk, July 15, 2019.

INVITED TALKS (CONT'D)

- *Computability of the Aldous–Hoover theorem,* Workshop on higher-order probabilistic computation, Bellairs Research Institute, Barbados. Invited workshop talk, March 18, 2019.
- *Model theory of invariant probabilistic constructions.* Logic Colloquium, UC Berkeley, September 8, 2017.
- *Some remarks on Category, Measure, and Invariance.* All Kinds of Mathematics Remind Me of You: Conference to celebrate the 70th Anniversary of Peter J. Cameron, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal. Invited conference talk, July 24, 2017.
- *Graph Turing machines.* Workshop on Logic, Language, Information and Computation (WoLLIC), London, UK. Invited to give talk based on refereed conference proceedings, July 19, 2017.
- *On the computability of graph Turing machines.* AMS Central Sectional Meeting, Special Session on Computability and Inductive Definability over Structures, Indiana University, Bloomington, Indiana. Invited special session talk, April 2, 2017.
- *Priors on propositions: towards probabilistic programming for theorem proving,* 2nd Conference on Artificial Intelligence and Theorem Proving (AITP 2017), Obergurgl, Austria. Invited conference talk, March 29, 2017.
- *Unique ergodicity and measures invariant under permutations of* $\mathbb{N}$*,* Workshop: Algorithmic Randomness Interacts with Analysis and Ergodic Theory, Banff International Research Station, Oaxaca, Mexico. Invited workshop talk, December 6, 2016.
- *Three Problems in Computable Probability Theory*, Workshop on Uncertainty in Computation, Simons Institute for the Theory of Computing, Berkeley, CA. Invited workshop talk, October 4, 2016.
- *Invariant measures via finite structures.* Workshop on Model Theory of Finite and Pseudofinite Structures, University of Leeds, Leeds, UK. Invited workshop talk, July 29, 2016.
- *Exchangeable constructions of countable structures.* Workshop on Bayesian Methods for Networks, Isaac Newton Institute, Cambridge, UK. Invited workshop talk, July 25, 2016.
- *Symmetric probabilistic constructions of countable structures.* 5th CSLI Workshop on Logic, Rationality, & Intelligent Interaction, Stanford University, Palo Alto. Invited workshop talk, May 29, 2016.
- *Exchangeable constructions of countable structures.* Workshop on Networks, Random Graphs, and Statistics, Columbia University, New York. Invited workshop talk, May 5, 2016.
- *Feedback Turing Computability, and Turing Computability as Feedback.* AMS Central Sectional Meeting, Special Session on Computability Theory and Applications, Loyola University, Chicago, Illinois. Invited special session talk, October 3, 2015.
- *The Weihrauch degrees of conditional distributions.* Workshop on Measuring the Complexity of Computational Content: Weihrauch Reducibility and Reverse Analysis, Schloss Dagstuhl, Germany. Invited workshop talk, September 25, 2015.
- *Exchangeable constructions via model theory.* Workshop on Logic and Random Graphs, Lorentz Center, Leiden. Invited workshop talk, August 31, 2015.
- *Ergodic Invariant Measures as Probabilistic Structures, Lecture 2: Connections with graphons.* LMS–EPSRC Durham Symposium on Permutation Groups and Transformation Semigroups, Durham, UK. Invited to give 3-lecture series with coauthors, July 23, 2015.
- *The topology of universal graphons.* Conference on Computability and Complexity in Analysis (CCA 2015), Meiji University, Tokyo. Invited to give talk based on conference proceedings, July 13, 2015.
- *Feedback Turing computability, and Turing computability as feedback.* Logic in Computer Science (LICS), Kyoto. Coauthor invited to give talk based on refereed conference proceedings, July 9, 2015.

INVITED TALKS (CONT'D)

- *Computability and complexity of conditioning and conditional independence.* Workshop on Logic and Computational Complexity (LCC15), Kyoto. Invited workshop talk, July 4, 2015.
- *Priors on exchangeable directed graphs.* 10th Conference on Bayesian Nonparametrics (BNP10), Raleigh, NC. Coauthor invited to give talk based on refereed conference proceedings, June 25, 2015.
- *Two "Vaught's Conjectures" for measures invariant under the logic action.* Second Workshop on Vaught's Conjecture, Berkeley. Invited workshop talk, June 3, 2015.
- *Computability of conditioning: approximate inference and conditional independence.* Workshop on Challenges and Trends in Probabilistic Programming, Schloss Dagstuhl, Germany. Invited workshop talk, April 29, 2015.
- *Infinitary model theory in the study of graphons.* AMS–ASL Joint Mathematics Meetings, Special Session on Beyond First-Order Model Theory, San Antonio. Invited special session talk, January 10, 2015.
- *Invariant measures concentrated on the Henson graph.* NII Shonan meeting on Algorithmic Randomness and Complexity, Shonan Village Center, Japan. Invited workshop talk, September 12, 2014.
- *On computability and disintegration.* Conference on Computability and Complexity in Analysis (CCA 2014), Technische Universität Darmstadt, Germany. Invited to give talk based on conference proceedings, July 22, 2014.
- *Computable invariant measures and algorithmically random structures.* Conference on Computability Theory and Foundations of Mathematics, Tokyo Institute of Technology, Japan. Invited conference talk, February 18, 2014
- *Unique invariant measures, with an application to algorithmically random structures.* Mini-conference on Analysis, Randomness, and Applications, University of South Africa, Pretoria. Invited conference talk, February 10, 2014.
- *Random symmetric constructions via inverse limits of finite structures.* AMS–ASL Joint Mathematics Meetings, Special Session on Logic and Probability, Baltimore. Invited special session talk, January 15, 2014.
- *A notion of a computational step for Partial Combinatory Algebras.* Conference on Theory and Applications of Models of Computation (TAMC 2013), University of Hong Kong. Invited to give talk based on refereed conference proceedings, May 21, 2013.
- *When is a graph random?* Workshop on Computability, Complexity, and Randomness, Schloss Dagstuhl, Germany. Invited workshop talk, January 9, 2012.
- *Aspects of randomness in analysis, graph theory, and probability theory.* Asian Logic Conference, Special Session on Algorithmic Randomness, Victoria University of Wellington, NZ. Invited special session talk, December 15, 2011.
- *Effective aspects of Lipschitz functions and variation.* Analysis and Randomness, Auckland. Invited workshop talk, December 12–13, 2011.
- *Invariant measures concentrated on countable structures.* AIM Workshop on Graph and Hypergraph limits, American Institute of Mathematics, Palo Alto. Invited workshop talk, August 16, 2011.
- *Invariant measures on countable structures.* KGRC Mini-workshop, Kurt Gödel Research Center, Vienna. Invited workshop talk, July 12, 2011.
- *The unreasonable effectiveness of statistical artificial intelligence.* Foundational Questions in the Mathematical Sciences, International Academy Traunkirchen, Austria. Invited workshop talk, July 9, 2011.
- *#P-complete conditional distributions.* Logic and Computational Complexity, Toronto. Invited workshop talk, June 25, 2011.

INVITED TALKS (CONT'D)

- *Noncomputable conditional distributions.* Logic in Computer Science (LICS), Toronto. Coauthor invited to give talk based on refereed conference proceedings, June 21, 2011.
- *Relativistic statistical arbitrage.* North American Financial Information Summit, New York City. Invited talk, May 24, 2011.
- *Invariant measures on countable models.* AMS–ASL Joint Mathematics Meetings, Special Session on Logic and Analysis, New Orleans. Invited special session talk, January 7, 2011.
- *The computability of exchangeable sequences.* Invited talk, MIT, February 10, 2010.
- *Mechanising mathematics.* Interactive Theorem Proving workshop, Cambridge, UK. Invited workshop talk, August 24, 2009.
- *Computable probability theory.* PROMYS 20th Year Celebration, Boston, MA. Invited conference talk, July 26, 2009.
- *Computable exchangeable sequences have computable de Finetti measures.* Computability in Europe (CiE 2009), Heidelberg, Germany. Invited to give talk based on refereed conference proceedings; joint work presented by D. M. Roy, July 20, 2009.
- *Computable de Finetti measures.* Mid-Atlantic Mathematical Logic Seminar, MAMLS @ Harvard: A meeting on the intersections of logic and mathematics, Cambridge, MA. Invited conference talk, May 9, 2009.
- *Models with high Scott rank.* AMS Eastern Sectional Meeting, Special Session on Computability Theory and Effective Algebra, Wesleyan University, Middletown, CT. Invited special session talk, October 11, 2008.

SEMINAR TALKS

- *Exchangeable constructions of countable structures.* Logic, Games, and Graphs Seminar, Pennsylvania State University, November 6, 2019.
- *Computability of exchangeable sequences, arrays, and graphs.* Logic Seminar, Harvard University, March 11, 2019.
- *Computability of exchangeable sequences, arrays, and graphs.* Logic Workshop, City University of New York, February 22, 2019.
- *Feedback computability.* Logic Seminar, Harvard University, October 17, 2017.
- *Symmetric random constructions in model theory.* Model Theory Seminar, City University of New York, April 4, 2014.
- *Symmetric random constructions in model theory.* Model Theory Seminar, UC Berkeley, October 23, 2013.
- *A model-theoretic approach to characterizing randomness notions.* Buenos Aires Semester in Computability, Complexity, and Randomness, March 13, 2013.
- *Model-theoretic methods in continuum limits of countable structures.* Connecticut Logic Seminar, Wesleyan University, April 30, 2012.
- *Model-theoretic constructions of limit structures.* Algorithms, Combinatorics, and Optimization Seminar, Carnegie Mellon University, March 29, 2012.
- *Invariant measure concentrated on countable structures.* Mathematical Logic Seminar, Carnegie Mellon University, March 27, 2012.
- *Computability and probabilistic symmetries.* Semester on Semantics and Syntax: A Legacy of Alan Turing, Isaac Newton Institute, Cambridge, UK, February 21, 2012.
- *Invariant measures on countable models.* Logic Workshop, Harvard University, October 27, 2010.
- *Invariant measures on countable models.* Logic Workshop, City University of New York, October 22, 2010.
- *Noncomputability of conditional probability.* Probability Seminar, MIT, March 1, 2010.

SEMINAR TALKS (CONT'D)

- *Computable probability theory and de Finetti's theorem.* Logic Seminar, University of Chicago, May 18, 2009.
- *Computable exchangeable sequences have computable de Finetti measures.* Logic and Computation Seminar, University of Pennsylvania, April 13, 2009.
- *Models with high Scott rank.* Logic Seminar, University of Notre Dame, November 29, 2007.

CONTRIBUTED TALKS

- *On the computability of graphons.* Association for Symbolic Logic, North American Annual Meeting, New York. May 21, 2019.
- *A classification of structures admitting a unique invariant measure.* Workshop on Homomorphisms and Graph Limits III, Hraničnı Zámeček, Hlohovec, Czech Republic. March 25, 2015.
- *Computable invariant measures and algorithmically random structures.* International Congress of Mathematicians, Seoul, August 14, 2014.
- *Computable invariant measures and algorithmically random structures.* Association for Symbolic Logic, North American Annual Meeting, Boulder. May 21, 2014.
- *Topological aspects of dense graph limits.* Arbeitsgemeinschaft: Limits of Structures, Mathematisches Forschungsinstitut Oberwolfach, Germany, April 4, 2013.
- *Computability and Conditional Probability.* Conference on Computability Theory and Foundations of Mathematics, Tokyo Institute of Technology, Japan, February 20, 2013.
- *Model-theoretic constructions of limit structures.* Workshop on Graph Homomorphisms, Limits, and Structures II, Hraničnı Zámeček, Hlohovec, Czech Republic. January 24, 2012.
- *Invariant measures concentrated on countable structures.* Workshop on Homogeneous Structures, Leeds. July 21, 2011.
- *Effective aspects of Lipschitz functions.* Association for Symbolic Logic, North American Annual Meeting, Berkeley. March 24, 2011.
- *The computability of conditional probability distributions.* AMS–ASL Joint Mathematics Meetings, San Francisco. January 15, 2010.
- *The complexity of computable conditional probability.* Computability in Europe (CiE 2009), Heidelberg, Germany. July 20, 2009.
- *Computable exchangeable sequences have computable de Finetti measures.* Association for Symbolic Logic, North American Annual Meeting, Notre Dame. May 20, 2009.

WORKSHOPS

Computability and Complexity

- *Algorithmic Randomness,* American Institute of Mathematics. Invited workshop participant, August 10–14, 2020, online.
- *Higher-order complexity and its applications,* NII Shonan Meeting. Invited speaker, October 7–10, 2019, Shonan Village Center, Japan.
- *Algorithmic Randomness Interacts with Analysis and Ergodic Theory,* Banff International Research Station for Mathematical Innovation and Discovery. Invited workshop participant, December 4–9, 2016, Oaxaca, Mexico.
- *Measuring the Complexity of Computational Content: Weihrauch Reducibility and Reverse Analysis,* Schloss Dagstuhl. Invited workshop participant, September 20–25, 2015, Dagstuhl, Germany.
- *Logic and Computational Complexity (LCC15)*. Invited speaker, July 4–5, 2015, Kyoto.
- *Algorithmic Randomness and Complexity,* NII Shonan Meeting. Invited speaker, September 8–12, 2014, Shonan Village Center, Japan.

WORKSHOPS (CONT'D)

Computability and Complexity (cont'd)

- *Logic, Probability, and Reflection IV,* MIRI. Workshop participant, September 7–13, 2013, Berkeley.
- *Computable Stability Theory,* American Institute of Mathematics. Workshop participant, August 12–19, 2013, Palo Alto.
- *Towards Efficient Homomorphic Encryption,* IdeaLab 2013, Institute for Computational and Experimental Research in Mathematics (ICERM). Workshop participant, July 15–19, 2013, Providence, RI.
- *Computability, Complexity, and Randomness,* Universidad de Buenos Aires. Invited special semester participant, March 12–18, 2013, Buenos Aires, Argentina.
- *Semantics and Syntax: A Legacy of Alan Turing,* Isaac Newton Institute. Special semester participant, February 5–26, 2012, Cambridge, UK.
- *Computability, Complexity, and Randomness,* Schloss Dagstuhl. Invited speaker and workshop participant, January 8–13, 2012, Dagstuhl, Germany.
- *Analysis and Randomness,* University of Auckland. Invited speaker, December 12–13, 2011, Auckland.
- *Reverse Mathematics Workshop,* University of Chicago. Workshop participant, September 16–18, 2011, Chicago.
- *Foundational Questions in the Mathematical Sciences,* International Academy Traunkirchen. Invited speaker, July 9-11, 2011, Traunkirchen, Austria.
- *Logic and Computational Complexity (LCC 2011)* at Logic in Computer Science (LICS 2011). Invited speaker, June 25, 2011, Toronto.

Bayesian Statistics, Machine Learning, and Probabilistic Programming

- *Higher-order probabilistic computation*, Bellairs Research Institute. Invited workshop participant, March 17–21, 2019, Barbados.
- *All of Bayesian Nonparametrics (BNP@NeurIPS)* at NeurIPS 2018. *Goodness-of-fit tests for high-dimensional discrete distributions with application to convergence diagnostics in Bayesian nonparametric inference,* poster, presented by coauthor F. A. Saad, December 7, 2018, Montreal, Canada.
- *Symposium on Advances in Approximate Bayesian Inference (AABI)* colocated with NeurIPS 2018. *Goodness-of-fit tests for high-dimensional discrete distributions with application to convergence diagnostics in approximate Bayesian inference,* poster, December 2, 2018, Montreal, Canada.
- *Probabilistic Programming Semantics* at Principles of Programming Languages (POPL 2017). Co-chair and workshop participant. *On computable representations of exchangeable data*, poster, and *Exchangeable Random Processes and Data Abstraction,* workshop paper, presented by coauthor S. Staton, January 17, 2017, Paris, France.
- *Uncertainty in Computation*, Simons Institute for the Theory of Computing. Invited workshop participant, October 4–7, 2016, Berkeley, CA.
- *Bayesian Methods for Networks,* Isaac Newton Institute. Invited speaker and workshop participant, July 25–27, 2016, Cambridge, UK.
- *Networks, Random Graphs, and Statistics,* Columbia University. Invited speaker and workshop participant, May 4–6, 2016, New York.
- *Probabilistic Programming Semantics* at Principles of Programming Languages (POPL 2016). Program Committee member and workshop participant. *Exchangeable random primitives,* workshop paper, presented by coauthor D. M. Roy, January 23, 2016, St. Petersburg, FL.

WORKSHOPS (CONT'D)

Bayesian Statistics, Machine Learning, and Probabilistic Programming (cont'd)

- *Bayesian Nonparametrics: The Next Generation* at Neural Information Processing Systems (NIPS 2015), *Priors on exchangeable directed graphs,* workshop paper, presented by coauthor D. Cai, December 12, 2015, Montreal, Canada.
- *Challenges and Trends in Probabilistic Programming,* Schloss Dagstuhl. Invited workshop participant, April 26–30, 2015, Dagstuhl, Germany.
- *Information and Entropy,* NIMBioS. Invited workshop participant, April 8–10, 2015, Knoxville, TN.
- *Probabilistic Programming* at Neural Information Processing Systems (NIPS 2013), *A tour through the theoretical foundations of probabilistic programming,* presented by coauthor D. M. Roy, December 7, 2012, South Lake Tahoe, NV.
- *Monte Carlo Methods for Modern Applications* at Neural Information Processing Systems (NIPS 2010), *When are probabilistic programs probably computationally tractable?*, presented by coauthor V. K. Mansinghka, December 10, 2010, Whistler, Canada.
- *Nonparametric Bayes Workshop* at Neural Information Processing Systems (NIPS 2009), *Predictive computable iff posterior computable*, workshop paper, presented by coauthor D. M. Roy, December 12, 2009, Whistler, Canada.
- *Learning Workshop,* Computational and Biological Learning Society. *Probabilistic programs, computability, and de Finetti measures,* workshop paper, presented by coauthor D. M. Roy, April 13–16, 2009, Clearwater, FL.

Combinatorics and Model Theory

- *Structure and Dynamics of Polish Groups,* École Polytechnique Fédérale de Lausanne (EPFL). *Entropy of Invariant Measures,* presented by coauthor N. L. Ackerman, March 19, 2018, Lausanne, Switzerland.
- *Model Theory and Combinatorics,* Institut Henri Poincaré. Invited workshop participant, January 29–February 2, 2018, Paris, France.
- *Model Theory of Finite and Pseudofinite Structures,* University of Leeds. Invited speaker and workshop participant, July 27–29, 2016, Leeds, UK.
- *Logic, Rationality, & Intelligent Interaction,* Center for the Study of Language and Information, Stanford University. Invited speaker and workshop participant, May 28–29, 2016, Palo Alto, CA.
- *Homogeneous Structures,* Banff International Research Station for Mathematical Innovation and Discovery. Invited workshop participant, November 8–13, 2015, Banff, Canada.
- *Logic and Random Graphs,* Lorentz Center. Invited speaker and workshop participant, August 31 – September 4, 2015, Leiden, The Netherlands.
- *LMS–EPSRC Durham Symposium on Permutation Groups and Transformation Semigroups,* Durham University. Invited speaker and workshop participant, July 20 – 30, 2015. Durham, UK.
- *Second Workshop on Vaught's Conjecture,* UC Berkeley. Invited workshop participant, June 1–5, 2015. Berkeley, CA.
- *Homomorphisms and Graph Limits III,* Hranični Zámeček. Invited workshop participant, March 22–27, 2015. Hlohovec, Czech Republic.
- *Workshop on Homogeneous Structures,* Trimester Program on Universality and Homogeneity, Hausdorff Research Institute for Mathematics, University of Bonn. Invited workshop and trimester program participant, October 27 – November 2, 2013, Bonn, Germany.
- *Arbeitsgemeinschaft: Limits of Structures,* Mathematisches Forschungsinstitut Oberwolfach. Workshop participant, March 31 – April 5, 2013, Oberwolfach, Germany.

WORKSHOPS (CONT'D)

Combinatorics and Model Theory (cont'd)

- *Graph Homomorphisms, Limits, and Structures II,* Hraničnı́ Zámeček. Invited workshop participant, January 23–27, 2012. Hlohovec, Czech Republic.
- *Graph and Hypergraph limits,* American Institute of Mathematics. Invited workshop participant, August 15–19, 2011, Palo Alto.
- *Workshop on Homogeneous Structures,* University of Leeds. Workshop participant, July 19–22, 2011, Leeds, UK.
- *KGRC Mini-workshop,* Kurt Gödel Research Center. Invited speaker, July 12–13, 2011, Vienna.
- *Model Theory of Fields,* Mathematics Research Community, American Mathematics Society. Invited workshop participant, June 19–25, 2010, Snowbird, UT.

Interactive Theorem Proving and Formalized Mathematics

- *Artificial Intelligence in Theorem Proving,* CNRS Centre Paul-Langevin. Program Committee member and workshop participant, March 25–30, 2018, Aussois, France.
- *Computer-aided mathematical proof,* Isaac Newton Institute. Invited workshop participant, July 10–14, 2017, Cambridge, UK
- *Artificial Intelligence in Theorem Proving,* University of Innsbruck. Invited speaker, March 26–30, 2017, Obergurgl, Austria.
- *Artificial Intelligence in Theorem Proving,* University of Innsbruck. Workshop participant, April 3–7, 2016, Obergurgl, Austria.
- *Interactive Theorem Proving Workshop,* Computer Laboratory, University of Cambridge. Invited speaker, August 24–25, 2009, Cambridge, UK.
- *Isabelle Theorem Prover Developer's Workshop,* Technische Universität München (TUM). Invited workshop participant, August 13–15, 2009, Munich.

SELECTED PRESS COVERAGE

Posterior distributions are computable from predictive distributions, 2010

- MIT News

Relativistic statistical arbitrage, 2010

- Wall Street Journal, IEEE Spectrum, Wired, New Scientist, BBC, The Atlantic, Fast Company

Causal entropic forces, 2013

- BBC, APS Physics Magazine, PhysOrg, Forbes, Quanta Magazine

The Fast Loaded Dice Roller, 2020

- Quanta Magazine, MIT News

TEACHING

University of Hawai'i at Mānoa

- *Math 243: Calculus III* (undergraduate course). Instructor, Spring 2011.
- *Math 649B: Logic* (graduate reading course on computational complexity). Co-instructor, Fall 2010.
- *Math 307: Linear Algebra and Differential Equations* (undergraduate course). Instructor, Fall 2010.

TEACHING (CONT'D)

Massachusetts Institute of Technology

· *6.885: Probabilistic Programming and Artificial Intelligence* (graduate course). Helped define problem sets and student projects; designed and delivered guest lectures, Spring 2020.
· *6.885: Probabilistic Programming and Artificial Intelligence* (graduate course). Helped define problem sets and student projects; designed and delivered guest lectures, Spring 2019.
· *18.515: Mathematical Logic* (graduate course). Instructor, Spring 2010.
· *18.03: Differential Equations* (undergraduate course). Recitation Instructor and Course Administrator, Fall 2009.
· *18.575: Model Theory* (graduate course). Instructor, Spring 2009.
· *18.03: Differential Equations* (undergraduate course). Recitation Instructor, Fall 2008.

Harvard University

· *QR 28: The Magic of Numbers* (undergraduate course). Head Teaching Fellow, Fall 2007.
· *Math 144: Model Theory and Algebra* (undergraduate course). Teaching Fellow, Spring 2007.
· *QR 28: The Magic of Numbers* (undergraduate course). Teaching Fellow, Fall 2006.
· *Math 141: Intro. to Mathematical Logic* (undergraduate course). Teaching Fellow, Fall 2006.
· *Math 143: Set Theory* (undergraduate course). Teaching Fellow, Spring 2006.
· *Math 141: Intro. to Mathematical Logic* (undergraduate course). Teaching Fellow, Fall 2005.
· *Math 137: Algebraic Geometry* (undergraduate course). Teaching Fellow, Spring 2005.
· *Math 144: Model Theory and Algebra* (undergraduate course). Teaching Fellow, Spring 2005.
· *Math 141: Intro. to Mathematical Logic* (undergraduate course). Teaching Fellow, Fall 2003.

Program in Mathematics for Young Scientists (PROMYS)

· *Undecidability and Hilbert's 10th Problem* (advanced seminar for high school students). Co-instructor, Summer 2020.

MENTORING

Doctoral

Boston University
· *Improved necessary and sufficient conditions for the existence of a subtle cardinal*, Peter Barendse, Ph.D. in Mathematics. Dissertation committee member, 2010.

Master's

Massachusetts Institute of Technology
· *Performance engineering in C and Julia for discrete distributions in probabilistic programming*, Anis Ehsani, M.Eng. student, Fall 2020 – Spring 2021.
African Institute for Mathematical Sciences (AIMS) Senegal
· *Logical methods in combinatorics*, Jean-Marc Mavugo, Master in Mathematical Sciences. External examiner, 2020.

Undergraduate

Massachusetts Institute of Technology
· *Evaluating inference quality for generative models using multivariate simulation based calibration*, Sharan Yalburgi, visiting student, Summer 2020 – Spring 2021.

MENTORING (CONT'D)

Undergraduate (cont'd)

University of Hawai'i at Mānoa

· *Computability and the Lovász Local Lemma,* Travis Hee Wai. NSF-funded undergraduate research mentor, Fall 2010. Hee Wai was selected as the Hawai'i Council of Engineering Societies' 2011 Student Engineer of the Year.

Undergraduate Research Opportunities Program (UROP), Massachusetts Institute of Technology

· *Combinatorial games and random structures,* Tamvana Makuluni. Undergraduate research mentor, Summer 2009.

· *Combinatorial games, linear orders, and logic,* Manuel Rivera. Undergraduate research mentor, Fall 2008 and IAP 2009.

Summer Program in Undergraduate Research (SPUR), Massachusetts Institute of Technology

· *A q-analogue of the Narayana numbers and a combinatorial interpretation,* Guilherme Issao Fujiwara. Research Experience for Undergraduates (REU) project mentor, 2005.

· *Some combinatorial results on subset sums of $\mathbb{Z}/n\mathbb{Z}$,* Kyungmin Kim. Research Experience for Undergraduates (REU) project mentor, 2005.

Freshman advising, Massachusetts Institute of Technology

· Advised 4 undergraduate math majors, Fall 2009.

High School

Research Science Institute (RSI)

· *Descriptive complexity of random bit strings,* Benjamin Dozier. High school mathematics research project mentor, 2007. Project led to Dozier being selected as a Finalist (top 40 nationwide) in the Intel Science Talent Search.

· *Cake-cutting with locally negative preference functions,* Winston Luo. High school mathematics research project mentor, 2007.

Program in Mathematics for Young Scientists (PROMYS)

· Co-instructor, *Undecidability and Hilbert's 10th Problem*, Advanced seminar for high school students, 2020.

· High school research project designer, 2008, 2011, and 2020.

· High school mathematics research project mentor, 2004–2007.

· Counselor, 2000–2003.

PROFESSIONAL ACTIVITIES

Board of Trustees

· Trustee, *PROMYS Foundation,* a 501(c)(3) organization with the goal of supporting the PROMYS Program at Boston University, 2011 – Present. (Secretary, 2011–2019.)

Steering Committee

· Steering Committee member, *Workshop on Languages for Inference (LAFI)*, formerly known as the Workshop on Probabilistic Programming Semantics (PPS), 2018–Present.

Program Committee Chair

· Program Committee co-chair, *Workshop on Probabilistic Programming Semantics (PPS 2018),* colocated with POPL, Los Angeles, January 9, 2018.

· Program Committee co-chair, *Workshop on Probabilistic Programming Semantics (PPS 2017),* colocated with POPL, Paris, France, January 17, 2017.

PROFESSIONAL ACTIVITIES (CONT'D)

Senior Program Committee / Area Chair

- Area Chair, *Artificial Intelligence and Statistics (AISTATS 2021),* online, April 13–15, 2021.
- Area Chair, *Artificial Intelligence and Statistics (AISTATS 2020),* online, August 26–28, 2020.
- Senior Program Committee member, *Uncertainty in Artificial Intelligence (UAI 2019),* Tel Aviv, Israel, July 22–25, 2019.

Program Committee

- Program Committee member, *Conference on Artificial Intelligence and Theorem Proving (AITP 2018),* Aussois, France, March 25–30, 2018.
- Program Committee member, *AAAI Conference on Artificial Intelligence (AAAI-17),* San Francisco, CA, February 4–9, 2017.
- Program Committee member, *Workshop on Practical Bayesian Nonparametrics (BNP @ NIPS),* Barcelona, Spain, December 9, 2016.
- Program Committee member, *Mathematical Foundations of Programming Semantics XXXII (MFPS 2016),* Pittsburgh, PA, May 23–26, 2016.
- Program Committee member, *Workshop on Probabilistic Programming Semantics (PPS 2016),* colocated with POPL, St. Petersburg, FL., January 23, 2016.
- Program Committee member, *Conference on Computability and Complexity in Analysis (CCA 2013)*, Nancy, France, July 8–10, 2013.

Organizer

- Co-organizer, *Special Session on Logic and Graph Limits*, Association for Symbolic Logic, North American Annual Meeting, University of California, Irvine, March 27, 2020 (cancelled).
- Co-organizer, Harvard–MIT Logic Seminar, 2012–2013.
- Co-organizer, *Special Session on Computability and Complexity*, American Mathematical Society Sectional Meeting, University of Hawai'i at Mānoa, March 3–4, 2012.
- Co-organizer, MIT Logic Seminar, 2008–2010.
- Local Committee, *MAMLS @ Harvard: A meeting on the intersections of logic and mathematics*, May 9–10, 2009.

REFEREEING AND REVIEWING

Journal refereeing

- Annals of Pure and Applied Logic
- Combinatorica
- Computability – The Journal of the Association CiE
- European Journal of Combinatorics
- Information Processing Letters
- Journal of Machine Learning Research
- Journal of Mathematical Logic
- Journal of Symbolic Logic
- Logical Methods in Computer Science
- Memoirs of the AMS
- SIAM Journal on Discrete Mathematics

REFEREEING AND REVIEWING (CONT'D)

Book refereeing

- A K Peters
- MIT Press
- O'Reilly

Conference reviewing

- Artificial Intelligence and Theorem Proving (AITP)
- Association for the Advancement of Artificial Intelligence (AAAI)
- Computability and Complexity in Analysis (CCA)
- Computability in Europe (CiE)
- Foundations of Computer Science (FOCS)
- International Conference on Machine Learning (ICML)
- Mathematical Foundations of Computer Science (MFCS)
- Mathematical Foundations of Programming Semantics (MFPS)
- Probabilistic Programming (PROBPROG)

Workshop reviewing

- Bayesian Nonparametrics (BNP@NeurIPS)
- Probabilistic Programming Semantics (PPS)

EXHIBIT 2

MATERIALS CONSIDERED

Description
Account Table (*representante*) provided by Huron
Invoice Table (*fatura*) provided by Huron
Transfer Table (*transferencia*) provided by Huron
Bonus Table (*bonus*) provided by Huron
Darr v. Dos Santos (In re TelexFree, LLC), 941 F.3d 576, 579 (1st Cir. 2019)
Memorandum of Decision on Class Defendants' Motion to Exclude Expert Witness Testimony of Timothy Martin, dated June 1, 2021
TelexFree Analysis of Damages, dated March 16, 2017
N. Ackerman, C. Freer, and D. Roy. On the computability of conditional probability, Journal of the ACM 66, no. 3, 23:1–23:40, 2019.
F. Saad, N. Ackerman, C. Freer, and V. Mansinghka. A family of exact goodness-of-fit tests for high-dimensional discrete distributions, Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, PMLR 89, 1640–1649, 2019.
D. Cai, N. Ackerman, and C. Freer. Priors on exchangeable directed graphs, Electronic Journal of Statistics, 10, no. 2, 3490–3515, 2016.
Rebuttal Expert Report of Joshua W. Dennis, dated July 31, 2020
Affidavit of Frantz Balan, dated July 29, 2020
Affidavit of Lusette Balan, dated July 29, 2020
Expert Report of Timothy J. Martin, dated April 13, 2018
Reply Expert Report of Timothy J. Martin, dated September 11, 2020
Trustee's Memorandum in Support of the Admissibility of Timothy Martin's Expert Opinion, dated November 18, 2020
Communication with Jean-Louis Sorondo of Huron
Summary Table (*rep_id_summary*) provided by Huron
Memorandum of Decision on Class Defendants' Motion to Exclude Expert Witness Testimony of Timothy Martin, dated June 22, 2021
Field Description Table (*10ksample.xlsx*) provided by Huron
Kabudula, C.W., Clark, B.D., Gómez-Olivé, F.X., Tollman, S., Menken, J. and Reniers, G. The promise of record linkage for assessing the uptake of health services in resource constrained settings: a pilot study from South Africa. BMC Med Res Methodol 14, 71 (2014)
"RFC 2822" (https://datatracker.ietf.org/doc/html/rfc2822)
"Address Verification Service" Wikipedia Page (https://en.wikipedia.org/wiki/Address_verification_service)
L. Getoor and A. Machanavajjhala, Entity Resolution: Tutorial, Advances in Social Network Analysis and Mining, 2012
Scott L DuVall, Richard A Kerber, Alun Thomas "Extending the Fellegi-Sunter probabilistic record linkage method for approximate field comparators", *J Biomed Inform*. 2010; 43(1):24-30.

Description
"FuzzyWuzzy: Fuzzy String Matching in Python" (https://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-in-python/)
RapidFuzz Library (https://github.com/maxbachmann/RapidFuzz)
Pattern Matching: The Gestalt Approach (https://www.drdobbs.com/database/pattern-matching-the-gestalt-approach/184407970?pgno=5)
"difflib" (https://docs.python.org/3/library/difflib.html)
"When to use which fuzz function to compare two strings" (https://stackoverflow.com/questions/31806695/when-to-use-which-fuzz-function-to-compare-2-strings)
Winkler, W. E.: String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. In: Proceedings of the Section on Survey Research. American Statistical Association, pp. 354–359, 1990.
Splink Python package (https://github.com/moj-analytical-services/splink)
Enamorado, T., Fifield, B., and Imai, K. (2019). Using a probabilistic model to assist merging of large-scale administrative records. American Political Science Review, 113(2), 353-371.
Apache Spark Engine (https://spark.apache.org/)
Fellegi, I. P. and Sunter, A. B. (1969). A theory for record linkage. Journal of the American Statistical Association, 64(328), 1183-1210.
Winkler, W. E. Using the EM Algorithm for Weight Computation in the Fellegi-Sunter Model of Record Linkage, American Statistical Association, Proceedings of the Section on Survey Research Methods, 667-671 (1988).
T. Enamorado and K. Imai. Validating Self-Reported Turnout by Linking Public Opinion Surveys with Administrative Records. Public Opinion Quarterly, Volume 83, Issue 4, Winter 2019.
Philip A. Collender, Zhiyue Tom Hu, Charles Li, Qu Cheng, Xintong Li, Yue You, Song Liang, Changhong Yang, and Justin V. Remais. Machine-learning classifiers for logographic name matching in public health applications: approaches for incorporating phonetic, visual, and keystroke similarity in large-scale probabilistic record linkage. arXiv:2001.01895, 2020.
Tom Jackson, Caris Greyson, Ian Rickard and Andromachi Tseloni. Data First: Criminal Courts Linked Data. An exploratory analysis of returning defendants and the potential of linked criminal courts data from 2011 to 2019 in England and Wales. Ministry of Justice Analytical Series, UK Ministry of Justice, 2022.
Pandas Python library (https://pandas.pydata.org/)
Wes McKinney. *Python for Data Analysis*, 2nd Edition, 2017.
Randall, S.M., Ferrante, A.M., Boyd, J.H., Semmens J.B. The effect of data cleaning on record linkage quality. BMC Med Inform Decis Mak 13, 64 (2013)
“North American Numbering Plan” (https://en.wikipedia.org/wiki/North_American_Numbering_Plan)
“Telephone numbers in Brazil” (https://en.wikipedia.org/wiki/Telephone_numbers_in_Brazil)

Description
Dusetzina SB, Tyree S, Meyer AM, Meyer A, Green L, and Carpenter WR. Linking Data for Health Services Research: A Framework and Instructional Guide. US Agency for Healthcare Research and Quality, Rockville, MD, 2014.
J. S. Murray, "Probabilistic Record Linkage and Deduplication after Indexing, Blocking, and Filtering". Journal of Privacy and Confidentiality (2015–2016)
"m and u probabilities" (https://observablehq.com/@robinl/m-and-u-probabilities)
Y. Zhu, Y. Matsuyama, Y. Ohashi, and S. Setoguchi. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. Journal of Biomedical Informatics 56 (2015): 80-86).
Yang, J. and Leskovec, J.. Defining and evaluating network communities based on ground-truth. Knowledge and Information Systems, 42(1): 2015, 181-213.
D. Resnick and J. Asher. Measurement of Type I and Type II Record Linkage Error. Proceedings of the American Statistical Association, Government Statistics Section; Arlington, VA, USA: American Statistical Association; 2019. pp. 293–311.
A. Ferrante and J. Boyd. A transparent and transportable methodology for evaluating Data Linkage software. Journal of Biomedical Informatics 45, no. 1 (2012): 165-172.
Y. Y. Yao. Information-theoretic measures for knowledge discovery and data mining. In Entropy Measures, Maximum Entropy Principle and Emerging Applications, pages 115–136. Karmeshu (ed.), Springer, 2003.
S. Emmons, S. Kobourov, M. Gallant, and K. Börner. Analysis of network clustering algorithms and cluster quality metrics at scale. PloS One, 11(7): (2016).
AggValidatioStep1_13_A.docx provided by Huron
updated testing script jls.sql provided by Huron
01052016 aggregation.sql provided by Huron
Aggregation Validation.docx provided by Huron
Christen, P. (2011). A survey of indexing techniques for scalable record linkage and deduplication. IEEE Transactions on Knowledge and Data Engineering, 24(9), 1537-1555
Parag Singla and Pedro Domingos, Multi-Relational Record Linkage. KDD-2004 Workshop on MultiRelational Data Mining, pp. 31-48
Olivier Binette and Rebecca C. Steorts, (Almost) All of Entity Resolution, arXiv:2008.04443, 2022
P. Christen, *Data Matching*, Data-Centric Systems and Applications, Springer-Verlag, 2012.
Ong TC, Duca LM, Kahn MG, and Crume TL. A hybrid approach to record linkage using a combination of deterministic and probabilistic methodology. J Am Med Inform Assoc. 2020;27(4):505-513
Radford M. Neal and Geoffrey E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. Learning in Graphical Models. Springer, 1998. 355-368
William E. Winkler and Yves Thibaudeau. An application of the Fellegi-Sunter model of record linkage to the 1990 US decennial census. US Bureau of the Census, 1991

Description
"Código de Endereçamento Postal" (https://en.wikipedia.org/wiki/C%C3%B3digo_de_Endere%C3%A7amento_Postal)
UK Office of National Statistics. (2021). Splink: MoJ's open source library for probabilistic record linkage at scale. (https://www.gov.uk/government/publications/joined-up-data-in-government-the-future-of-data-linking-methods/splink-mojs-open-source-library-for-probabilistic-record-linkage-at-scale)
"National identification number" (https://en.wikipedia.org/wiki/National_identification_number)
Hernández, M.A. and Stolfo, S.J. Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem. Data Mining and Knowledge Discovery 2, 9–37 (1998)
Charpentier, Bertrand. Multi-scale clustering in graphs using modularity. Master's thesis, KTH, 2019
X Yan, LGS Jeub, A Flammini, F Radicchi, and S Fortunato. Weight thresholding on complex networks. Physical Review E 98.4 (2018): 042304
Uday Kamath and John Liu. *Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning*. Springer, 2021
Sneath, P.H.A and Sokal, R.R.. *Numerical Taxonomy: the Principles and Practice of Numerical Classification*. Freeman, New York (1973)
"CPF number" (https://en.wikipedia.org/wiki/CPF_number)

EXHIBIT 3

Cluster ID	Total Net Equity	Total Number of User Accounts
132785	$4,087,745.19	471

	rep_id	ag_rep_id	rep_nome	rep_email	rep_cel	rep_fone	rep_end	rep_ numero	rep_cep	rep_login	rep_pwd_secondary	rep_cpf	rep_datanasc	net_equity
1	132785	132785	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	1630 Worcester rd # 626 C	1630	01702	ttmilagres	c99fb769d8bcee3287936bf71335976e	[redacted]		$ 2,580,340.17
2	4471758	4471758	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett rd	2b	01702	ttstars	c99fb769d8bcee3287936bf71335976e		1953-10-15	$ 1,581,837.02
3	11244377	11244377	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett rd	2B	01702	lilimax	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$ 237,712.10
4	10526539	10526539	Maria Neves	zieffj@yahoo.com	5089043810	5089043810	2B Brackett Rd	.	.	ttmilagres01	fd7b8fce445cb01fb358a4ac2b5cb4ab		1953-10-15	$ 85,842.00
5	11277987	11277987	Maria Teresa m Neves	hbmaltaro@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	ttnews1	51990a482be6a725cff5cdb88c070750		1953-10-15	$ 27,174.80
6	162734	162734	Maria T Neves	ttmilagres@gmail.com	5089043810	5089043810	1630 Worcester rd # 626	626	01702	ttmilagres3	e36b0cc8105b559ded106ffb5008533a			$ 24,531.00
7	206326	190471	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	ttmilagres9	c99fb769d8bcee3287936bf71335976e			$ 11,589.00
8	235817	235817	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd	2B	01702	callandsave	c99fb769d8bcee3287936bf71335976e			$ 6,894.80
9	190471	190471	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd	2B	01702	ttmilagres6	c99fb769d8bcee3287936bf71335976e			$ 1,574.60
10	196685	196685	maria neves	mmariashousecleaning@gmail.com	8572367654	8572367654	998 pleasant st	998	01701	prosperity	6b3784d2708615ed49a1ae177b7ad757			$ 399.20
11	8658454	8503060	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	callandsave11	c99fb769d8bcee3287936bf71335976e		1953-10-15	$ 0.00
12	18694481	18693825	Mar ives	ttmilagres@gmail.com	5557862314	5557862314	f		01702	loriloricl4				$ (49.90)
13	11303053	11303053	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett	.	.	lilimax3cl				$ (49.90)
14	3470118	3470118	Maria Teresa M Neves	ttmilagres@gmail.com	.	5089043810	2B Brackett rd	.	01702	ttclient2				$ (49.90)
15	18044647	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews153c				$ (49.90)
16	8670314	8670314	Maria T. M. Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	.	callcl				$ (49.90)
17	17951851	17951851	Maria Neves	anaevangelista2000@hotmail.com	7743860576	7743860576	Beaver st	126 a	01702	1mndaniel2009				$ (49.90)
18	13405463	13405463	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	ttmilagres04cl				$ (49.90)
19	17953997	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews142cl				$ (49.90)
20	17952007	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews142c				$ (49.90)
21	17953397	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews141c				$ (49.90)
22	17969729	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews148c				$ (49.90)
23	18693999	18693825	Mar ives	ttmilagres@gmail.com	5557862314	5557862314	f		01702	loriloricl2				$ (49.90)
24	17969509	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews147c				$ (49.90)
25	27599942	27599942	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	vgjoy3u3ukn1o76				$ (49.90)
26	27599740	27599740	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	0nx63zyha8c31at				$ (49.90)
27	27599736	27599736	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	hrq0asgjnb15bb0				$ (49.90)
28	27599878	27599878	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	cck5ixsvf21pi6a				$ (49.90)
29	27599816	27599816	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	xiacthp3gphh6bc				$ (49.90)
30	18040213	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews144c				$ (49.90)
31	18049149	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews154c				$ (49.90)
32	27599818	27599818	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	0zmv2yqstrc70jy				$ (49.90)
33	27599914	27599914	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	ydwgwf18xw06rvr				$ (49.90)
34	16637951	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews18voip				$ (49.90)
35	27600072	27600072	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	dw11xgg71rpqmx7				$ (49.90)
36	16735401	16735401	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	..	.	.	galonews96cl				$ (49.90)
37	27600056	27600056	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	p24oeotezu66bxw				$ (49.90)
38	27600042	27600042	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	86296agwmj8gc5h				$ (49.90)
39	27599980	27599980	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	lvz0l7gshybujsf				$ (49.90)
40	27599962	27599962	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	sspnmk8a50t69os				$ (49.90)
41	27599944	27599944	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	3gmcej4hcj6570i	ecf7dba3645ffe49b7b30edbcb0dec32			$ (49.90)
42	18066859	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews169c				$ (49.90)
43	16640697	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews56voip				$ (49.90)

44	27599808	27599808	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	nqyi059r853r6lf		[redacted]		$	(49.90)
45	16625101	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews11voip				$	(49.90)
46	16633881	16628681	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	2 B brackett rd	.	01702	galonews41voip				$	(49.90)
47	16783083	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews105cl				$	(49.90)
48	16773575	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews07cl				$	(49.90)
49	16780763	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews13cl				$	(49.90)
50	16773901	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews89cl				$	(49.90)
51	16776559	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews15cl				$	(49.90)
52	16780373	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews112cl				$	(49.90)
53	16732883	16732883	Maria milagres	ttmilagres@gmail.com	5089043810	5089043810	.	.	.	galonews08cl				$	(49.90)
54	10741927	10715815	Teresa neves	teresamilagres@hotmail.com	5089043810	5089043810	.	.	.	ttcliente13				$	(49.90)
55	10741803	10715815	Teresa neves	teresamilagres@hotmail.com	5089043810	5089043810	.	.	.	ttcliente12				$	(49.90)
56	10526503	10526503	maria Teresa Milagres	ttmilagres@gmail.com	5089043810	5089043810	.	.	.	ttcliente9				$	(49.90)
57	11501403	11501403	maria lares	m.adame43@yahoo.com	9132599771	9132599771	2101 w fuller ave	2101	76133	mary999				$	(49.90)
58	16337571	16337571	maria neves	mmariashousecleaning@gmail.com	8572367654	8572367654	155 pleasant st	155	01752	prosperityvoip				$	(49.90)
59	17966355	17962023	maria neves	ttmilagres@gmail.com	5089043810	5089043810	268 a windaham rd	268	bh14qu	galonews183c				$	(49.90)
60	17237597	13405463	maria neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	ttmilagres00cl				$	(49.90)
61	16631955	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews38voip				$	(49.90)
62	16644253	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews63voip				$	(49.90)
63	16622511	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews6voip				$	(49.90)
64	16629235	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews30voip				$	(49.90)
65	16646383	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews71voip				$	(49.90)
66	16634173	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.....	01702	galonews47voip				$	(49.90)
67	16631271	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews37voip				$	(49.90)
68	16626347	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews21voip				$	(49.90)
69	16641989	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews59voip				$	(49.90)
70	16640355	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews55voip				$	(49.90)
71	16626785	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews22voip				$	(49.90)
72	16621363	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews7				$	(49.90)
73	16639051	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews14voip				$	(49.90)
74	16628681	16628681	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	2 B brackett rd	.	01702	galonews25voip				$	(49.90)
75	16630149	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews12voip				$	(49.90)
76	16642523	16621363	Maria Teresa Milagres Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	galonews60voip				$	(49.90)
77	16644521	16644521	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	galonews68voip				$	(49.90)
78	16646637	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews76voip				$	(49.90)
79	27599762	27599762	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	02b brackett rd	2	01702	8lrjcvemfzsmpo7				$	(49.90)
80	15376207	15376207	rita lares	calderonr27@yahoo.com	8178455929	8176924229	4409 merida st	4409	76133	1piruleto				$	(49.90)
81	19161311	19159783	Mara ves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		01702	ttstars0cl7				$	(49.90)
82	19160801	19159783	Mara ves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		01702	ttstars0cl6				$	(49.90)
83	19160197	19159783	Mara ves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		01702	ttstars0cl3				$	(49.90)
84	18990527	18990383	Mara ives	ttmilagres@gmail.com	7787862314	4447862314	2 bhgh rd		01702	conectingcl3				$	(49.90)
85	18990647	18990383	Mara ives	ttmilagres@gmail.com	7787862314	4447862314	2 bhgh rd		01702	conectingcl				$	(49.90)
86	19161765	19159783	Mara ves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		01702	ttstars0cl10				$	(49.90)
87	19159783	19159783	Mara ves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		01702	ttstars0cl1				$	(49.90)
88	16620383	16620383	Maria Neves	hbmaltaro@gmail.com	5089043810	5089043810	hbmaltaro@gmail.com	.	.	galonews4client				$	(49.90)
89	19161623	19159783	Mara ves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		01702	ttstars0cl9				$	(49.90)
90	18990959	18990383	Mara ives	ttmilagres@gmail.com	7787862314	4447862314	2 bhgh rd		01702	conectingcl5				$	(49.90)
91	18990811	18990383	Mara ives	ttmilagres@gmail.com	7787862314	4447862314	2 bhgh rd		01702	conectingcl4				$	(49.90)
92	19159877	19159783	Mara ves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		01702	ttstars0cl2				$	(49.90)

93	18990383	18990383	Mara ives	ttmilagres@gmail.com	7787862314	4447862314	2 bhgh rd		01702	conectingcl2		[redacted]		$	(49.90)
94	19161461	19159783	Mara ves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		01702	ttstars0cl8				$	(49.90)
95	16438329	14057783	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	thth7cliente				$	(49.90)
96	18052713	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews184c				$	(49.90)
97	19160613	19159783	Mara ves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		01702	ttstars0cl5				$	(49.90)
98	18694363	18693825	Mar ives	ttmilagres@gmail.com	5557862314	5557862314	f		01702	loriloricl3				$	(49.90)
99	19160413	19159783	Mara ves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		01702	ttstars0cl4				$	(49.90)
100	18069637	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews170c				$	(49.90)
101	18990151	18990151	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	2 bhgh rd		01702	conectingcl1				$	(49.90)
102	18693825	18693825	Mar ives	ttmilagres@gmail.com	5557862314	5557862314	f		01702	loriloricl1				$	(49.90)
103	10798821	10798821	Maria Teresa Milagres Neves	sullyg24@gmail.com	5089043810	5089043810	2b Brackett rd	2b	01702	ttstarss	2f11def4b449903f0c287715ddbe17fb		1953-10-15	$	(50.00)
104	10528279	10528279	Maria T. M. Neves	andradesrv@hotmail.com	5089043810	5089043810	.	.	.	ttstars1	54effaca3d18b37e7b3732525efa0b62		1953-10-15	$	(50.00)
105	11149467	11149467	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2b brackett Rd,	2	01702	simplemoney	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(50.00)
106	14618193	14618193	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	ttmilagres06	e489d262de4bcf401a017219dcfaf958		1953-10-15	$	(50.00)
107	14787301	14786381	Maria Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	lilimax23voip				$	(99.80)
108	14004009	14004009	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	htht62cl				$	(99.80)
109	14792677	14786381	Maria Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	lilimax35voip				$	(99.80)
110	14791361	14786381	Maria Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	lilimax34voip				$	(99.80)
111	10495711	10495711	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	..	ttstarsclient				$	(99.80)
112	14057783	14057783	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	thth16cl				$	(99.80)
113	14789721	14786381	Maria Neves	ttmilagres@gmail.com	.	5089043810	.	.	01702	lilimax28voip				$	(99.80)
114	12467711	10495711	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	.	ttstarscl4				$	(99.80)
115	296483	296413	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brakett rd	2	01702	ttcliente2				$	(99.80)
116	2261362	2261362	Maria Teresa Neves	teresamilagres@hotmail.com	5089043810	5089043810	2B Brackett Rd	2B	01702	ttnevescliente				$	(99.80)
117	2060788	2060788	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2	01702	ttcliente3				$	(99.80)
118	18679463	18679463	Maria Teresa Neves	ttmilagres@gmail.com	4447862314	4447862314	02b bracketrd		01702	ttmilgres0cl1				$	(99.80)
119	14008259	14007419	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2b brackett Rd	2	01702	thth7cl				$	(99.80)
120	14574063	14574063	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	.	callandsave6cl				$	(99.80)
121	11599805	11599805	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	.	skysun4cliente				$	(99.80)
122	14786381	14786381	Maria Neves	ttmilagres@gmail.com	.	5089043810	..	.	01702	lilimax16voip				$	(99.80)
123	11766001	11766001	Marai Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2 b Brackett rd	2	01702	ttstarscl03				$	(99.80)
124	18679979	18679827	Mara Tes Nev	ttmilagres@gmail.com	4447862314	4447862314	02b bracketrd		01702	ttmilgres0cl3				$	(99.80)
125	18679827	18679827	Mara Tes Nev	ttmilagres@gmail.com	4447862314	4447862314	02b bracketrd		01702	ttmilgres0cl2				$	(99.80)
126	14788601	14788601	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	lilimax6voip				$	(99.80)
127	17237291	13405463	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	ttmilagres0cl				$	(99.80)
128	4473096	4473096	Maria Teresa M Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	.	01702	ttstarscliente				$	(149.70)
129	4426442	4425960	Maria T. M. Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	.	ttclientes5				$	(149.70)
130	10434967	10434967	Maria T. Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	.	callandsavecl				$	(149.70)
131	4425960	4425960	Maria T. M. Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	.	ttclientes2				$	(149.70)
132	13029217	12913029	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett rd	2	01702	ttmilagres02cl				$	(199.60)
133	10715741	10715285	Teresa neves	teresamilagres@hotmail.com	5089043810	5089043810	.	.	.	ttstarscl3				$	(199.60)
134	2523888	2523888	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	ttcliente4				$	(199.60)
135	11247261	11247261	Maria Teresa M Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	lilimaxcl				$	(299.40)
136	18629015	18629015	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett rd	2	01702	ttstars0	c99fb769d8bcee3287936bf71335976e		1956-09-16	$	(339.60)
137	29035450	28988768	Maria milagres	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	conecting3	c99fb769d8bcee3287936bf71335976e		1965-12-15	$	(339.60)
138	28988768	28988768	Maria milagres	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	conecting2	c99fb769d8bcee3287936bf71335976e		1065-12-15	$	(339.60)
139	280593	280593	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brakett rd	2	01702	callcliente				$	(349.30)
140	10715815	10715815	Teresa neves	teresamilagres@hotmail.com	5089043810	5089043810	.	.	.	ttcliente10				$	(349.30)
141	10741745	10715815	Teresa neves	teresamilagres@hotmail.com	5089043810	5089043810	.	.	.	ttcliente11				$	(349.30)

142	296413	296413	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brakett rd	2	01702	ttcliente1		[redacted]		$	(349.30)
143	10678285	10678285	Teresa m. Neves	teresamilagres@hotmail.com	5089043810	5089043810	.	.	.	ttstarscl				$	(399.20)
144	10715285	10715285	Teresa neves	teresamilagres@hotmail.com	5089043810	5089043810	..	.	.	ttstarscl1				$	(399.20)
145	10715653	10715285	Teresa neves	teresamilagres@hotmail.com	5089043810	5089043810	.	.	.	ttstarscl2				$	(399.20)
146	10450919	10450919	Maria T. Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	.	ttcliente8				$	(399.20)
147	10347917	10347917	Maria T. Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	ttcliente5				$	(399.20)
148	10450569	10450569	Maria T. Neves	ttmilagres@gmail.com	5089043810	5089043810	ttmilagres@gmail.com	.	.	ttcliente7				$	(449.10)
149	10337975	10337975	Maria T. M. Neves	ttmilagres@gmail.com	5089043810	5089043810	.	.	01702	ttcliente6				$	(449.10)
150	132790	132790	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	1630 Worcester rd # 626 C	1630	36570	ttmilagres1	c99fb769d8bcee3287936bf71335976e			$	(677.20)
151	12913483	12913029	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett Rd	2	01702	ttmilagres02	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,086.00)
152	8594574	8503060	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	callandsave3	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,225.40)
153	8503060	8503060	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	callandsave2	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,225.40)
154	190135	190135	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd	00000	01702	ttmilagres5	c99fb769d8bcee3287936bf71335976e			$	(1,275.20)
155	8649192	8503060	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	callandsave4	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,275.30)
156	198875	190471	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	ttmilagres7	c99fb769d8bcee3287936bf71335976e			$	(1,295.20)
157	198981	190471	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	ttmilagres8	c99fb769d8bcee3287936bf71335976e			$	(1,305.30)
158	133361	133361	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	1630 Worcester rd # 626 C	626	01702	ttmilagres2	c99fb769d8bcee3287936bf71335976e			$	(1,325.10)
159	8651026	8503060	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	callandsave5	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,325.20)
160	1902924	1902924	Maria Neves	mmariashousecleaning@gmail.com	8572367654	8572367654	155 Pleasent st # B9	B9	01752	prosperity2	d310469b4398208a9737689a4ffc204a			$	(1,375.00)
161	181393	181393	MariaTeresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd	0000	01702	ttmilagres4	c99fb769d8bcee3287936bf71335976e			$	(1,375.00)
162	8653610	8503060	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	callandsave6	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,375.10)
163	11246147	11245137	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	lilimax9	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,375.10)
164	13997415	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,375.10)
165	16640849	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews66	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
166	16642597	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews73	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
167	18069855	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews175	2ce4b2e3efcc473a6ed2d70b01c4d7e9			$	(1,425.00)
168	16627303	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews30	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
169	18053979	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews196	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
170	16640913	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews67	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
171	14787889	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax27	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
172	14790219	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax33	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
173	16644599	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews79	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
174	14787633	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax25	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
175	16642681	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews74	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
176	16644435	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews78	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
177	16640991	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews68	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
178	16642753	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews75	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
179	16224593	16224593	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd		01702	thth29	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
180	16645827	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews80	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
181	16626439	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews23	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
182	16630641	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews38	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
183	16630403	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews36	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
184	16630273	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews35	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
185	16629427	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews34	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
186	16629357	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews33	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
187	16640621	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews64	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
188	16642847	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews76	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
189	16638957	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews56	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
190	16639135	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews57	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)

191	16639317	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews59	2ce4b2e3efcc473a6ed2d70b01c4d7e9	[redacted]	1953-10-15	$	(1,425.00)
192	16639225	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews58	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
193	16634075	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews49	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
194	16640555	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews63	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
195	16640493	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	..	01702	galonews62	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
196	16640425	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews61	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
197	16639397	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews60	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
198	16638775	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews55	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
199	16638679	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews54	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
200	16630567	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews37	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
201	16631457	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews39	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
202	18053585	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews193	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
203	16631555	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews40	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
204	18049761	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews163	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
205	18070019	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews176	2ce4b2e3efcc473a6ed2d70b01c4d7e9			$	(1,425.00)
206	18070913	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews180	2ce4b2e3efcc473a6ed2d70b01c4d7e9			$	(1,425.00)
207	16626509	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews24	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
208	16621589	16621589	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd	.	01702	galonews07	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
209	16625989	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews22	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
210	14005767	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth9	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
211	16625875	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews21	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
212	16634831	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews52	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
213	16634433	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews51	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
214	16634311	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews50	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
215	16640765	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews65	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
216	16632323	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews46	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
217	16632151	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews44	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
218	16632081	16621589	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd	.	01702	galonews43	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
219	16631769	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews42	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
220	16631679	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews41	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
221	16642367	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews72	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
222	14793097	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax41	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
223	14791613	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax37	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
224	16646549	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews82	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
225	16782407	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews108	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
226	16782041	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews107	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
227	16919369	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews128	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
228	16778925	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews115	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
229	16776937	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews113	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
230	16782593	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews110	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
231	16781959	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews106	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
232	16780961	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews105	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
233	8657066	8503060	Maria Teresa Neves	ttmilagres@gmail.com	95089043810	5089043810	2B Brackett rd	2B	01702	callandsave8	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,425.00)
234	16919563	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews131	8f1aed01410875d49cb2e6f6c48db787		1953-10-15	$	(1,425.00)
235	16919493	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews130	8f1aed01410875d49cb2e6f6c48db787		1953-10-15	$	(1,425.00)
236	16919423	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews129	8f1aed01410875d49cb2e6f6c48db787		1953-10-15	$	(1,425.00)
237	16774669	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews125	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
238	16774267	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews120	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
239	16776853	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews112	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)

240	16782511	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews109	2ce4b2e3efcc473a6ed2d70b01c4d7e9	[redacted]	1953-10-15	$	(1,425.00)
241	16919151	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews127	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
242	16782635	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews111	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
243	17967281	17962023	maria neves	ttmilagres@gmail.com	5089043810	5089043810	268 a windaham rd	268	bh14qu	galonews185	81dc9bdb52d04dc20036dbd8313ed055		1953-10-15	$	(1,425.00)
244	17968095	17962023	maria neves	ttmilagres@gmail.com	5089043810	5089043810	268 a windaham rd	268	bh14qu	galonews186	81dc9bdb52d04dc20036dbd8313ed055		1953-10-15	$	(1,425.00)
245	17969093	17962023	maria neves	ttmilagres@gmail.com	5089043810	5089043810	268 a windaham rd	268	bh14qu	galonews188	81dc9bdb52d04dc20036dbd8313ed055		1953-10-15	$	(1,425.00)
246	17968609	17962023	maria neves	ttmilagres@gmail.com	5089043810	5089043810	268 a windaham rd	268	bh14qu	galonews187	81dc9bdb52d04dc20036dbd8313ed055		1953-10-15	$	(1,425.00)
247	17969533	17962023	maria neves	ttmilagres@gmail.com	5089043810	5089043810	268 a windaham rd	268	bh14qu	galonews189	81dc9bdb52d04dc20036dbd8313ed055		1953-10-15	$	(1,425.00)
248	17962023	17962023	maria neves	ttmilagres@gmail.com	5089043810	5089043810	268 a windaham rd	268	bh14qu	galonews183	81dc9bdb52d04dc20036dbd8313ed055		1953-10-15	$	(1,425.00)
249	11245307	11245307	Mara Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2	01702	lilimax4	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
250	17969923	17962023	maria neves	ttmilagres@gmail.com	5089043810	5089043810	268 a windaham rd	268	bh14qu	galonews190	81dc9bdb52d04dc20036dbd8313ed055		1953-10-15	$	(1,425.00)
251	16774561	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews123	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
252	17965087	17962023	maria neves	ttmilagres@gmail.com	5089043810	5089043810	268 a windaham rd	268	bh14qu	galonews182	81dc9bdb52d04dc20036dbd8313ed055		1953-10-15	$	(1,425.00)
253	4467402	4467402	maria Teresa Milagres neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett rd	2b	01702	ttnews	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,425.00)
254	16632537	16632537	Maria Teresa` Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews47	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
255	11599085	11599085	Maria Teresa m Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett rd	2	01702	ttnews26	88bfc20129245733083b323adf88845a		1953-10-15	$	(1,425.00)
256	16619321	16619097	Maria Teresa m Neves	ttmilagres@gmail.com	5089043810	5089043810	2b brackett rd,	2	01702	galonews6	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
257	16619097	16619097	Maria Teresa m Neves	ttmilagres@gmail.com	5089043810	5089043810	2b brackett rd,	2	01702	galonews5	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
258	16774625	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews124	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
259	16779121	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews117	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
260	16779195	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews118	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
261	16626707	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews26	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
262	16642203	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews71	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
263	8657580	8503060	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	callandsave9	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,425.00)
264	12913029	12913029	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett Rd	2	01702	ttmilagre	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,425.00)
265	11244775	11244775	Maria Teresa Milagres neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett rd	.	01702	lilimax1	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
266	16622845	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews15	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
267	16626593	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews25	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
268	16628873	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews32	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
269	16919619	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews132	8f1aed01410875d49cb2e6f6c48db787		1953-10-15	$	(1,425.00)
270	16646463	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews81	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
271	16647067	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews83	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
272	16646723	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews84	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
273	18070767	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews179	2ce4b2e3efcc473a6ed2d70b01c4d7e9			$	(1,425.00)
274	16642069	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews69	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
275	16642125	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews70	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
276	8656054	8503060	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	callandsave7	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,425.00)
277	16779081	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews116	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
278	16779243	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews119	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
279	12972683	12913029	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett Rd	2	01702	ttmilagres03	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,425.00)
280	16774465	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews122	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
281	16780933	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews104	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
282	16778859	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews114	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
283	16774345	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews121	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
284	12972911	12972911	Maria Teresa Neves	ttmilagres@gmail.com	50890438100	50890438100	2 B Brackett rd	2	01702	ttmilagres04	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,425.00)
285	16919105	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews126	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
286	16920413	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews139	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
287	8658008	8503060	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	callandsave10	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,425.00)
288	16920343	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews138	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)

289	16920287	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews137	2ce4b2e3efcc473a6ed2d70b01c4d7e9	[redacted]	1953-10-15	$	(1,425.00)
290	16920249	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews136	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
291	16920191	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews135	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
292	16920141	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews134	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
293	16919873	16773575	Maria Teresa Neves	ttmilagres@gmail.com	5089043810	5089043810	02 b Brackett Rd		01702	galonews133	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
294	16621755	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews08	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
295	11458817	11458613	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	.	01702	ttnews19	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
296	16644345	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews77	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
297	16622023	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.....	01702	galonews10	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
298	17951611	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews142	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
299	14007419	14007419	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett rd	2	01702	thth15	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
300	14006589	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth14	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
301	17955709	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews143	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-13	$	(1,425.00)
302	18042489	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews154	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
303	17956685	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews144	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-13	$	(1,425.00)
304	18051837	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews184	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
305	17963637	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews146	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
306	17963863	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews147	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
307	17967931	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews148	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
308	18041747	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews149	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
309	14056517	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth20	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
310	14056323	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth17	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
311	18053875	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews195	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
312	18070545	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews178	2ce4b2e3efcc473a6ed2d70b01c4d7e9			$	(1,425.00)
313	16708453	16708453	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	galonews88	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
314	14787753	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax26	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
315	18067585	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews174	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
316	18067483	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews173	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
317	17946195	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews141	81dc9bdb52d04dc20036dbd8313ed055		1953-10-15	$	(1,425.00)
318	18042023	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews151	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
319	18045529	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews159	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
320	14006037	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth11	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
321	13833805	13833805	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Bracket Rd	2B	01702	telexusaworld1	bb50b2c078929bd053d47b7103fa4f91		1953-10-15	$	(1,425.00)
322	14003525	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth8	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
323	14056541	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth21	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
324	16225057	16224593	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd		01702	thth34	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
325	14006539	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth13	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
326	14786847	14786847	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2 B	01702	lilimax23	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
327	14056475	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth19	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
328	16225129	16224593	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd		01702	thth35	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
329	14056719	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth24	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
330	18044849	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews155	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
331	14056827	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth27	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
332	14056683	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth23	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
333	14056937	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth28	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
334	14056599	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth22	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
335	16225021	16224593	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd		01702	thth33	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
336	14006247	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth12	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
337	18062923	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews165	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)

338	14007511	14007419	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett rd	2	01702	thth16	cfd12d74bca9357022eb7d8367bcab26	[redacted]	1953-10-15	$	(1,425.00)
339	18067375	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews172	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
340	18067073	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews171	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
341	18065865	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews170	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
342	18042375	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews153	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
343	14790893	14789917	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2	01702	lilimax34	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
344	18045005	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews156	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
345	18045111	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews157	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
346	18049887	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews164	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
347	18049635	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews162	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
348	18049463	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews161	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
349	14790053	14789917	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2	01702	lilimax32	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
350	18045631	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews160	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
351	18045279	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews158	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
352	16732631	16708453	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	galonews96	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
353	14789975	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax31	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
354	14789917	14789917	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2	01702	lilimax30	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
355	16732683	16708453	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	galonews97	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
356	14789165	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax29	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
357	14789003	14789003	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2 B	01702	lilimax28	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
358	14791763	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax39	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
359	16224641	16224593	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd		01702	thth30	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
360	18064021	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews167	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
361	14790991	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax35	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
362	16708565	16708453	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	galonews89	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
363	18064249	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews169	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
364	14793319	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax43	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
365	18064147	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews168	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
366	18053331	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews192	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
367	18051957	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews181	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
368	18041863	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews150	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
369	18042185	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews152	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
370	18063861	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews166	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
371	16734193	16708453	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	galonews98	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
372	14005989	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth10	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
373	18061061	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews0191	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
374	16734953	16708453	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	galonews101	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
375	16225177	16224593	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd		01702	thth36	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
376	14792977	14789917	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2	01702	lilimax40	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
377	14791705	14789917	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2	01702	lilimax38	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
378	16734277	16708453	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	galonews99	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
379	14791539	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax36	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
380	16734349	16708453	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	galonews100	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
381	16734471	16708453	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	galonews102	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
382	16734537	16708453	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	galonews103	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
383	13998753	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth6	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
384	14003267	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth7	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
385	13998693	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth5	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
386	11245531	11245531	Maria Teresa Milagres Neves	ttmilagres@gmail.com	50890438105	5089043810	2B Brackett rd	2B	01702	lilimax6	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)

387	11458835	11458613	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	.	01702	ttnews20	cfd12d74bca9357022eb7d8367bcab26	[redacted]	1953-10-15	$	(1,425.00)
388	11458871	11458613	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	.	01702	ttnews21	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
389	11458917	11458613	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	.	01702	ttnews22	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
390	11458933	11458613	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	.	01702	ttnews23	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
391	11458953	11458613	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	.	01702	ttnews24	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
392	11459027	11458613	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	.	01702	ttnews25	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
393	18070263	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews177	2ce4b2e3efcc473a6ed2d70b01c4d7e9			$	(1,425.00)
394	11245415	11245137	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	lilimax5	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
395	11866855	11866855	Maria Teresa Milagres Neves	hbmaltaro@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	ttnews31	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
396	11246255	11245137	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	lilimax10	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
397	2239822	2239822	Maria Teresa Milagres Neves	teresamilagres@hotmail.com	5089043810	5089043810	2B Brackett rd	2	01702	ttneves	c25a4a343798d77ac374ce05e6e3649b		1953-10-15	$	(1,425.00)
398	18053751	17946195	Maria Neves	ttmilagres@gmail.com	4447862314	4447862314	268 windham rd		bh14qu	galonews194	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
399	16224667	16224593	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd		01702	thth31	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
400	11867071	11866855	Maria Teresa Milagres Neves	hbmaltaro@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	ttnews33	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
401	14793225	14787633	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2	01702	lilimax42	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
402	15143621	15143621	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	ttmilagres07	9e4c97a7933a162ecc44b7beb298ace6		1953-10-15	$	(1,425.00)
403	11867017	11866855	Maria Teresa Milagres Neves	hbmaltaro@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	ttnews32	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
404	8502674	235817	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	callandsave1	1139841c9dede68a9d851bfbc512a4bd		1953-10-15	$	(1,425.00)
405	11458767	11458613	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	.	01702	ttnews18	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
406	11458747	11458613	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	.	01702	ttnews17	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
407	13998527	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth4	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
408	16622975	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews16	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
409	16623059	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews17	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
410	16622131	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews11	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
411	16622269	16621589	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd	.	01702	galonews12	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
412	16622641	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews13	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
413	16626895	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews27	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
414	16627189	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd		01702	galonews29	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
415	16627073	16621589	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd	.	01702	galonews28	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
416	16622737	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews14	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
417	16623149	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews18	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
418	16224699	16224593	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd		01702	thth32	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
419	16628755	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews31	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
420	16625215	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews19	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
421	16625357	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews20	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
422	16621887	16621755	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett Rd	.	01702	galonews09	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
423	11245659	11245137	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	lilimax7	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
424	11245137	11245137	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	lilimax3	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
425	11245995	11245137	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	lilimax8	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
426	17949747	17949747	Maria Neves	Anaevangelista2000@hotmail.com	7743860576	7743860576	Beaver st	126 a	01702	mndaniel2009	4733cdf81f59dc9589616b89e5e8a1c1		1977-02-14	$	(1,425.00)
427	11867267	11866855	Maria Teresa Milagres Neves	hbmaltaro@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	ttnews35	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
428	11867141	11866855	Maria Teresa Milagres Neves	hbmaltaro@gmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	ttnews34	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
429	11458613	11458613	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett rd	.	01702	ttnews16	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
430	11599497	11599389	Maria T M Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett rd	2	01702	ttnews28	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
431	17235927	17235927	Maria T Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd	2	01702	ttmilagres00	c99fb769d8bcee3287936bf71335976e		1953-10-23	$	(1,425.00)
432	11278205	11278205	Maria T M neves	hbmaltaro@gmail.com	5089043810	5089043810	2b brackett rd	.	01702	ttnews3	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
433	11278539	11278539	Maria T M Neves	hbmaltaro@gmail.com	5088085532	5088085532	2b Brackett Rd	.	01702	ttnews6	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
434	11278563	11278205	Maria T M Neves	hbmaltaro@gmail.com	5089043810	5089043810	2b Brackett rd	.	01702	ttnews7	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
435	11599565	11599565	Maria T M Neves	hbmaltaro@gmail.com	5089043810	5089043810	2b Brackett rd	2	01702	ttnews29	95cae36133785b81a4dd0f51b40d1b7b		1953-10-15	$	(1,425.00)

436	11926571	11925697	Maria T M Neves	teresamilagres@hotmail.com	5089043810	5089043810	2b Brackett rd	.	01702	lilimax15	ba168c2e6d5fa0cc0b72c48454335a67	[redacted]	1953-10-15	$	(1,425.00)
437	11278247	11278247	Maria T M Neves	hbmaltaro@gmail.com	5089043810	5089043810	2b Bracket rd	.	01702	ttnews4	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
438	11925697	11925697	Maria T M Neves	teresamilagres@hotmail.com	5089043810	5089043810	2b Brackett rd	.	01702	lilimax14	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
439	11278609	11278205	Maria T M Neves	hbmaltaro@gmail.com	5089043810	5089043810	2b brackett rd	.	01702	ttnews9	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
440	11278037	11278037	Maria T. M Neves	hbmaltaro@gmail.com	5089043810	5089043810	2b Brackett rd	.	01702	ttnews2	a2eb8060261955661 22a76e94219759f		1953-10-15	$	(1,425.00)
441	11599389	11599389	Maria T M Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett rd	2	01702	ttnews27	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
442	11278587	11278205	Maria T M Neves	hbmaltaro@gmail.com	5089043810	5089043810	2b Brackett rd	.	01702	ttnews8	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
443	11599763	11599389	Maria T M Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett rd	2	01702	ttnews30	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
444	11278655	11278205	Maria T M Neves	hbmaltaro@gmail.com	5089043810	5089043810	2b Brackett rd	.	01702	ttnews10	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
445	11278295	11278205	Maria T M Neves	hbmaltaro@gmail.com	5089043810	5089043810	2b Brackett rd	.	01702	ttnews5	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
446	13998145	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth1	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
447	13998187	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth2	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
448	13998239	13997415	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	2	01702	thth3	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
449	14786759	14786759	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 b Bracket Rd.	2 B	01702	lilimax22	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
450	14056797	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth26	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
451	14787529	14787529	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	2b	01702	lilimax24	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
452	14056761	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth25	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
453	16638275	16638275	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2B Brackett Rd.	.	01702	galonews53	2ce4b2e3efcc473a6ed2d70b01c4d7e9		1953-10-15	$	(1,425.00)
454	11308793	11308607	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett rd	2	01702	ttnews12	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
455	12472939	12472939	Maria Teresa M. Neves	telexfreeindependent@gmail.com	5089043810	5089043810	2 B Brackett Rd	2	01702	lilimax18	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
456	12473507	12472939	Maria Teresa M. Neves	telexfreeindependent@gmail.com	5089043810	5089043810	2 B Brackett Rd	2	01702	lilimax19	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
457	12419111	12419111	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett rd	2	01702	lilimax16	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
458	11308863	11308607	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett rd	2	01702	ttnews13	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
459	11309101	11308607	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett rd	2	01702	ttnews15	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
460	12473729	12472939	Maria Teresa M. Neves	telexfreeindependent@gmail.com	5089043810	5089043810	2 B Brackett Rd	2	01702	lilimax20	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
461	11308607	11308607	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett rd	2	01702	ttnews11	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
462	12419299	12419111	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett rd	2	01702	lilimax17	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
463	11308937	11308607	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett rd	2	01702	ttnews14	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
464	11301795	11301631	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett	2 B	01702	lilimax12	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
465	11301855	11301855	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett Rd	2 B	01702	lilimax13	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
466	12473919	12472939	Maria Teresa M. Neves	telexfreeindependent@gmail.com	5089043810	5089043810	2 B Brackett Rd	2	01702	lilimax21	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
467	14056375	14006539	Maria Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brakett Rd	.	01702	thth18	cfd12d74bca9357022eb7d8367bcab26		1953-10-15	$	(1,425.00)
468	11301631	11301631	Maria Teresa M. Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B Brackett	2 B	01702	lilimax11	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,425.00)
469	17235929	17235929	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2b Brackett rd	.	01702	ttmilagres0	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,425.00)
470	11244877	11244775	Maria Teresa Milagres Neves	ttmilagres@gmail.com	5089043810	5089043810	2 B brackett rd	.	01702	lilimax2	ba168c2e6d5fa0cc0b72c48454335a67		1953-10-15	$	(1,495.00)
471	2240074	2240074	Maria Teresa Milagres Neves	teresamilagres@hotmail.com	5089043810	5089043810	2B Brackett rd	2B	01702	ttneves1	c99fb769d8bcee3287936bf71335976e		1953-10-15	$	(1,495.00)

EXHIBIT 4

rep_end gamma cutoff selection

#	clean_rep_end_l	clean_rep_end_r	end_jwsim
1	137a winthrop st apt11a	137a winthrop st apt 11a	99.17
2	calle juan buendía 758 urb. patazca	calle juan buendia 758 urb. patazca	98.86
3	3 swans pond ave	3 swains pond ave	98.82
4	barrio de la cruz 20 3d	barrio de la cruz20 3d	98.22
5	shenyagn	shenyang	97.50
6	454 meridian st apt #2	454 meridian st #2	96.36
7	803 sw morrison st ste b2	803 sw morrison st 2b	95.85
8	jr. los nogales 230	jr. los nogales	95.79
9	87 - 48 de 114 street, 2do piso - richmond hill	87 - 48 de 114 street 2 piso - richmond hill	94.63
10	flat270 norwood road	flat 270 norwood road	94.05
11	78 neil st	78 neil	94.00
12	calle 21 sur 22-63	clle 21 sur # 22 63	93.41
13	arzovispo xelmirez	alzobispo xelmirez	93.33
14	pasaje ibiza, 2265	pasaje ibiza	93.33
15	rua x3	rua x4	93.33
16	6532 46th st	6532 6th street	93.00
17	caminho velho da igreja	caminho velho da igreja n°16 garachico	92.11
18	149w selden st	149 w selden st	91.78
19	#29h	29	91.11
20	av sampaio	av. sampaio, n 668	89.63
21	rm 803,8/f,wah oi hse,wah kwai est,aberdeen,h.k.	rm 803, wah oi house, wah kwai est,	89.32
22	14812 sw 17 sreet	14812 s.w 171 terrace	89.06
23	r. 1	r. a	88.33
24	r. 1	r. a	88.33
25	rua da longueira,51 moradia assomada park	rua da longueira	87.80
26	chican y yangana	chican	87.50
27	av. aviacion 415 - 2do. piso 2204	av.aviacíon 415 2do.piso 2204	87.33
28	caminho	caminho da achada 107	86.67
29	nangangqu xuefuli68hao 701shi	nangang	84.83
30	roundtree blvd	36 roundtree blvd	84.59
31	rua anadia	r anadia	82.75
32	av chone entrada pita km 34	avenida chone km 34	80.82
33	1209	127	77.78
34	av. manco capac	jr. manco inca	76.03
35	an hui	rang hu lu qu	73.93
36	calle 104 # 25a -11	10 calle 18 av b local 101	71.50
37	mz 64 lte 04 cte16	ciudad nueva mz 64 lt 05 cte 16	71.33
38	nangang	heilongjiang	71.03
39	calle c no 12 vista el valle	calle d	69.05
40	estrada dr joao abel freitas	estrada francisco alvares nobrega , edificio park , lote 6 bloco 1 fraçao e	68.83
41	henan	nangang	67.62
42	cra 70 n 80-26	cll 20c 76 07 apto 301	65.63
43	calle 32 no 4m 33	calle san francisco de asis	65.45
44	rua drº alfredo ferreira júnior n.º 34	rua da escola	65.13
45	44 cypress st apt 2	spruce street	64.79
46	changchun	wuhan	64.44
47	diagonal 18n 25ai 08	carrera 89 c 68 a 105	63.07
48	la brida	cll 50b sur # 89a 51	62.50
49	cr 41-c # 11-35	cll 132 a no 89-51 bloque 6 casa 14	62.19
50	nangang	guangxi	61.90
51	daoliqu shanghailu6hao 407shi	zhan lan guan xi lu 907 hao	61.43
52	mario avila	manzana92	60.27

53	beijing	shenyang	60.12
54	angangn	ranghulu	60.12
55	rua antonio gomes filho	rua beirute	59.93
56	changchun	shenyang	59.35
57	shenyang	changchun	59.35
58	changchun	shenyang	59.35
59	rua nova esperança	caminho das courelas 40 a	59.28
60	dongcheng	neimeng	58.86
61	changchun	guangxi	58.73
62	ps caravely	los rosales	58.59
63	santiago 2	angel aguirre	57.61
64	changchun	shen	57.41
65	shen	changchun	57.41
66	kouqian	huludaoshilianshanquhaiyuehuayuan	56.85
67	caminho do terã§o	vivien st	55.99
68	rua santo antonio	calle 1ra no.973 chicago apto. 4 piso 1 edificio f & j	55.94
69	75 n martin st	sandralopes@hotmail.co.uk	55.65
70	carrera 71 b bis numero 12-30	col primavera	55.64
71	calle guayaquil	las palmas	55.56
72	123	213	55.56
73	rua ponta da cruz	rua josã© galvã£o n3 3dto.	55.52
74	hutai	shenyang	55.00
75	calle 52a no. 31 - 101 apto 201	kr 2 a sur n 18 35	54.57
76	desiderio arias	jose audilo santana	54.31
77	38 arlington	192 garden st	54.27
78	matadera 9 v noord aruba	casa 1	54.17
79	8848 beacon lakes dr	el rubio san jose de las matas	54.06
80	huioho	vhgvhgv	53.97
81	3215 villa strada way	rua: jose almeida macedo	53.95
82	sichuan chengdou	beijing	53.87
83	av 5 de agosto mz a1 lt 2	4375 york blvd apt#108	53.63
84	av 04 de julho	rua das hortencias	53.44
85	rua dr pita n67	calle libertador, edif zozzoro	53.33
86	kilometro 8 acaray, av. san blas y colombia	jr. leoncio prado 163	53.22
87	chorerras	jr. francisco pizarro	53.12
88	1234	125 lakeview lane apt 2	52.90
89	rua do serrado edif santa teresa ac1	av: goiania	52.48
90	nangang	beijing	52.38
91	huadian	beijing	52.38
92	divonzir luciano	rua santana	52.23
93	#42,street289,beong kok ii commune, toul kork district	#155, prey chisak village	52.06
94	av. central	manion st	51.94
95	rua antã³nio assunã§ã£o	av duque de avila n23	51.58
96	pedro vicente maldonado y simón bolivar	cra 86 d # 51 -18 sur	51.10
97	beijing	guanganjie61hao34-2-201	51.06
98	gabriela mistral	carrera 122 n 12-33	50.80
99	beijing	guangzhou	50.26
100	changchun	beijing	50.26

EXHIBIT 5

Summary of TelexFree User Accounts and Participants by Tier

Participant Type	Tier: Net Equity of Participant (in abs $)	User Accounts	Avg. User Accounts per Cluster	Avg. Net Equity per User Account	Participant Clusters	Participant %	Avg. Net Equity per Participant	Total Equity	Total Equity %	Cumulative Total Equity %
All	a. 0-100	**796,250**	1	n/a	**691,906**	44.2%	n/a	**$ (37,385,020)**	n/a	n/a
All	b. 100-1,000	**1,770,661**	5	n/a	**343,128**	21.9%	n/a	**$ (133,974,754)**	n/a	n/a
All	c. 1,000-5,000	**3,431,226**	8	n/a	**432,725**	27.6%	n/a	**$ (755,795,467)**	n/a	n/a
All	d. 5,000-10,000	**1,615,100**	31	n/a	**52,733**	3.4%	n/a	**$ (223,489,952)**	n/a	n/a
All	e. 10,000-25,000	**1,714,439**	55	n/a	**31,153**	2.0%	n/a	**$ (148,591,541)**	n/a	n/a
All	f. 25,000-50,000	**662,789**	85	n/a	**7,784**	0.5%	n/a	**$ 74,482,784**	n/a	n/a
All	g. 50,000-100,000	**417,684**	114	n/a	**3,660**	0.2%	n/a	**$ 158,163,187**	n/a	n/a
All	h. 100,000-250,000	**327,596**	148	n/a	**2,210**	0.1%	n/a	**$ 289,167,091**	n/a	n/a
All	i. 250,000-1,000,000	**162,467**	162	n/a	**1,004**	0.1%	n/a	**$ 428,560,678**	n/a	n/a
All	j. 1,000,000-2,000,000	**18,057**	287	n/a	**63**	0.0%	n/a	**$ 83,708,816**	n/a	n/a
All	k. 2,000,000-5,000,000	**71,348**	4,197	n/a	**17**	0.0%	n/a	**$ 38,963,930**	n/a	n/a
All	l. 5,000,000+	-	-	n/a	**0**	0.0%	n/a	**$ -**	n/a	n/a
	Total All	**10,987,617**	**7**	n/a	**1,566,383**	**100.0%**	n/a	**$ (226,190,248)**	n/a	n/a

Participant Type	Tier: Net Losses of Participant (in $)	User Accounts	Avg. User Accounts per Cluster	Avg. Net Equity per User Account	Participant Clusters	Participant %	Avg. Net Equity per Participant	Total Equity	Total Equity %	Cumulative Total Equity %
Net Loser	a. 0-100	**771,466**	1	$ (49)	**685,517**	46.2%	$ (55)	**$ (37,672,375)**	2.1%	2.1%
Net Loser	b. 100-1,000	**1,639,567**	5	$ (87)	**324,580**	21.9%	$ (439)	**$ (142,363,031)**	8.0%	10.1%
Net Loser	c. 1,000-5,000	**3,109,900**	8	$ (264)	**406,715**	27.4%	$ (2,019)	**$ (821,020,850)**	46.1%	56.2%
Net Loser	d. 5,000-10,000	**1,395,861**	33	$ (210)	**42,850**	2.9%	$ (6,856)	**$ (293,769,015)**	16.5%	72.7%
Net Loser	e. 10,000-25,000	**1,365,292**	64	$ (224)	**21,205**	1.4%	$ (14,436)	**$ (306,105,826)**	17.2%	89.8%
Net Loser	f. 25,000-50,000	**404,201**	138	$ (240)	**2,924**	0.2%	$ (33,153)	**$ (96,937,959)**	5.4%	95.3%
Net Loser	g. 50,000-100,000	**221,933**	311	$ (215)	**714**	0.0%	$ (66,717)	**$ (47,635,924)**	2.7%	97.9%
Net Loser	h. 100,000-250,000	**134,652**	769	$ (180)	**175**	0.0%	$ (138,226)	**$ (24,189,464)**	1.4%	99.3%
Net Loser	i. 250,000-1,000,000	**44,110**	2,206	$ (189)	**20**	0.0%	$ (416,451)	**$ (8,329,019)**	0.5%	99.8%
Net Loser	j. 1,000,000-2,000,000	**1,248**	1,248	$ (816)	**1**	0.0%	$ (1,018,291)	**$ (1,018,291)**	0.1%	99.8%
Net Loser	k. 2,000,000-5,000,000	**62,725**	62,725	$ (50)	**1**	0.0%	$ (3,129,978)	**$ (3,129,978)**	0.2%	100.0%
Net Loser	l. 5,000,000+	-	-	$ -	**0**	0.0%	$ -	**$ -**	0.0%	100.0%
	Total Net Losers	**9,150,955**	**6**	$ (195)	**1,484,702**	**100.0%**	$ (1,200)	**$ (1,782,171,730)**	**100.0%**	**100.0%**

Participant Type	Tier: Net Gains of Participant (in $)	User Accounts	Avg. User Accounts per Cluster	Avg. Net Equity per User Account	Participant Clusters	Participant %	Avg. Net Equity per Participant	Total Equity	Total Equity %	Cumulative Total Equity %
Net Winner	a. 0-100	**24,784**	4	$ 12	**6,389**	7.8%	$ 45	**$ 287,355**	0.0%	0.0%
Net Winner	b. 100-1,000	**131,094**	7	$ 64	**18,548**	22.7%	$ 452	**$ 8,388,276**	0.5%	0.6%
Net Winner	c. 1,000-5,000	**321,326**	12	$ 203	**26,010**	31.8%	$ 2,508	**$ 65,225,383**	4.2%	4.7%
Net Winner	d. 5,000-10,000	**219,239**	22	$ 321	**9,883**	12.1%	$ 7,111	**$ 70,279,063**	4.5%	9.3%
Net Winner	e. 10,000-25,000	**349,147**	35	$ 451	**9,948**	12.2%	$ 15,834	**$ 157,514,285**	10.1%	19.4%
Net Winner	f. 25,000-50,000	**258,588**	53	$ 663	**4,860**	5.9%	$ 35,272	**$ 171,420,743**	11.0%	30.4%
Net Winner	g. 50,000-100,000	**195,751**	66	$ 1,051	**2,946**	3.6%	$ 69,857	**$ 205,799,111**	13.2%	43.6%
Net Winner	h. 100,000-250,000	**192,944**	95	$ 1,624	**2,035**	2.5%	$ 153,984	**$ 313,356,555**	20.1%	63.8%
Net Winner	i. 250,000-1,000,000	**118,357**	120	$ 3,691	**984**	1.2%	$ 443,994	**$ 436,889,697**	28.1%	91.8%
Net Winner	j. 1,000,000-2,000,000	**16,809**	271	$ 5,041	**62**	0.1%	$ 1,366,566	**$ 84,727,107**	5.4%	97.3%
Net Winner	k. 2,000,000-5,000,000	**8,623**	539	$ 4,882	**16**	0.0%	$ 2,630,869	**$ 42,093,908**	2.7%	100.0%
Net Winner	l. 5,000,000+	-	-	$ -	**0**	0.0%	$ -	**$ -**	0.0%	100.0%
	Total Net Winners	**1,836,662**	**22**	$ 847	**81,681**	**100.0%**	$ 19,049	**$ 1,555,981,482**	**100.0%**	**100.0%**

EXHIBIT 6

Provided natively as *aggregation_output.csv*

EXHIBIT 7

Provided natively as *aggregation_summary.csv*

EXHIBIT 3

COMMITTEE ON RULES OF PRACTICE AND PROCEDURE
OF THE
JUDICIAL CONFERENCE OF THE UNITED STATES
WASHINGTON, D.C. 20544

JOHN D. BATES
CHAIR

CHAIRS OF ADVISORY COMMITTEES

JAY S. BYBEE
APPELLATE RULES

DENNIS R. DOW
BANKRUPTCY RULES

ROBERT M. DOW, JR.
CIVIL RULES

RAYMOND M. KETHLEDGE
CRIMINAL RULES

PATRICK J. SCHILTZ
EVIDENCE RULES

MEMORANDUM

TO: Honorable John D. Bates, Chair
Standing Committee on Rules of Practice and Procedure

FROM: Honorable Patrick J. Schiltz, Chair
Advisory Committee on Evidence Rules

RE: Report of the Advisory Committee on Evidence Rules

DATE: May 15, 2022

I. Introduction

The Advisory Committee on Evidence Rules (the "Committee") met in Washington, D.C., on May 6, 2021. At the meeting the Committee discussed and gave final approval to three proposed amendments that had been released for public comment. The Committee also considered and approved six proposed amendments with the recommendation that they be released for public comment.

The Committee made the following determinations at the meeting:

● It unanimously approved proposed amendments to Rules 106, 615, and 702, and recommends to the Standing Committee that they be transmitted to the Judicial Conference.

● It unanimously approved proposals to amend Rules 611 (adding two new subdivisions), 613(b), 801(d)(2), 804(b)(3), and 1006, and recommends to the Standing Committee that these proposed amendments be released for public comment.

A full description of all of these matters can be found in the draft minutes of the Committee meeting, attached to this Report. The proposed amendments can also be found as attachments to this Report.

II. Action Items

A. Proposed Amendment to Rule 106, for Final Approval

At the suggestion of Judge Paul Grimm, the Committee has for the last five years considered and discussed whether Rule 106 --- the rule of completeness --- should be amended. Rule 106 provides that if a party introduces all or part of a written or recorded statement in such a way as to be misleading, the opponent may introduce a completing statement that would correct the misimpression. The Committee has considered whether Rule 106 should be amended in two respects: 1) to provide that a completing statement is admissible over a hearsay objection; and 2) to expand the rule to cover unrecorded oral statements, as well as written and recorded statements.

The courts are not uniform in their treatment of these issues. On the hearsay question, some courts have held that when a party introduces a portion of a statement that is misleading, that party can still object, on hearsay grounds, to completing evidence that corrects the misimpression. Other courts have held essentially that if a party introduces a portion of a statement in a manner that misleads the factfinder, that party forfeits the right to object to introduction of other portions of that statement when that is necessary to remedy the misimpression. As to unrecorded oral statements, most courts have found that when necessary to complete, such statements are admissible either under Rule 611(a) or under the common law rule of completeness.

After much discussion and consideration, the Committee in Spring, 2021 unanimously approved an amendment for release for public comment. The proposal released for public comment allows the completing statement to be admitted over a hearsay objection and covers unrecorded oral statements.

The overriding goal of the amendment is to treat all questions of completeness in a single rule. That is particularly important because completeness questions often arise at trial, and so it is important for the parties and the court to be able to refer to a single rule to govern admissibility.

What has been particularly confusing to courts and practitioners is that Rule 106 has been considered a "partial codification" of the common law --- meaning that the parties must be aware that common law may still be invoked. As stated in the Committee Note, the amendment is intended to displace the common law, just as the common law has been displaced by all of the other Federal Rules of Evidence.

As to admissibility of out-of-court statements, the amendment takes the position that the proponent, by introducing part of a statement in a misleading manner, forfeits the right to foreclose admission of a remainder that is necessary to remedy the misimpression. Simple notions of fairness, already embodied in Rule 106, dictate that a misleading presentation cannot stand unrebutted. The amendment leaves it up to the court to determine whether the completing remainder will be admissible to prove a fact (a hearsay use) or simply to provide context (a non-hearsay use). Either usage is encompassed within the rule terminology --- that the completing remainder is admissible "over a hearsay objection."

As to unrecorded oral statements, most courts already admit such statements when necessary to complete --- they just do so under a different evidence rule or under the common law. The Committee was convinced that covering unrecorded oral statements under Rule 106 would be a user-friendly change, especially because the existing hodgepodge of coverage of unrecorded statements presents a trap for the unwary. As stated above, the fact that completeness questions almost always arise at trial means that parties cannot be expected to quickly get an answer from the common law, or from a rule such as Rule 611(a) that does not specifically deal with completeness.

It is important to note that nothing in the amendment changes the basic rule, which applies only to the narrow circumstances in which a party has created a misimpression about the statement, and the adverse party proffers a completing statement that in fact corrects the misimpression. So, the mere fact that a statement is probative and contradicts a statement offered by the opponent is not enough to justify completion under Rule 106.

The Committee received only a few public comments on the proposed changes to Rule 106. All comments were in favor of the proposed amendment, with a couple of comments providing some suggestions for minor changes. After considering the public comment, the Committee unanimously approved a slight change to the proposal: deletion of the phrase "written or oral," which makes clear that Rule 106 applies to all statements, including those that are not written or oral. The Committee determined that statements made through conduct, or through sign language, should be covered by the rule of completeness, as there was no reason to distinguish such statements from those that are written or oral. The proposed Committee Note was slightly revised to accord with the change in text.

At its Spring 2022 meeting, the Committee unanimously gave final approval to the proposed amendment to Rule 106. The Committee recommends that the proposed amendment, and the accompanying Committee Note, be approved by the Standing Committee and referred to the Judicial Conference.

The proposed amendment to Rule 106, together with the proposed Committee Note, the GAP report, and the summary of public comment, is attached to this Report.

B. Proposed Amendment to Rule 615, for Final Approval

Rule 615 provides for court orders excluding witnesses so that they "cannot hear other witnesses' testimony." The Committee determined that there are problems raised in the case law and in practice regarding the scope of a Rule 615 order: does it apply only to exclude witnesses from the courtroom (as stated in the text of the rule) or does it extend outside the confines of the courtroom to prevent prospective witnesses from obtaining or being provided trial testimony? Most courts have held that a Rule 615 order extends to prevent access to trial testimony outside of the courtroom, because exclusion from the courtroom is not sufficient to protect against the risk of witnesses tailoring their testimony after obtaining access to trial testimony. But other courts have read the rule as it is written.

After extensive consideration and research over four years, the Committee agreed on an amendment that would clarify the extent of an order under Rule 615. Committee members have noted that where parties can be held in contempt for violating a court order, due process requires that the order be clear if it seeks to do more than exclude witnesses from the courtroom. The Committee's investigation of this problem is consistent with its ongoing efforts to ensure that the Evidence Rules are keeping up with technological advancement, given the increased possibility of witness access to information about testimony through news, social media, YouTube, or daily transcripts.

At its Spring, 2021 meeting the Committee unanimously voted in favor of an amendment to Rule 615. That amendment, released for public comment in August, 2021, limits an exclusion order to just that --- exclusion of witnesses from the courtroom. But a new subdivision provides that the court has discretion to issue further orders to "(1) prohibit disclosure of trial testimony to witnesses who are excluded from the courtroom; and (2) prohibit excluded witnesses from accessing trial testimony." In other words, if a court wants to do more than exclude witnesses from the courtroom, the court must say so.

The Committee also considered whether an amendment to Rule 615 should address orders that prohibit counsel from referring to trial testimony while preparing prospective witnesses. The Committee resolved that any amendment to Rule 615 should not mention trial counsel in text, because the question of whether counsel can use trial testimony to prepare witnesses raises issues of professional responsibility and the right to counsel that are beyond the purview of the Evidence Rules. Judges must address these issues on a case-by-case basis.

Finally, the Committee approved an additional amendment to the existing provision that allows an entity-party to designate "an officer or employee" to be exempt from exclusion. There is some dispute in the courts about whether the entity-party is limited to one such exemption or is entitled to more than one. The amendment clarifies that the exemption is limited to one officer or

employee. The rationale is that the exemption is intended to put entities on a par with individual parties, who cannot be excluded under Rule 615. Allowing the entity more than one exemption is inconsistent with that rationale.

As noted, these proposed changes to Rule 615 were released for public comment in August, 2021. Only a few public comments were received. All were supportive of the amendment, with two comments suggesting minor changes. In response to the public comment, the Committee made two minor changes the Committee Note to the proposed amendment.

At its Spring 2022 meeting, the Committee unanimously gave final approval to the proposed amendment to Rule 615. The Committee recommends that the proposed amendment, and the accompanying Committee Note, be approved by the Standing Committee and referred to the Judicial Conference.

The proposed amendment to Rule 615, together with the Committee Note, the GAP report, and the summary of public comment, is attached to this Report.

C. Proposed Amendment to Rule 702, for Final Approval

The Committee has been researching and discussing the possibility of an amendment to Rule 702 for five years. The project began with a Symposium on forensic experts and *Daubert,* held at Boston College School of Law in October, 2017. That Symposium addressed, among other things, the challenges to forensic evidence raised in a report by the President's Council of Advisors on Science and Technology. A Subcommittee on Rule 702 was appointed to consider possible treatment of forensic experts, as well as the weight/admissibility question discussed below. The Subcommittee, after extensive discussion, recommended against certain courses of action. The Subcommittee found that: 1) It would be difficult to draft a freestanding rule on forensic expert testimony, because any such amendment would have an inevitable and problematic overlap with Rule 702; and 2) It would not be advisable to set forth detailed requirements for forensic evidence either in text or Committee Note because such a project would require extensive input from the scientific community, and there is substantial debate about what requirements are appropriate.

The full Committee agreed with these suggestions. But the Subcommittee did express interest in considering an amendment to Rule 702 that would focus on one important aspect of forensic expert testimony --- the problem of overstating results (for example, an expert claiming that her opinion has a "zero error rate", where that conclusion is not supportable by the expert's methodology). The Committee heard extensively from DOJ on the important efforts it is now employing to regulate the testimony of its forensic experts, and to limit possible overstatement.

The Committee considered a proposal to add a new subdivision (e) to Rule 702 that would essentially prohibit any expert from drawing a conclusion overstating what could actually be concluded from a reliable application of a reliable methodology. But a majority of the members decided that the amendment would be problematic, because Rule 702(d) already requires that the expert must reliably apply a reliable methodology. If an expert overstates what can be reliably

concluded (such as a forensic expert saying the rate of error is zero) then the expert's opinion should be excluded under Rule 702(d). The Committee was also concerned about the possible unintended consequences of adding an overstatement provision that would be applied to all experts, not just forensic experts.

The Committee, however, unanimously favored a slight change to existing Rule 702(d) that would emphasize that the court must focus on the expert's opinion, and must find that the opinion actually proceeds from a reliable application of the methodology. The Committee unanimously approved a proposal—released for public comment in August, 2021--- that would amend Rule 702(d) to require the court to find that "the expert's opinion reflects a reliable application of the principles and methods to the facts of the case." As the Committee Note elaborates: "A testifying expert's opinion must stay within the bounds of what can be concluded by a reliable application of the expert's basis and methodology." The language of the amendment more clearly empowers the court to pass judgment on the conclusion that the expert has drawn from the methodology. Thus the amendment is consistent with *General Electric Co., v. Joiner,* 522 U.S. 136 (1997), in which the Court declared that a trial court must consider not only the expert's methodology but also the expert's conclusion; that is because the methodology must not only be reliable, it must be reliably applied.

Finally, the Committee resolved to respond to the fact that many courts have declared that the reliability requirements set forth in Rule 702(b) and (d) --- that the expert has relied on sufficient facts or data and has reliably applied a reliable methodology --- are questions of weight and not admissibility, and more broadly that expert testimony is presumed to be admissible. These statements misstate Rule 702, because its admissibility requirements must be established to a court by a preponderance of the evidence. The Committee concluded that in a fair number of cases, the courts have found expert testimony admissible even though the proponent has not satisfied the Rule 702(b) and (d) requirements by a preponderance of the evidence --- essentially treating these questions as ones of weight rather than admissibility, which is contrary to the Supreme Court's holdings that under Rule 104(a), admissibility requirements are to be determined by court under the preponderance standard.

Initially, the Committee was reluctant to propose a change to the text of Rule 702 to address these mistakes as to the proper standard of admissibility, in part because the preponderance of the evidence standard applies to almost all evidentiary determinations, and specifying that standard in one rule might raise negative inferences as to other rules. But ultimately the Committee unanimously agreed that explicitly weaving the Rule 104(a) standard into the text of Rule 702 would be a substantial improvement that would address an important conflict among the courts. While it is true that the Rule 104(a) preponderance of the evidence standard applies to Rule 702 as well as other rules, it is with respect to the reliability requirements of expert testimony that many courts are misapplying that standard. Moreover, it takes some effort to determine the applicable standard of proof --- Rule 104(a) does not mention the applicable standard of proof, requiring a resort to case law. And while *Daubert* mentions the standard, *Daubert* does so only in a footnote in the midst of much discussion about the liberal standards of the Federal Rules of Evidence. Consequently, the Committee unanimously approved an amendment for public comment that

would explicitly add the preponderance of the evidence standard to Rule 702(b)-(d). The language of the proposal released for public comment required that "the proponent has demonstrated by a preponderance of the evidence" that the reliability requirements of Rule 702 have been met. The Committee Note to the proposal made clear that there is no intent to raise any negative inference regarding the applicability of the Rule 104(a) standard of proof to other rules --- emphasizing that incorporating the preponderance standard into the text of Rule 702 was made necessary by the decisions that have failed to apply it to the reliability requirements of Rule 702.

More than 500 comments were received on the proposed amendments to Rule 702. In addition, a number of comments were received at a public hearing held on the rule. Many of the comments were opposed to the amendment, and almost all of the fire was directed toward the term "preponderance of the evidence." Some thought that "preponderance of the evidence" would limit the court to considering only *admissible* evidence at the *Daubert* hearing. Others thought that the term represented a shift from the jury to the judge as factfinder. By contrast, commentators who supported the amendment argued that the amendment should go further and clarify that it is the court, not the jury, that decides admissibility.

The Committee carefully considered the public comments. The Committee does not agree that the preponderance of the evidence standard would limit the court to considering only admissible evidence; the plain language of Rule 104(a) allows the court deciding admissibility to consider inadmissible evidence. Nor did the Committee believe that the use of the term preponderance of the evidence would shift the factfinding role from the jury to the judge, for the simple reason that, when it comes to making preliminary determinations about admissibility, the judge *is* and *always has been* a factfinder.

But while disagreeing with these comments, the Committee recognized that it would be possible to replace the term "preponderance of the evidence" with a term that would achieve the same purpose while not raising the concerns (valid or not) mentioned by many commentators. The Committee unanimously agreed to change the proposal as issued for public comment to provide that the proponent must establish that it is "*more likely than not*" that the reliability requirements are met. This standard is substantively identical to "preponderance of the evidence" but it avoids any reference to "evidence" and thus addresses the concern that the term "evidence" means only admissible evidence.

The Committee was also convinced by the suggestion in the public comment that the rule should clarify that it is the court and not the jury that must decide whether it is more likely than not that the reliability requirements of the rule have been met. Therefore, the Committee unanimously agreed with a change requiring that the proponent establish "*to the court*" that it is more likely than not that the reliability requirements have been met. The proposed Committee Note was amended to clarify that nothing in amended Rule 702 requires a court to make any findings about reliability in the absence of a proper objection.

With those changes, and a few stylistic and corresponding changes to the Committee Note, the Committee unanimously voted in favor of adopting the amendments to Rule 702, for final approval.

At the Spring 2022 meeting, the Committee unanimously gave final approval to the proposed amendment to Rule 702. The Committee recommends that the proposed amendment, and the accompanying Committee Note, be approved by the Standing Committee and referred to the Judicial Conference.

The proposed amendment to Rule 702, together with the proposed Committee Note, GAP report, summary of public comment, and summary of the public hearing, is attached to this Report.

D. Possible Amendment to Rule 611 on Illustrative Aids, for Release for Public Comment

At the Spring meeting, the Committee unanimously approved a proposal to add a new Rule 611(d) to regulate the use of illustrative aids at trial. The distinction between "demonstrative evidence" (admitted into evidence and used substantively to prove disputed issues at trial) and "illustrative aids" (not admitted into evidence but used solely to assist the jury in understanding other evidence) is sometimes a difficult one to draw, and is a point of confusion in the courts. In addition, the standards for allowing illustrative aids to be presented --- and particularly whether illustrative aids may be used by the jury during deliberations --- are not made clear in the case law. The Committee has determined that it would be useful to set forth uniform standards to regulate the use of illustrative aids, and in doing so clarify the distinction between illustrative aids and demonstrative evidence.

The proposed amendment would distinguish illustrative aids --- presentations that are not evidence but offered only to help the factfinder understand evidence --- from demonstrative evidence offered to prove a fact. The amendment would allow illustrative aids to be used at trial after the court balances the utility of the aid against the risk of unfair prejudice, confusion, and delay.

Because illustrative aids are not evidence, adverse parties do not receive pretrial discovery of such aids. The proposed rule would require notice to be provided, unless the court for good cause orders otherwise. The Committee determined that advance notice is important so that the court can rule on whether the aid has sufficient utility before it is displayed to the jury. (After all, you can't unring a bell.) The Committee Note recognizes that the timing of the notice will depend on the circumstances.

Finally, because illustrative aids are not evidence, the proposed rule provides that the aids should not be allowed into the jury room during deliberations, unless the court orders otherwise. The Committee Note specifies that if the court does allow an illustrative aid to go to the jury room, the court should instruct the jury that the aid is not evidence.

It is important to note that the proposed rule is not intended to regulate PowerPoints or other aids that an attorney uses merely to guide the jury through an opening or closing argument. Again, illustrative aids assist the jury in understanding *evidence*; something that assists the jury in following an *argument* is therefore not an illustrative aid.

The Committee strongly believes that the rule on illustrative aids will provide an important service to courts and litigants. Illustrative aids are used in almost every trial, and yet nothing in the evidence rules specifically addresses their use. This amendment rectifies that problem.

At the Spring 2022 meeting, the Committee unanimously approved the proposed amendment to add Rule 611(d) to regulate the use of illustrative aids at a trial. The Committee recommends that the proposed amendment, and the accompanying Committee Note, be released for public comment.

The proposed amendment to add Rule 611(d), together with the proposed Committee Note, is attached to this Report.

E. Proposed Amendment to Rule 1006, for Release for Public Comment[1]

Evidence Rule 1006 provides that a summary can be admitted as evidence if the underlying records are admissible and too voluminous to be conveniently examined in court. The Committee has determined that the courts are in dispute about a number of issues regarding admissibility of summaries of evidence under Rule 1006 --- and that much of the problem is that some courts do not properly distinguish between summaries of evidence under Rule 1006 (which are themselves admitted into evidence) and summaries that are illustrative aids (which are not evidence at all). Some courts have stated that summaries admissible under Rule 1006 are "not evidence," which is incorrect. Other courts have stated that all of the underlying evidence must be admitted before the summary can be admitted; that, too, is incorrect. Still other courts state that the summary is inadmissible if any of the underlying evidence *has* been admitted; that is also wrong.

After extensive research and discussion, the Committee unanimously approved an amendment to Rule 1006 that would provide greater guidance to the courts on the admissibility and proper use of summary evidence under Rule 1006.

The proposal to amend Rule 1006 dovetails with the proposal to establish a rule on illustrative aids, discussed above. These two rules serve to distinguish a summary of voluminous evidence (which is itself evidence and governed by Rule 1006) from a summary that is designed to help the trier of fact understand evidence that has already been presented (which is not itself evidence and would be governed by new Rule 611(d)). The proposed amendment to Rule 1006 would clarify that a summary is admissible whether or not the underlying evidence has been

[1] This rule is taken out of numerical sequence, because it is of a piece with the proposed amendment on illustrative aids.

admitted. The Committee believes that the proposed amendment will provide substantial assistance to courts and litigants in navigating this confusing area.

At the Spring 2022 meeting, the Committee unanimously approved the proposed amendment to Rule 1006. The Committee recommends that the proposed amendment, and the accompanying Committee Note, be released for public comment.

The proposed amendment to Rule 1006, together with the proposed Committee Note, is attached to this Report.

F. Proposed Rule 611(e), Setting Forth Safeguards When Allowing Jurors to Submit Questions for Witnesses, for Release for Public Comment.

There is controversy in the courts over whether jurors should be allowed to question witnesses at trial. The Committee is not seeking to resolve that controversy in a rule amendment. But the Committee has determined that it would be useful to set forth the minimum safeguards that should be applied if the trial court does decide to allow jurors to question witnesses. Standards regulating the practice can be found in some court of appeals cases, but the Committee has unanimously determined that it would be useful to set forth a single set of safeguards in an Evidence Rule --- specifically, in a new subdivision 611(e). The proposed Rule 611(e) requires the court to instruct jurors, among other things, that they must submit questions in writing; that they are not to draw negative inferences if their question is rephrased or does not get asked; and that they must maintain their neutrality. The proposed rule also provides that the court must consult with counsel when jurors submit questions, and that counsel must be allowed to object to such questions outside the jury's hearing.

The Committee Note to proposed Rule 611(e) emphasizes that the rule is agnostic about whether a court decides to permit jurors to submit questions.

At the Spring 2022 meeting, the Committee unanimously approved the proposed amendment to add a new Rule 611(e). The Committee recommends that the proposed amendment, and the accompanying Committee Note, be released for public comment.

The proposed amendment to add Rule 611(e), together with the proposed Committee Note, is attached to this Report.

G. Proposed Amendment to Rule 613(b), for Release for Public Comment.

The common law provided that before a witness could be impeached with extrinsic evidence of a prior inconsistent statement, the adverse party was required to give the witness an opportunity to explain or deny the statement. Rule 613(b) rejects that "prior presentation" requirement. It provides that extrinsic evidence of the inconsistent statement is admissible so long

as the witness is given an opportunity to explain or deny the statement at some point in the trial. It turns out, though, that many (perhaps most) courts have retained the common law "prior presentation" requirement. These courts have found that a prior presentation requirement saves time, because a witness will almost always concede that she made the inconsistent statement, and that makes it unnecessary for anyone to introduce extrinsic evidence. The prior presentation requirement also avoids unfair surprise and the difficulties inherent in calling a witness back to the stand to give her an opportunity at some later point to explain or deny a prior statement that has been proven through extrinsic evidence.

After discussion at three Committee meetings, the Committee unanimously determined that the better rule is to require a prior opportunity to explain or deny the statement, with the court having discretion to allow a later opportunity (for example, when the prior inconsistent statement is not discovered until after the witness testifies). This will bring the rule into alignment with what appears to be the practice of most trial judges --- a practice that the Committee concluded is superior to the practice described in the current rule.

The Committee unanimously approved the following change:

> Extrinsic evidence of a witness's prior inconsistent statement ~~is admissible only if~~ may not be admitted until after the witness is given an opportunity to explain or deny the statement and an adverse party is given an opportunity to examine the witness about it, unless the court orders otherwise ~~or if justice so requires~~. This subdivision (b) does not apply to an opposing party's statement under Rule 801(d)(2).

At the Spring 2022 meeting, the Committee unanimously approved the proposed amendment to Rule 613(b). The Committee recommends that the proposed amendment, and the accompanying Committee Note, be released for public comment.

The proposed amendment to Rule 613(b), together with the proposed Committee Note, is attached to this Report.

H. Proposed Amendment to Rule 801(d)(2) Governing Successors-in-Interest, for Release for Public Comment

Rule 801(d)(2) provides a hearsay exemption for statements of a party opponent. Courts are split about the applicability of this exemption in the following situation: a declarant makes a statement that would have been admissible against him as a party-opponent, but he is not the party-opponent because his claim or defense has been transferred to another (either by agreement or by operation of law), and it is the transferee that is the party-opponent. Some circuits would permit the statements made by the declarant to be offered against the successor as a party-opponent statement under Rule 801(d)(2), while others would foreclose admissibility because the statement was made by one who is technically not the party-opponent in the case.

At its Spring, 2002 meeting, after previous discussion, the Committee determined that the dispute in the courts about the admissibility of party-opponent statements against successors should be resolved by a rule amendment, because the problem arises with some frequency in a variety of successor/predecessor situations (most commonly, decedent and estate in a claim brought for damages under 42 U.S.C. § 1983). The Committee unanimously determined that the appropriate result should be that a hearsay statement should be admissible against the successor-in-interest. The Committee reasoned that admissibility was fair when the successor-in-interest is standing in the shoes of the declarant --- because the declarant is in substance the party-opponent. Moreover, a contrary rule results in random application of Rule 801(d)(2), and possible strategic action, such as assigning a claim in order to avoid admissibility of a statement. The Committee approved the following addition to Rule 801(d)(2):

> If a party's claim or potential liability is directly derived from a declarant or the declarant's principal, a statement that would be admissible against the declarant or the principal under this rule is also admissible against the party.

The proposed Committee Note would emphasize that to be admissible against the successor, the declarant must have made the statement before the transfer of the claim or defense.

At its Spring 2022 meeting, the Committee unanimously approved the proposed amendment to Rule 801(d)(2). The Committee recommends that the proposed amendment, and the accompanying Committee Note, be released for public comment.

The proposed amendment to Rule 801(d)(2), together with the proposed Committee Note, is attached to this Report.

I. Proposed Amendment to the Rule 804(b)(3) Corroborating Circumstances Requirement, for Release for Public Comment

Rule 804(b)(3) provides a hearsay exception for declarations against interest. In a criminal case in which a declaration against penal interest is offered, the rule requires that the proponent provide "corroborating circumstances that clearly indicate the trustworthiness" of the statement. There is a dispute in the courts about the meaning of the "corroborating circumstances" requirement. Most federal courts consider both the inherent guarantees of trustworthiness underlying a particular declaration against interest as well as independent evidence corroborating (or refuting) the accuracy of the statement. But some courts do not permit inquiry into independent evidence --- limiting judges to consideration of the inherent guarantees of trustworthiness surrounding the statement. This latter view --- denying consideration of independent corroborative evidence --- is inconsistent with the 2019 amendment to Rule 807 (the residual exception), which requires courts to look at corroborative evidence in determining whether a hearsay statement is sufficiently trustworthy under that exception. The rationale is that corroborative evidence can shore up concerns about the potential unreliability of a statement --- a rationale that is applied in

many other contexts, such as admissibility of co-conspirator hearsay, and tips from informants in determining probable cause.

At its Spring, 2022 meeting, the Committee unanimously approved an amendment to Rule 804(b)(3) that would parallel the language in Rule 807, and require the court to consider the presence or absence of corroborating evidence in determining whether "corroborating circumstances" exist. The proposed language for the amendment, which is recommended for release for public comment, is as follows:

Rule 804(b)(3) Statement Against Interest.

A statement that:

> (A) A reasonable person in the declarant's position would have made only if the person believed it to be true because, when made, it was so contrary to the declarant's proprietary or pecuniary interest or had so great a tendency to invalidate the declarant's claim against someone else or to expose the declarant to civil or criminal liability; and
>
> (B) if offered in a criminal case as one that tends to expose the declarant to criminal liability, the court finds it ~~is~~ supported by corroborating circumstances that clearly indicate trustworthiness --- after considering the totality of circumstances under which it was made and evidence, if any, corroborating it. ~~if offered in a criminal case as one that tends to expose the declarant to criminal liability~~.

The Committee believes that it is important to rectify the dispute among the circuits about the meaning of "corroborating circumstances" and that requiring consideration of corroborating evidence not only avoids inconsistency with the residual exception, but is also supported by logic and by the legislative history of Rule 804(b)(3).

At its Spring 2022 meeting, the Committee unanimously approved the proposed amendment to Rule 80(4)(b)(3). The Committee recommends that the proposed amendment, and the accompanying Committee Note, be released for public comment.

The proposed amendment to Rule 804(b)(3), together with the proposed Committee Note, is attached to this Report.

III. Minutes of the Spring, 2022 Meeting

The draft of the minutes of the Committee's Spring, 2022 meeting is attached to this report. These minutes have not yet been approved by the Committee.

EXHIBIT 4

April 24, 2023

Honorable Kevin McCarthy
Speaker, United States House of Representatives
Washington, DC 20515

Dear Mr. Speaker:

I have the honor to submit to the Congress amendments to the Federal Rules of Appellate Procedure that have been adopted by the Supreme Court of the United States pursuant to Section 2072 of Title 28, United States Code.

Accompanying the amended rules are the following materials that were submitted to the Court for its consideration pursuant to Section 331 of Title 28, United States Code: a transmittal letter to the Court dated October 19, 2022; a blackline version of the rules with committee notes; an excerpt from the September 2022 report of the Committee on Rules of Practice and Procedure to the Judicial Conference of the United States; and excerpts from the May 2022 reports of the Advisory Committee on Appellate Rules.

Sincerely,

/s/ John G. Roberts, Jr.

April 24, 2023

Honorable Kamala D. Harris
President, United States Senate
Washington, DC 20510

Dear Madam President:

I have the honor to submit to the Congress amendments to the Federal Rules of Appellate Procedure that have been adopted by the Supreme Court of the United States pursuant to Section 2072 of Title 28, United States Code.

Accompanying the amended rules are the following materials that were submitted to the Court for its consideration pursuant to Section 331 of Title 28, United States Code: a transmittal letter to the Court dated October 19, 2022; a blackline version of the rules with committee notes; an excerpt from the September 2022 report of the Committee on Rules of Practice and Procedure to the Judicial Conference of the United States; and excerpts from the May 2022 reports of the Advisory Committee on Appellate Rules.

Sincerely,

/s/ John G. Roberts, Jr.

April 24, 2023

SUPREME COURT OF THE UNITED STATES

ORDERED:

1. The Federal Rules of Evidence are amended to include amendments to Rules 106, 615, and 702.

[*See infra* pp. ___ ___ ___.]

2. The foregoing amendments to the Federal Rules of Evidence shall take effect on December 1, 2023, and shall govern in all proceedings thereafter commenced and, insofar as just and practicable, all proceedings then pending.

3. THE CHIEF JUSTICE is authorized to transmit to the Congress the foregoing amendments to the Federal Rules of Evidence in accordance with the provisions of Section 2074 of Title 28, United States Code.

JUDICIAL CONFERENCE OF THE UNITED STATES

WASHINGTON, D.C. 20544

THE CHIEF JUSTICE
OF THE UNITED STATES
Presiding

HONORABLE ROSLYNN R. MAUSKOPF
Secretary

October 19, 2022

MEMORANDUM

To: The Chief Justice of the United States
The Associate Justices of the Supreme Court

From: Judge Roslynn R. Mauskopf Roslyn R. Mauskopf

RE: TRANSMITTAL OF PROPOSED AMENDMENTS TO THE FEDERAL RULES OF EVIDENCE

By direction of the Judicial Conference of the United States, pursuant to the authority conferred by 28 U.S.C. § 331, I transmit for the Court's consideration proposed amendments to Rules 106, 615, and 702 of the Federal Rules of Evidence, which have been approved by the Judicial Conference. The Judicial Conference recommends that the amendments be adopted by the Court and transmitted to Congress pursuant to law.

For your assistance in considering the proposed amendments, I am transmitting (i) clean and blackline copies of the amended rules along with committee notes; (ii) an excerpt from the September 2022 report of the Committee on Rules of Practice and Procedure to the Judicial Conference; and (iii) an excerpt from the May 2022 report of the Advisory Committee on Evidence Rules.

Attachments

PROPOSED AMENDMENTS TO THE FEDERAL RULES OF EVIDENCE[1]

Rule 702. Testimony by Expert Witnesses

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if the proponent demonstrates to the court that it is more likely than not that:

(a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;

(b) the testimony is based on sufficient facts or data;

(c) the testimony is the product of reliable principles and methods; and

(d) the ~~expert has reliably applied~~ expert's opinion reflects a reliable application of the

[1] New material is underlined; matter to be omitted is lined through.

principles and methods to the facts of the case.

Committee Note

Rule 702 has been amended in two respects:

(1) First, the rule has been amended to clarify and emphasize that expert testimony may not be admitted unless the proponent demonstrates to the court that it is more likely than not that the proffered testimony meets the admissibility requirements set forth in the rule. *See* Rule 104(a). This is the preponderance of the evidence standard that applies to most of the admissibility requirements set forth in the evidence rules. *See Bourjaily v. United States*, 483 U.S. 171, 175 (1987) ("The preponderance standard ensures that before admitting evidence, the court will have found it more likely than not that the technical issues and policy concerns addressed by the Federal Rules of Evidence have been afforded due consideration."); *Huddleston v. United States*, 485 U.S. 681, 687 n.5 (1988) ("preliminary factual findings under Rule 104(a) are subject to the preponderance-of-the-evidence standard"). But many courts have held that the critical questions of the sufficiency of an expert's basis, and the application of the expert's methodology, are questions of weight and not admissibility. These rulings are an incorrect application of Rules 702 and 104(a).

There is no intent to raise any negative inference regarding the applicability of the Rule 104(a) standard of proof for other rules. The Committee concluded that emphasizing the preponderance standard in Rule 702 specifically was made necessary by the courts that have failed to apply correctly the reliability requirements of that

rule. Nor does the amendment require that the court make a finding of reliability in the absence of objection.

The amendment clarifies that the preponderance standard applies to the three reliability-based requirements added in 2000—requirements that many courts have incorrectly determined to be governed by the more permissive Rule 104(b) standard. But it remains the case that other admissibility requirements in the rule (such as that the expert must be qualified and the expert's testimony must help the trier of fact) are governed by the Rule 104(a) standard as well.

Some challenges to expert testimony will raise matters of weight rather than admissibility even under the Rule 104(a) standard. For example, if the court finds it more likely than not that an expert has a sufficient basis to support an opinion, the fact that the expert has not read every single study that exists will raise a question of weight and not admissibility. But this does not mean, as certain courts have held, that arguments about the sufficiency of an expert's basis always go to weight and not admissibility. Rather it means that once the court has found it more likely than not that the admissibility requirement has been met, any attack by the opponent will go only to the weight of the evidence.

It will often occur that experts come to different conclusions based on contested sets of facts. Where that is so, the Rule 104(a) standard does not necessarily require exclusion of either side's experts. Rather, by deciding the disputed facts, the jury can decide which side's experts to credit. "[P]roponents 'do not have to demonstrate to the judge by a preponderance of the evidence that the assessments of their experts are correct, they only have to demonstrate by a preponderance of evidence that their opinions are reliable. . . . The evidentiary requirement of

reliability is lower than the merits standard of correctness.'" Advisory Committee Note to the 2000 amendment to Rule 702, quoting *In re Paoli R.R. Yard PCB Litigation*, 35 F.3d 717, 744 (3d Cir. 1994).

Rule 702 requires that the expert's knowledge "help" the trier of fact to understand the evidence or to determine a fact in issue. Unfortunately, some courts have required the expert's testimony to "appreciably help" the trier of fact. Applying a higher standard than helpfulness to otherwise reliable expert testimony is unnecessarily strict.

(2) Rule 702(d) has also been amended to emphasize that each expert opinion must stay within the bounds of what can be concluded from a reliable application of the expert's basis and methodology. Judicial gatekeeping is essential because just as jurors may be unable, due to lack of specialized knowledge, to evaluate meaningfully the reliability of scientific and other methods underlying expert opinion, jurors may also lack the specialized knowledge to determine whether the conclusions of an expert go beyond what the expert's basis and methodology may reliably support.

The amendment is especially pertinent to the testimony of forensic experts in both criminal and civil cases. Forensic experts should avoid assertions of absolute or one hundred percent certainty—or to a reasonable degree of scientific certainty—if the methodology is subjective and thus potentially subject to error. In deciding whether to admit forensic expert testimony, the judge should (where possible) receive an estimate of the known or potential rate of error of the methodology employed, based (where appropriate) on studies that reflect how often the method produces accurate results. Expert opinion testimony regarding the weight of feature comparison evidence (i.e., evidence that a set of

features corresponds between two examined items) must be limited to those inferences that can reasonably be drawn from a reliable application of the principles and methods. This amendment does not, however, bar testimony that comports with substantive law requiring opinions to a particular degree of certainty.

Nothing in the amendment imposes any new, specific procedures. Rather, the amendment is simply intended to clarify that Rule 104(a)'s requirement applies to expert opinions under Rule 702. Similarly, nothing in the amendment requires the court to nitpick an expert's opinion in order to reach a perfect expression of what the basis and methodology can support. The Rule 104(a) standard does not require perfection. On the other hand, it does not permit the expert to make claims that are unsupported by the expert's basis and methodology.

Excerpt from the September 2022 Report of the Committee on Rules of Practice and Procedure

* * * * *

REPORT OF THE JUDICIAL CONFERENCE

COMMITTEE ON RULES OF PRACTICE AND PROCEDURE

TO THE CHIEF JUSTICE OF THE UNITED STATES AND MEMBERS OF THE JUDICIAL CONFERENCE OF THE UNITED STATES:

* * * * *

FEDERAL RULES OF EVIDENCE

Rules Recommended for Approval and Transmission

The Advisory Committee on Evidence Rules recommended for final approval proposed amendments to Evidence Rules 106, 615, and 702.

Rule 106 (Remainder of or Related Writings or Recorded Statements)

The proposed amendment to Rule 106 – the rule of completeness – would allow any completing statement to be admitted over a hearsay objection and would cover all statements, whether or not recorded. The overriding goal of the amendment is to treat all questions of completeness in a single rule. That is particularly important because completeness questions often arise at trial, and so it is important for the parties and the court to be able to refer to a single rule to govern admissibility. The amendment is intended to displace the common law, just as the common law has been displaced by all of the other Federal Rules of Evidence.

The Advisory Committee received only a few public comments on the proposed changes to Rule 106. As published, the amendment would have inserted the words "written or oral" before "statement" so as to address the rule's applicability to unrecorded oral statements. After public comment, the Advisory Committee deleted the phrase "written or oral" to make clear that

NOTICE
NO RECOMMENDATIONS PRESENTED HEREIN REPRESENT THE POLICY OF THE JUDICIAL CONFERENCE UNLESS APPROVED BY THE CONFERENCE ITSELF.

Rule 106 applies to all statements, including statements – such as those made through conduct or through sign language – that are neither written nor oral.

Rule 615 (Excluding Witnesses)

The proposed amendments to Rule 615 would limit an exclusion order under the existing rule (which would be re-numbered Rule 615(a)) to exclusion of witnesses from the courtroom, and would add a new subdivision (b) that would provide that the court has discretion to issue further orders to "(1) prohibit disclosure of trial testimony to witnesses who are excluded from the courtroom; and (2) prohibit excluded witnesses from accessing trial testimony." Under the proposed amendments, if a court wants to do more than exclude witnesses from the courtroom, the court must so order. In addition, the proposed amendments would clarify that the existing provision that allows an entity-party to designate "an officer or employee" to be exempt from exclusion is limited to one officer or employee. The rationale is that the exemption is intended to put entities on par with individual parties, who cannot be excluded under Rule 615. Allowing the entity more than one exemption is inconsistent with that rationale. In response to public comments, the Advisory Committee made two minor changes to the committee note (replacing the word "agent" with the word "representative" and deleting a case citation). The Standing Committee, in turn, revised three sentences in the committee note (including the sentence addressing orders governing counsel's disclosure of testimony for witness preparation).

Rule 702 (Testimony by Expert Witnesses)

The proposed amendments to Rule 702's first paragraph and to Rule 702(d) are the product of Advisory Committee work dating back to 2016. As amended, Rule 702(d) would require the proponent to demonstrate to the court that "the expert's opinion reflects a reliable application of the principles and methods to the facts of the case." This language would more clearly empower the court to pass judgment on the conclusion that the expert has drawn from the

methodology. In addition, the proposed amendments as published would have required that "the proponent has demonstrated by a preponderance of the evidence" that the requirements in Rule 702(a) – (d) have been met. This language was designed to reject the view of some courts that the reliability requirements set forth in Rule 702(b) and (d) – that the expert has relied on sufficient facts or data and has reliably applied a reliable methodology to the facts – are questions of weight and not admissibility, and more broadly that expert testimony is presumed to be admissible. With this language, the Advisory Committee sought to explicitly weave the Rule 104(a) standard into the text of Rule 702.

More than 500 comments were received on the proposed amendments to Rule 702. In addition, a number of comments were received at a public hearing. Many of the comments opposed the amendment, and the opposition was especially directed toward the phrase "preponderance of the evidence." Another suggestion in the public comment was that the rule should clarify that it is the court and not the jury that must decide whether it is more likely than not that the reliability requirements of the rule have been met. The Advisory Committee carefully considered the public comments and determined to replace "the proponent has demonstrated by a preponderance of the evidence" with "the proponent demonstrates to the court that it is more likely than not" that the reliability requirements are met. The Advisory Committee also made a number of changes to the committee note, and the Standing Committee, in its turn, made one minor edit to the committee note.

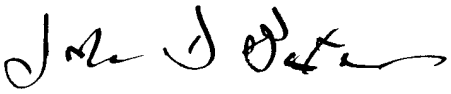
After making the changes, noted above, to the committee notes for Rules 615 and 702, the Standing Committee unanimously approved the proposed amendments to Rules 106, 615, and 702.

> **Recommendation:** That the Judicial Conference approve the proposed amendments to Evidence Rules 106, 615, and 702, as set forth in Appendix E, and transmit them to the Supreme Court for consideration with a recommendation that

they be adopted by the Court and transmitted to Congress in accordance with the law.

* * * * *

Respectfully submitted,



John D. Bates, Chair

Elizabeth J. Cabraser	Troy A. McKenzie
Jesse M. Furman	Patricia Ann Millett
Robert J. Giuffra, Jr.	Lisa O. Monaco
Frank Mays Hull	Gene E.K. Pratter
William J. Kayatta, Jr.	Kosta Stojilkovic
Peter D. Keisler	Jennifer G. Zipps
Carolyn B. Kuhl	

* * * * *

COMMITTEE ON RULES OF PRACTICE AND PROCEDURE
OF THE
JUDICIAL CONFERENCE OF THE UNITED STATES
WASHINGTON, D.C. 20544

JOHN D. BATES
CHAIR

CHAIRS OF ADVISORY COMMITTEES

JAY S. BYBEE
APPELLATE RULES

DENNIS R. DOW
BANKRUPTCY RULES

ROBERT M. DOW, JR.
CIVIL RULES

RAYMOND M. KETHLEDGE
CRIMINAL RULES

PATRICK J. SCHILTZ
EVIDENCE RULES

MEMORANDUM

TO: Honorable John D. Bates, Chair
Standing Committee on Rules of Practice and Procedure

FROM: Honorable Patrick J. Schiltz, Chair
Advisory Committee on Evidence Rules

RE: Report of the Advisory Committee on Evidence Rules

DATE: May 15, 2022

I. Introduction

The Advisory Committee on Evidence Rules (the "Committee") met in Washington, D.C., on May 6, 2021. At the meeting the Committee discussed and gave final approval to three proposed amendments that had been released for public comment. The Committee also considered and approved six proposed amendments with the recommendation that they be released for public comment.

The Committee made the following determinations at the meeting:

● It unanimously approved proposed amendments to Rules 106, 615, and 702, and recommends to the Standing Committee that they be transmitted to the Judicial Conference.

* * * * *

A full description of all of these matters can be found in the draft minutes of the Committee meeting, attached to this Report. The proposed amendments can also be found as attachments to this Report.

II. Action Items

A. Proposed Amendment to Rule 106, for Final Approval

At the suggestion of Judge Paul Grimm, the Committee has for the last five years considered and discussed whether Rule 106 --- the rule of completeness --- should be amended. Rule 106 provides that if a party introduces all or part of a written or recorded statement in such a way as to be misleading, the opponent may introduce a completing statement that would correct the misimpression. The Committee has considered whether Rule 106 should be amended in two respects: 1) to provide that a completing statement is admissible over a hearsay objection; and 2) to expand the rule to cover unrecorded oral statements, as well as written and recorded statements.

The courts are not uniform in their treatment of these issues. On the hearsay question, some courts have held that when a party introduces a portion of a statement that is misleading, that party can still object, on hearsay grounds, to completing evidence that corrects the misimpression. Other courts have held essentially that if a party introduces a portion of a statement in a manner that misleads the factfinder, that party forfeits the right to object to introduction of other portions of that statement when that is necessary to remedy the misimpression. As to unrecorded oral statements, most courts have found that when necessary to complete, such statements are admissible either under Rule 611(a) or under the common law rule of completeness.

After much discussion and consideration, the Committee in Spring, 2021 unanimously approved an amendment for release for public comment. The proposal released for public comment allows the completing statement to be admitted over a hearsay objection and covers unrecorded oral statements.

The overriding goal of the amendment is to treat all questions of completeness in a single rule. That is particularly important because completeness questions often arise at trial, and so it is important for the parties and the court to be able to refer to a single rule to govern admissibility. What has been particularly confusing to courts and practitioners is that Rule 106 has been considered a "partial codification" of the common law --- meaning that the parties must be aware that common law may still be invoked. As stated in the Committee Note, the amendment is intended to displace the common law, just as the common law has been displaced by all of the other Federal Rules of Evidence.

As to admissibility of out-of-court statements, the amendment takes the position that the proponent, by introducing part of a statement in a misleading manner, forfeits the right to foreclose admission of a remainder that is necessary to remedy the misimpression. Simple notions of fairness, already embodied in Rule 106, dictate that a misleading presentation cannot stand unrebutted. The amendment leaves it up to the court to determine whether the completing remainder will be admissible to prove a fact (a hearsay use) or simply to provide context (a non-hearsay use). Either usage is encompassed within the rule terminology --- that the completing remainder is admissible "over a hearsay objection."

As to unrecorded oral statements, most courts already admit such statements when necessary to complete --- they just do so under a different evidence rule or under the common law. The Committee was convinced that covering unrecorded oral statements under Rule 106 would be a user-friendly change, especially because the existing hodgepodge of coverage of unrecorded statements presents a trap for the unwary. As stated above, the fact that completeness questions almost always arise at trial means that parties cannot be expected to quickly get an answer from the common law, or from a rule such as Rule 611(a) that does not specifically deal with completeness.

It is important to note that nothing in the amendment changes the basic rule, which applies only to the narrow circumstances in which a party has created a misimpression about the statement, and the adverse party proffers a completing statement that in fact corrects the misimpression. So, the mere fact that a statement is probative and contradicts a statement offered by the opponent is not enough to justify completion under Rule 106.

The Committee received only a few public comments on the proposed changes to Rule 106. All comments were in favor of the proposed amendment, with a couple of comments providing some suggestions for minor changes. After considering the public comment, the Committee unanimously approved a slight change to the proposal: deletion of the phrase "written or oral," which makes clear that Rule 106 applies to all statements, including those that are not written or oral. The Committee determined that statements made through conduct, or through sign language, should be covered by the rule of completeness, as there was no reason to distinguish such statements from those that are written or oral. The proposed Committee Note was slightly revised to accord with the change in text.

At its Spring 2022 meeting, the Committee unanimously gave final approval to the proposed amendment to Rule 106. The Committee recommends that the proposed amendment, and the accompanying Committee Note, be approved by the Standing Committee and referred to the Judicial Conference.

The proposed amendment to Rule 106, together with the proposed Committee Note, the GAP report, and the summary of public comment, is attached to this Report.

B. Proposed Amendment to Rule 615, for Final Approval

Rule 615 provides for court orders excluding witnesses so that they "cannot hear other witnesses' testimony." The Committee determined that there are problems raised in the case law

and in practice regarding the scope of a Rule 615 order: does it apply only to exclude witnesses from the courtroom (as stated in the text of the rule) or does it extend outside the confines of the courtroom to prevent prospective witnesses from obtaining or being provided trial testimony? Most courts have held that a Rule 615 order extends to prevent access to trial testimony outside of the courtroom, because exclusion from the courtroom is not sufficient to protect against the risk of witnesses tailoring their testimony after obtaining access to trial testimony. But other courts have read the rule as it is written.

After extensive consideration and research over four years, the Committee agreed on an amendment that would clarify the extent of an order under Rule 615. Committee members have noted that where parties can be held in contempt for violating a court order, due process requires that the order be clear if it seeks to do more than exclude witnesses from the courtroom. The Committee's investigation of this problem is consistent with its ongoing efforts to ensure that the Evidence Rules are keeping up with technological advancement, given the increased possibility of witness access to information about testimony through news, social media, YouTube, or daily transcripts.

At its Spring, 2021 meeting the Committee unanimously voted in favor of an amendment to Rule 615. That amendment, released for public comment in August, 2021, limits an exclusion order to just that --- exclusion of witnesses from the courtroom. But a new subdivision provides that the court has discretion to issue further orders to "(1) prohibit disclosure of trial testimony to witnesses who are excluded from the courtroom; and (2) prohibit excluded witnesses from accessing trial testimony." In other words, if a court wants to do more than exclude witnesses from the courtroom, the court must say so.

The Committee also considered whether an amendment to Rule 615 should address orders that prohibit counsel from referring to trial testimony while preparing prospective witnesses. The Committee resolved that any amendment to Rule 615 should not mention trial counsel in text, because the question of whether counsel can use trial testimony to prepare witnesses raises issues of professional responsibility and the right to counsel that are beyond the purview of the Evidence Rules. Judges must address these issues on a case-by-case basis.

Finally, the Committee approved an additional amendment to the existing provision that allows an entity-party to designate "an officer or employee" to be exempt from exclusion. There is some dispute in the courts about whether the entity-party is limited to one such exemption or is entitled to more than one. The amendment clarifies that the exemption is limited to one officer or employee. The rationale is that the exemption is intended to put entities on a par with individual parties, who cannot be excluded under Rule 615. Allowing the entity more than one exemption is inconsistent with that rationale.

As noted, these proposed changes to Rule 615 were released for public comment in August, 2021. Only a few public comments were received. All were supportive of the amendment, with two comments suggesting minor changes. In response to the public comment, the Committee made two minor changes the Committee Note to the proposed amendment.

At its Spring 2022 meeting, the Committee unanimously gave final approval to the proposed amendment to Rule 615. The Committee recommends that the proposed amendment, and the accompanying Committee Note, be approved by the Standing Committee and referred to the Judicial Conference.

The proposed amendment to Rule 615, together with the Committee Note, the GAP report, and the summary of public comment, is attached to this Report.

C. Proposed Amendment to Rule 702, for Final Approval

The Committee has been researching and discussing the possibility of an amendment to Rule 702 for five years. The project began with a Symposium on forensic experts and *Daubert,* held at Boston College School of Law in October, 2017. That Symposium addressed, among other things, the challenges to forensic evidence raised in a report by the President's Council of Advisors on Science and Technology. A Subcommittee on Rule 702 was appointed to consider possible treatment of forensic experts, as well as the weight/admissibility question discussed below. The Subcommittee, after extensive discussion, recommended against certain courses of action. The Subcommittee found that: 1) It would be difficult to draft a freestanding rule on forensic expert testimony, because any such amendment would have an inevitable and problematic overlap with Rule 702; and 2) It would not be advisable to set forth detailed requirements for forensic evidence either in text or Committee Note because such a project would require extensive input from the scientific community, and there is substantial debate about what requirements are appropriate.

The full Committee agreed with these suggestions. But the Subcommittee did express interest in considering an amendment to Rule 702 that would focus on one important aspect of forensic expert testimony --- the problem of overstating results (for example, an expert claiming that her opinion has a "zero error rate", where that conclusion is not supportable by the expert's methodology). The Committee heard extensively from DOJ on the important efforts it is now employing to regulate the testimony of its forensic experts, and to limit possible overstatement.

The Committee considered a proposal to add a new subdivision (e) to Rule 702 that would essentially prohibit any expert from drawing a conclusion overstating what could actually be concluded from a reliable application of a reliable methodology. But a majority of the members decided that the amendment would be problematic, because Rule 702(d) already requires that the expert must reliably apply a reliable methodology. If an expert overstates what can be reliably concluded (such as a forensic expert saying the rate of error is zero) then the expert's opinion should be excluded under Rule 702(d). The Committee was also concerned about the possible unintended consequences of adding an overstatement provision that would be applied to all experts, not just forensic experts.

The Committee, however, unanimously favored a slight change to existing Rule 702(d) that would emphasize that the court must focus on the expert's opinion, and must find that the opinion actually proceeds from a reliable application of the methodology. The Committee unanimously approved a proposal—released for public comment in August, 2021--- that would amend Rule 702(d) to require the court to find that "the expert's opinion reflects a reliable application of the principles and methods to the facts of the case." As the Committee Note

elaborates: "A testifying expert's opinion must stay within the bounds of what can be concluded by a reliable application of the expert's basis and methodology." The language of the amendment more clearly empowers the court to pass judgment on the conclusion that the expert has drawn from the methodology. Thus the amendment is consistent with *General Electric Co., v. Joiner,* 522 U.S. 136 (1997), in which the Court declared that a trial court must consider not only the expert's methodology but also the expert's conclusion; that is because the methodology must not only be reliable, it must be reliably applied.

Finally, the Committee resolved to respond to the fact that many courts have declared that the reliability requirements set forth in Rule 702(b) and (d) --- that the expert has relied on sufficient facts or data and has reliably applied a reliable methodology --- are questions of weight and not admissibility, and more broadly that expert testimony is presumed to be admissible. These statements misstate Rule 702, because its admissibility requirements must be established to a court by a preponderance of the evidence. The Committee concluded that in a fair number of cases, the courts have found expert testimony admissible even though the proponent has not satisfied the Rule 702(b) and (d) requirements by a preponderance of the evidence --- essentially treating these questions as ones of weight rather than admissibility, which is contrary to the Supreme Court's holdings that under Rule 104(a), admissibility requirements are to be determined by court under the preponderance standard.

Initially, the Committee was reluctant to propose a change to the text of Rule 702 to address these mistakes as to the proper standard of admissibility, in part because the preponderance of the evidence standard applies to almost all evidentiary determinations, and specifying that standard in one rule might raise negative inferences as to other rules. But ultimately the Committee unanimously agreed that explicitly weaving the Rule 104(a) standard into the text of Rule 702 would be a substantial improvement that would address an important conflict among the courts. While it is true that the Rule 104(a) preponderance of the evidence standard applies to Rule 702 as well as other rules, it is with respect to the reliability requirements of expert testimony that many courts are misapplying that standard. Moreover, it takes some effort to determine the applicable standard of proof --- Rule 104(a) does not mention the applicable standard of proof, requiring a resort to case law. And while *Daubert* mentions the standard, *Daubert* does so only in a footnote in the midst of much discussion about the liberal standards of the Federal Rules of Evidence. Consequently, the Committee unanimously approved an amendment for public comment that would explicitly add the preponderance of the evidence standard to Rule 702(b)-(d). The language of the proposal released for public comment required that "the proponent has demonstrated by a preponderance of the evidence" that the reliability requirements of Rule 702 have been met. The Committee Note to the proposal made clear that there is no intent to raise any negative inference regarding the applicability of the Rule 104(a) standard of proof to other rules --- emphasizing that incorporating the preponderance standard into the text of Rule 702 was made necessary by the decisions that have failed to apply it to the reliability requirements of Rule 702.

More than 500 comments were received on the proposed amendments to Rule 702. In addition, a number of comments were received at a public hearing held on the rule. Many of the comments were opposed to the amendment, and almost all of the fire was directed toward the term "preponderance of the evidence." Some thought that "preponderance of the evidence" would limit the court to considering only *admissible* evidence at the *Daubert* hearing. Others thought that the

term represented a shift from the jury to the judge as factfinder. By contrast, commentators who supported the amendment argued that the amendment should go further and clarify that it is the court, not the jury, that decides admissibility.

The Committee carefully considered the public comments. The Committee does not agree that the preponderance of the evidence standard would limit the court to considering only admissible evidence; the plain language of Rule 104(a) allows the court deciding admissibility to consider inadmissible evidence. Nor did the Committee believe that the use of the term preponderance of the evidence would shift the factfinding role from the jury to the judge, for the simple reason that, when it comes to making preliminary determinations about admissibility, the judge *is* and *always has been* a factfinder.

But while disagreeing with these comments, the Committee recognized that it would be possible to replace the term "preponderance of the evidence" with a term that would achieve the same purpose while not raising the concerns (valid or not) mentioned by many commentators. The Committee unanimously agreed to change the proposal as issued for public comment to provide that the proponent must establish that it is "*more likely than not*" that the reliability requirements are met. This standard is substantively identical to "preponderance of the evidence" but it avoids any reference to "evidence" and thus addresses the concern that the term "evidence" means only admissible evidence.

The Committee was also convinced by the suggestion in the public comment that the rule should clarify that it is the court and not the jury that must decide whether it is more likely than not that the reliability requirements of the rule have been met. Therefore, the Committee unanimously agreed with a change requiring that the proponent establish "*to the court*" that it is more likely than not that the reliability requirements have been met. The proposed Committee Note was amended to clarify that nothing in amended Rule 702 requires a court to make any findings about reliability in the absence of a proper objection.

With those changes, and a few stylistic and corresponding changes to the Committee Note, the Committee unanimously voted in favor of adopting the amendments to Rule 702, for final approval.

At the Spring 2022 meeting, the Committee unanimously gave final approval to the proposed amendment to Rule 702. The Committee recommends that the proposed amendment, and the accompanying Committee Note, be approved by the Standing Committee and referred to the Judicial Conference.

The proposed amendment to Rule 702, together with the proposed Committee Note, GAP report, summary of public comment, and summary of the public hearing, is attached to this Report.

* * * * *

EXHIBIT 5

UNITED STATES BANKRUPTCY COURT
DISTRICT OF MASSACHUSETTS

In re: TELEXFREE, LLC, TELEXFREE, INC., and TELEXFREE FINANCIAL, INC., Debtors	Chapter 11 Cases 14–40987–MSH 14–40988–MSH 14–40989–MSH Jointly Administered
STEPHEN B. DARR, TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Plaintiff, v. BENJAMIN ARGUETA, A REPRESENTATIVE OF A CLASS OF DEFENDANT NET WINNERS, Defendants.	Adversary Proceeding No. 16-4006
STEPHEN B. DARR AS TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Plaintiff, v. MARCO PUZZARINI, A REPRESENTATIVE OF A CLASS OF DEFENDANT NET WINNERS, Defendants.	Adversary Proceeding No. 16-4007

REBUTTAL EXPERT REPORT OF JOSHUA W. DENNIS

Respectfully submitted this 31st day of July, 2020

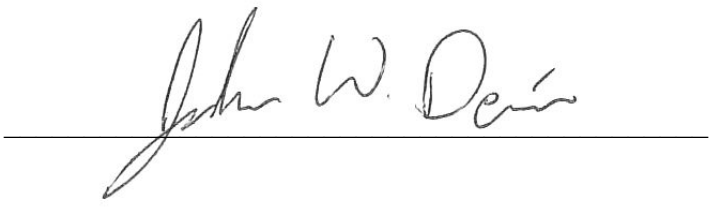


Table of Contents

I. INTRODUCTION 1
II. BACKGROUND 2
A. TelexFree User Accounts and Participants 2
B. TelexFree Transactions 4
III. ASSIGNMENT 7
IV. INFORMATION CONSIDERED 7
V. UNDERSTANDING OF TRUSTEE'S METHODOLOGY FOR DETERMINING NET WINNERS AND NET LOSERS 9
A. Trustee Aggregation Process 9
B. Trustee Net Equity Calculation 13
VI. SUMMARY OF ANALYSIS AND FINDINGS 14
A. The Aggregation Process is Unreliable 16
i. Mr. Martin's Application of Deterministic Record Linkage to the Aggregation Algorithm Does to Provide Reliable Results 16
ii. The ePOC Data Does Not Validate the Aggregation Process 20
iii. Aggregation Errors Can Result in a Significant Overstatement of Participant Net Equity 25
B. The Calculation of Net Equity is Unreliable 28
i. The Net Equity Calculation Fails to Account for Participant Investments and Receipts Related to Transfers of Credit 28
ii. The Amount of Cash Invested/Received by Participants as Part of Triangular Transactions or Credits Transfers Can Differ from Amount of Credits Recorded in SIG 32
VII. CONCLUSION 37

List of Exhibits

Exhibit Number	Description
1	Curriculum Vitae of Joshua W. Dennis
2	Documents Considered
3	User Accounts and Participants by Entity and Type
4	Summary of TelexFree Net Equity by Tier and Transaction Type
5	Summary of TelexFree User Accounts and Participants by Tier
6	Summary of TelexFree Top Net Loser Matched to Net Winners
7	Top 100 TelexFree Net Losers
8	Top 100 TelexFree Net Winners
9.1-9.31	Top Net Loser and Matched Net Winner User Account Detail
10	Frantz Balan and Lusette Balan User Account Detail

I. INTRODUCTION

1. I, Joshua W. Dennis, am a Partner in the Boston office of StoneTurn Group LLP ("StoneTurn"). I provide financial and economic consulting services – including the analysis of complex data – to attorneys and companies. I have more than 15 years of experience working with clients and counsel on matters requiring complex data analysis, such as economic damages, forensic accounting, compliance monitoring, and valuation engagements.

2. In addition to experience in data collection, mining, anomaly detection, and reporting of transactional and trading data, I have extensive expertise in financial modeling to quantify and assess lost profits, diminution of business value, and other areas of economic loss. My analyses also frequently include evaluations related to data sufficiency and integrity testing, integration of disparate datasets, data exploration, and visual modeling.

3. In conjunction with my experience related to the quantification of economic damages, I have gained expertise through my certifications and trainings. I have earned the Certified Valuation Analyst ("CVA")[1] designation granted by the National Association of Certified Valuation Analysts ("NACVA")[2] and a Certificate in Intellectual Property Law from Suffolk University.

4. A copy of my current curriculum vitae, which summarizes my qualifications and professional experience, is attached as **Exhibit 1.** I have co-authored one article within the last ten years: *The Role of Data Analytics in Regulatory Inquiries*, The Society of Corporate Compliance and Ethics (SCCE), March 2020.

5. StoneTurn charges $425 per hour for time I spend consulting and developing expert opinions, as well as time that may be spent testifying related to my opinions. Neither my, nor my

[1] NACVA trains and certifies CVAs to perform business valuations as a service to both the consulting community and the users of their services. Through training and rigorous testing, CVAs demonstrate they are qualified to provide capable and professionally executed valuation services. NACVA'S valuation designations are the only valuation credentials accredited by the National Commission for Certifying Agencies. (www.nacva.com.)

[2] NACVA, an organization comprised of more than 7,000 CPAs and other valuation and consulting professionals, supports the users of business and intangible asset valuation services and financial forensic services, including damages determinations of all kinds and fraud detection and prevention, by training and certifying financial professionals in these disciplines. (www.nacva.com.)

staff's, nor StoneTurn's, compensation is affected in any way by either the conclusions I have reached or the outcome of this lawsuit.

II. BACKGROUND

6. On June 6, 2014, the Plaintiff, Stephen Darr, was appointed Chapter 11 Trustee ("Trustee") of the bankruptcy estates of TelexFree, LLC, TelexFree, Inc., and TelexFree Financial, Inc., (collectively, the "Company," "TelexFree," or "Debtors").[3] The Trustee alleges that TelexFree operated a massive Ponzi and pyramid scheme that defrauded hundreds of thousands of individuals in numerous countries out of billions of dollars.[4]

7. The Trustee alleges that roughly 15,000 individuals or entities residing in the U.S. are "Net Winners"[5] who illegally profited from the fraud. These 15,000 Net Winners comprise a class of Defendants ("U.S. Net Winner Class").[6] Additionally, the Trustee alleges that roughly 80,000 individuals or entities residing outside the U.S. are "Net Winners" who illegally profited from the fraud.[7] These 80,000 Net Winners comprise a class of defendants ("Non-U.S. Net Winner Class," collectively with the U.S. Net Winner Class, hereinafter referred to as the "Net Winner Class").

A. TelexFree User Accounts and Participants

8. I understand that an individual could join TelexFree (as a "Participant") by purchasing either a membership plan or individual voice over Internet ("VoIP") phone packages. However, VoIP phone packages, which cost $49.90 per package,[8] constituted only a minor portion of the Company's business, and TelexFree primarily generated returns from the membership fees associated with recruiting new Participants.[9]

[3] Amended Complaint, Adversary Proceeding No. 16-04006, April 14, 2016 ("Amended Complaint") at 3.

[4] Amended Complaint at 21.

[5] For the specific definition of Net Winner, see Section V of this report.

[6] Class Certification Order and Approval of Class Counsel, Adversary Proceeding No. 16-4006, October 6, 2016 ("Class Cert. 16-4006") at 3.

[7] Class Certification Order and Approval of Class Counsel, Adversary Proceeding No. 16-4007, August 3, 2017 ("Class Cert. 16-4007") at 2.

[8] Motion by Chapter 11 Trustee for Entry of Order Finding that Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, October 7, 2015 ("Motion for Entry of Order that Debtors Engaged in Ponzi Scheme") at 6.

[9] Amended Complaint at 21.

9. Each time a Participant purchased a membership plan or VoIP Package, the Participant established a "User Account" with the Company and was directed to provide a name, address, email address, phone numbers, passcodes, and certain other information.

10. The User Accounts were used to track the activity of the Participant, including "Credits" issued to the Participant for becoming a "Promoter" of the business by placing advertisements on certain websites, as well as recruiting additional Participants. The User Accounts were also used to track certain other activities, such as bonuses, commissions, payments of invoices, use of Credits to satisfy invoices, transfers of Credits, and cash receipts. A Participant could establish more than one User Account, and in practice many Participants had multiple User Accounts, sometimes numbering in the thousands.[10]

11. The User Accounts and related data were stored in TelexFree's Management Information System (i.e., SIG).[11] The SIG database contained data for approximately 17 million User Accounts, of which approximately 15 million were associated with at least one paid invoice.[12]

12. I understand that an affiliated entity of TelexFree, Ympactus Commercial Ltda. ("Ympactus"), reportedly operated a similar scheme in Brazil, and also stored its data in SIG. I understand Ympactus was seized by the Brazilian authorities in June 2013.[13] To assist in separating out the Ympactus data from the TelexFree data within SIG, I understand that the currency field was utilized, which identified whether an invoice was denominated in Brazilian reais ("Reais" or "BRL") or United States dollars ("USD"). I understand that User Accounts with Brazilian contact information were denominated in Reais, and thus associated with Ympactus. Conversely, invoices associated with User Accounts with non-Brazilian contact information were denominated in USD and assumed to be associated with TelexFree. Approximately 11 million User Accounts with at least one paid invoice remained after excluding the User Accounts associated with Ympactus.[14]

[10] Amended Complaint at 22-23. Also, see e.g., **Exhibit 7** and **Exhibit 8**.

[11] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 11. ("SIG stands for Sistemas de Informacoes Gerenciais, which is from the Portuguese language and translates roughly to 'Information Management System.'")

[12] Martin Report at 11. Also, see **Exhibit 3**.

[13] Martin Report at 7; Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 14-16.

[14] Martin Report at 11. Also, see **Exhibit 3**.

13. The Company provided Participants with two membership plan options, each of which allowed for the generation of Credits. A one-year membership plan could be purchased by Participants for $339 (for an AdCentral Plan) or $1,425 (for an AdCentral Family Plan), as follows:[15]

- **"AdCentral Plan":** $339 for a one-year contract ($50 membership fee plus $289 contract fee). This contract entitled the User Account holder with the right to sell ten VoIP Packages, for which a Participant could receive a commission if the packages were sold. Participants were required to place one internet ad per day and, for each complete week in which the Participant placed the required ads, he/she was entitled to one additional VoIP Package, which could be sold or exchanged for $20 in credits with the Debtors. Thus, Participants who posted the required ads were eligible to receive $20 per week for 52 weeks, for a total return of $1,040.

- **"AdCentral Family Plan":** $1,425 for a one-year contract ($50 membership fee plus $1,375 contract fee). This contract entitled the User Account holder with the right to sell fifty VoIP Packages, for which a Participant could receive a commission if the packages were sold. Participants were required to place five internet ads per day and, for each week in which the Participant placed the required ads, he/she was entitled to five additional VoIP Packages, which could be sold or exchanged for $100 in credits with the Debtors. Thus, the Participants who posted the required ads were eligible to receive $100 per week for 52 weeks, for a total return of $5,200.

B. TelexFree Transactions

14. I understand that Invoices issued by the Company to Participants for the purchase of a membership plan could be satisfied in at least the following ways. A Participant could satisfy the invoice by payment in cash directly to the Company (a "Direct Transaction"), or a Participant could satisfy the TelexFree invoice in other ways, such as by having the recruiting Participant pay the invoice by redeeming his/her accumulated Credits, and TelexFree marked the invoice as paid and reduced the recruiting Participant's Credits by a corresponding amount (a "Triangular Transaction"). The Trustee alleges that Participants frequently brought on new recruits through

[15] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 7-8; Memorandum in Support of Proposed Supplemental Order Respecting Motion by Chapter 11 Trustee for Entry of Order Finding that Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, January 19, 2016 ("Support of Proposed Supplemental Order") at 3.

Triangular Transactions as one means to monetize their accumulated Credits, often in place of redeeming Credits directly with the Company for cash.[16]

15. I understand that while invoices associated with the sale of membership plans or VoIP Packages had a face value of approximately $3.07 billion, only $360 million, or roughly twelve percent (12%) of that amount, was paid in cash to TelexFree. The balance of these invoices was satisfied using Participants' Credits.[17]

16. In the case of a Direct Transaction where a Participant satisfies his/her own invoice by payment in cash to the Company, the process worked, generally, as follows:[18]

- The new Participant established an online account;
- The Company's database recorded the data entered by the new Participant and the identity of the recruiting Participant, and assigned an identification number to the newly created User Account;
- The Company recorded the purchase, issued a numbered invoice, and marked the invoice as 'pending';
- The new Participant could pay the invoice by cash, check, cashier's check, money order, wire transfer, or through a third-party online payment processing account. When the invoice was paid, the Company would update the invoice;
- The new Participant could then start building a pyramid underneath the newly created User Account by recruiting other members (or by purchasing new User Accounts himself/herself or through associates), which in turn generated bonuses and commissions for the new Participant.

17. In the case of a Triangular Transaction where a new User Account is opened using accumulated Credits in another User Account, the process worked, generally, as follows:[19]

- A recruited Participant created a new online account, or used his/her existing account information, to establish a new User Account;

[16] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 9-10; Support of Proposed Supplemental Order at 3-4.

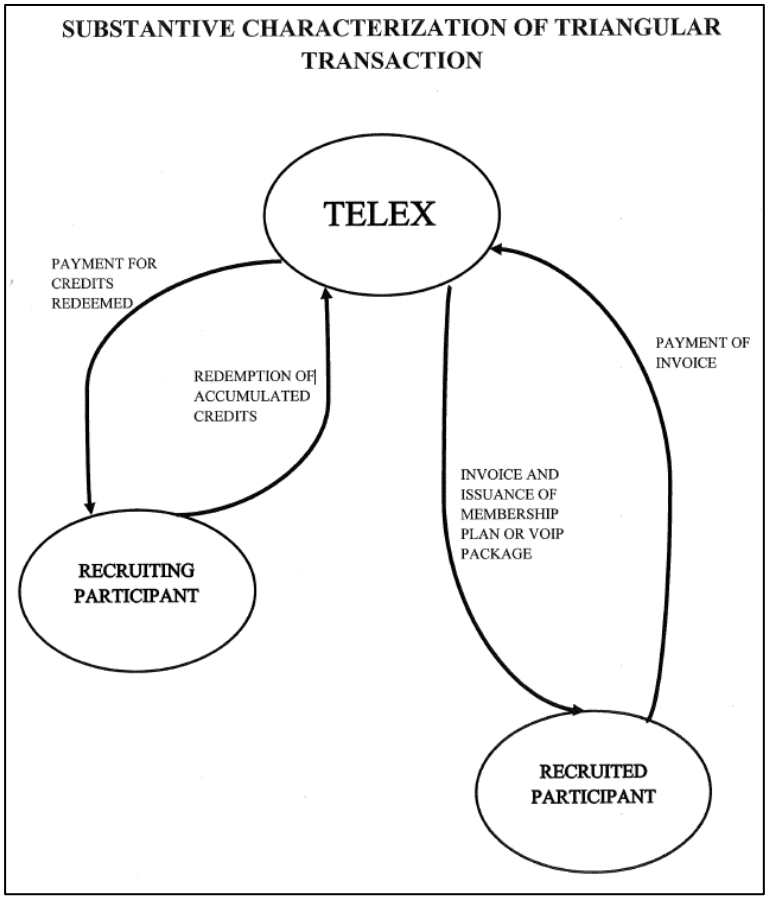
[17] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 10.

[18] Amended Complaint at 23-24.

[19] Amended Complaint at 24.

- The Company's database recorded the details entered by the recruited Participant and assigned an identification number to the new User Account;
- The Company recorded the purchase, issued an invoice to the new User Account, and marked the invoice as 'pending';
- The new recruited Participant forwarded the invoice to the recruiting Participant or a to a third-party Participant, who satisfied the invoice with accumulated Credits in the existing User Account;
- The Trustee alleges that the new recruited Participant always paid the invoice amount to the recruiting Participant (in those cases where there were two separate Participants involved). However, no evidence of such payments exists in SIG.

18. A graphical illustration of a Triangle Transaction is reflected below:[20]



[20] Support of Proposed Supplemental Order, Exhibit C.

III. ASSIGNMENT

19. StoneTurn[21] has been retained by Milligan Rona Duran & King LLC ("Counsel") to assess the methodology developed and employed the Trustee in calculating the supposed net funds received by members of the Net Winner Class as a result of their participation in TelexFree.

20. My opinions are based on consideration of case-specific facts and circumstances, as well as on my education, training, and experience. I am not offering any legal opinions.

21. While my work on this matter is on-going, this report summarizes my current opinions given the information available to date. I may consider any additional materials if produced subsequent to the issuance of this report and modify or supplement my analysis or opinions as necessary. I also reserve the right to address any comments related to my work in this matter, as outlined in this report and associated exhibits, that may be received from an expert retained by the Trustee.

22. In connection with my anticipated trial testimony in this action, I may create, from various documents produced in this litigation, demonstrative exhibits which refer or relate to the matters discussed in this report, or in my deposition testimony (if I were to be deposed). I have not yet created any such exhibits as of the date of this report.

IV. INFORMATION CONSIDERED

23. I have considered information from a variety of sources, including (i) information produced or filed by the parties in this proceeding, (ii) information provided by relevant parties, and (iii) information from independent research.

24. I have considered information produced in these proceedings, including legal filings.

25. Additionally, I have considered information provided by relevant parties, including reports submitted by experts retained on behalf of the Trustee in this matter, such as the Expert Report of

[21] As used herein, other than references to my education and experience, "I" ,"We" or "StoneTurn" shall mean either me personally or those StoneTurn professionals acting under my supervision. Also, "My", "Our", and "Us" shall refer to actions taken by me personally or by those StoneTurn professionals acting under my direct supervision. All of the conclusions and opinions expressed herein are my own.

Timothy J. Martin, April 13, 2018 ("Martin Report"). I have also considered the following information provided by Mr. Martin:

- TelexFree Sistemas de Informacoes Gerenciais ("SIG") data tables, the management information systems used by TelexFree to track the activity for Participants by User Account.
- TelexFree data tables utilized and modified by Mr. Martin as part of his Aggregation Process and Net Equity computation[22] ("Huron Data Tables"), including:[23]
    - **rep_id_summary:** This table contains participant account and aggregation information developed by Mr. Martin for the Trustee;
    - **claim_statement:** This table contains the transactions extracted from the TelexFfree database, which were then used to determine and calculate Net Equity;
    - **telexfree_accounts:** This table contains the same participant account and aggregation information, but excludes the Ympactus accounts; and
    - **telexfree_transactions:** This table contains the transactions extracted from the TelexFree database to calculate Net Equity, but includes only transactions related to participant Net Equity.
- Electronic Proof of Claim ("ePOC") data tables containing claims and other related information provided by Participants through a portal developed by the Trustee and his advisors;
- A summary description of the components used to calculate Net Equity, the mechanism for computing Net Equity using the TelexFree database, and the method used to aggregate User Accounts;[24] and
- Data summaries of Mr. Martin's concluded Net Equity by Participant and User Account.[25]

26. Lastly, I have also considered information obtained through independent research, including trade press articles and other publicly available sources.

[22] For the specific definition of Aggregation Process and Net Equity, see Section V of this report, below.

[23] Transmittal letter from Andrew Lizotte to Ilyas Rona, January 5, 2017.

[24] Transmittal letter from Andrew Lizotte to Ilyas Rona, January 5, 2017; Telexfree Net Equity Calculation and Aggregation.pdf.

[25] TelexFree - User Account Detail.xlsx; Domestic and Foreign Net Winners v2.xlsx.

27. The information I considered is listed in **Exhibit 2**, as well as the footnotes and other citations in this report. I understand that discovery is ongoing in this matter. I reserve the ability to consider further discovery in this case, including additional expert reports, transcripts of depositions, and documents produced or made available.

V. UNDERSTANDING OF TRUSTEE'S METHODOLOGY FOR DETERMINING NET WINNERS AND NET LOSERS

28. I understand that the Trustee has retained Huron Consulting Services LLC ("Huron"), including Mr. Timothy Martin, a Managing Director in the Business Advisory Practice of Huron, to act as a financial and accounting advisor to the Trustee.[26] As part of this retention, I understand that Huron and Mr. Martin assisted in the development of a methodology that attempts to identify and quantify TelexFree "Net Winners" and "Net Losers," as defined herein.

29. I understand that Mr. Martin considers a Participant to be a "Net Winner" when the net amount of funds received by the Participant exceeds the amounts expended. Conversely, Mr. Martin considers a Participant to be a "Net Loser" when the net amount of funds expended exceeds the amounts received. The nominal dollar value, whether a net gain or net loss, associated with a given Participant is his/her "Net Equity," which I understand the Court in this matter has defined as follows:[27]

> The amount invested by a Participant into the Debtors, including amounts paid by the Participant pursuant to the Triangular Transactions, less amounts received by a Participant from the Debtors' scheme, including amounts received by the Participant pursuant to the Triangular Transactions.

30. I understand that the Court's definition of Net Equity does not include the value of labor for the time spent by Participants or any expense incurred by Participants (e.g., rent, conferences, travel, etc.) resulting from their involvement in TelexFree.

A. Trustee Aggregation Process

31. I understand that in order to analyze the data and determine who is a Net Winner or Net Loser based on Mr. Martin's definitions, Huron developed a computer program (the "Aggregation

[26] Martin Report at 3-4.

[27] Supplemental Order Respecting Motion by Chapter 11 Trustee for Entry of Order Finding that Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, January 26, 2016 ("Supplemental Order"), at 2-3.

Algorithm") to identify owners of TelexFree accounts ("User Accounts") and supposedly aggregate multiple User Accounts by Participant (the "Aggregation Process").

32. The Aggregation Process utilizes information supplied by Participants when registering User Accounts, such as name, email address, street address, cell phone number, and other information, which was stored in SIG ("Potential Identifiers").[28] Specifically, I understand that the Aggregation Algorithm relies upon Mr. Martin's application of a "Deterministic Record Linkage" process using the follow data fields ("Aggregation Fields"):[29]

Table 1: Aggregation Fields

Field	Description
Name	Name (usually full name) as entered into SIG
Email Address	Email address associated with User Account in SIG
Home Phone	Home phone number entered into SIG
Cell Phone	Cell phone number entered into SIG
Physical Address	Street address entered into SIG
Log-In	Unique one-word name created by Participant
Password	User created Password used to access User Account details
Rep ID	Unique number generated by SIG

33. Translated to the actual field names per the rep_id_summary table, Mr. Martin uses the following: [30]

[28] Martin Report at 3.

[29] Mr. Martin describes Deterministic Record Linkage as "generat[ing] links based on the number of individual identifiers that match among the available data sets. Two records are considered a match under a deterministic record linkage procedure if all, or a predetermined number of identifiers are identical." (Martin Report at 14-15).

[30] Martin Report, Addendum A; Telexfree Net Equity Calculation and Aggregation.pdf.

TelexFree/SIG Field	Description	Additional Information	Trustee Designation For Modified data from TelexFree/SIG Field ("Aggregation Field")
rep_nome	Full Name	Aggregation field removes spaces and other extraneous characters and excludes names with fewer than five characters.	name_key
rep_email	Email Address	Aggregation field identifies properly formatted permanent email addresses and excludes other entries.	email_key
rep_cel	Mobile Phone	Aggregation field removes non-numeric values and excludes entries with fewer than eight digits.	cel_key
rep_fone	Home Phone	Aggregation field removes non-numeric values and excludes entries with fewer than eight digits.	fone_key
rep_login	Login	Aggregation field only uses first three characters of login.	login_key
rep_end	Address 1	Aggregation field only uses first five characters of address field with spaces removed.	end_key
rep_pwd_secondary	Secondary Password	Aggregation field uses the Secondary Password field where a password was provided.	hash_key
N/A	N/A	Aggregation field is the first four characters of the name_key field.	part_name_key
N/A	N/A	Aggregation field is the first character of the name_key field concatenated with the last four characters of the name_key field.	part_name_key2
N/A	N/A	Aggregation field is the last five digits of the cel_key field concatenated with the last five digits of the fone_key field.	part_phone_key

34. Other Potential Identifiers available in the rep_id_summary table that do not appear to have been utilized by Mr. Martin as part of the Aggregation Algorithm include: house number (rep_numero);[31] neighborhood (rep_barrio); city (rep_cidade); state_code (rep_uf); zip code (rep_cep); and country code (rep_pais). Additionally, Potential Identifiers available in the telexfree_accounts table that do not appear to have been utilized by Mr. Martin as part of the Aggregation Algorithm include: social security number (cpf_id); date of birth (date_of_birth); and company name (company_name). The Martin Report neither discusses why these additional Potential Identifiers were not utilized as part of the Aggregation Algorithm, nor the extent to which their inclusion may have impacted the overall results of Aggregation Process.

35. Mr. Martin uses his selected Aggregation Fields as part of a 13-step Aggregation Process, which attempts to confirm which User Accounts belong to the same Participant based upon

[31] The "rep_end" field the Mr. Martin utilizes for the Address within the Aggregation Algorithm only starts with a numerical value (i.e., street number) 19% of the time for TelexFree User Accounts. By comparison, "rep_numero" field (i.e., the separate field used to capture street number) is filled in 78% of the time for TelexFree User Accounts. (Based on queries of Huron Data Tables.)

matching various combinations of Aggregation Fields, or reformatted version thereof, as reflected in the table below:[32]

Aggregation Step	Aggregation Fields Used
1	name_key, email_key
2	name_key, cel_key
3	name_key, fone_key
4	name_key, part_phone_key
5	name_key, login_key, end_key,
6	part_name_key, email_key, part_phone_key
7	part_name_key, part_phone_key, login_key, end_key
8	part_name_key, email_key, login_key, end_key
9	part_name_key, login_key, hash_key
10	part_name_key2, email_key, part_phone_key
11	part_name_key2, part_phone_key, login_key, end_key
12	part_name_key2, email_key, login_key, end_key
13	part_name_key2, login_key, hash_key

36. I understand that the Aggregation Algorithm was performed using the above criteria sequentially, starting with User Accounts sharing the same full name and conforming email address. Step 2 was then run independently of Step 1, and where an overlap of two or more User Accounts existed between the Aggregations in Step 1 and Step 2 the User Accounts were combined into a new aggregation. This process continued through all 13 steps.[33]

37. As a result of the Aggregation Process employed by Mr. Martin, I understand that approximately 14.99 million User Accounts were consolidated to approximately 3.33 million Participants. Removal of the Ympactus User Accounts resulted in 10.99 million TelexFree User Accounts associated with 2.14 million Participants.[34]

38. The result of the Trustee's Aggregation Process based upon the Huron Data Tables provided by Mr. Martin is summarized by entity below:[35]

[32] Martin Report, Addendum A; Telexfree Net Equity Calculation and Aggregation.pdf.
[33] Martin Report at 15-18, Addendum A.
[34] Martin Report, Addendum A at 2.
[35] See **Exhibit 3**.

Entity	User Accounts	Participants	Net Equity
Hybrid[36]	138	113	$8,279,788
TelexFree	10,987,607	2,137,074	($226,190,248)
Ympactus	4,006,249	1,234,149	($425,439,930)
TOTAL	14,993,994	3,371,336	($643,350,389)
Unique Count	14,993,994	3,327,992	($643,350,389)

B. Trustee Net Equity Calculation

39. In the determination of Net Equity ("Net Equity Calculation"), I understand the Trustee considered the following four primary transaction types:[37]

- **Direct Payment:** Transactions in which a Participant makes a payment (via cash, wire transfer, check, payment processor, etc.) to TelexFree in satisfaction of an invoice.
- **Direct Receipt:** Transactions in which a Participant receives a payment from TelexFree or payment processor.
- **Triangular Payment:** Transactions in which a recruiting Participant satisfies the membership fee for a newly recruited Participant using the recruiting Participant's accumulated Credits.
- **Triangular Receipt:** Transactions in which a recruiting Participant is assumed to have receives cash from a newly recruited Participant in exchange for using his/her accumulated Credits to satisfy an invoice for the newly recruited Participant.

40. In addition to these four primary transaction types noted in the Martin Report, I understand the Net Equity Calculation also includes two other types of transactions: (i) chargebacks; and (ii) manual/purchased Credits. Chargebacks relate to bank or credit card refunds ("Chargebacks") while manual Credits, or purchased Credits as they are referred to in the Huron Data Tables, relate to instances where a participant paid TelexFree directly or through a bank or merchant to purchase system Credits ("Purchased Credits").[38]

[36] I understand that a "Hybrid" entity reflects instances where a User Account transacted in both USD and BRL. (See e.g., Telexfree Analysis of Damages, Huron, March 9, 2017 ("Huron Analysis of Damages") at 26.)

[37] Martin Report at 11-12; Huron Analysis of Damages at 33-35.

[38] Telexfree Net Equity Calculation and Aggregation.pdf; TelexFree Analysis of Damages at 33-35.

41. More specifically, Purchased Credits were Credits issued to User Accounts unrelated to the purchase of a membership plan and not resulting from the placement of advertisements or other components of the compensation scheme. Although I understand that the Trustee alleges that significant amounts of Purchased Credits were issued without consideration, I understand there were other instances where Purchased Credits were issued to User Accounts in exchange for cash payments to TelexFree.[39] As such, Purchased Credits were excluded from the Net Equity Calculation by Mr. Martin unless SIG indicated that a payment was made to a banking institution or via cashier's check, or the value was below $20,000.[40]

42. There were also exchanges of Credits between User Accounts unassociated with the issuance and satisfaction of TelexFree invoices ("Credit Transfers").[41] These Credit Transfers have also been excluded from the Net Equity Calculation by Mr. Martin, and thus the Trustee.

43. The results by Net Winner and Net Loser based upon the Huron Data Tables provided by Mr. Martin are summarized below:[42]

Participant Type	User Accounts	Participants	Total Net Equity
Net Loser	9,196,908	2,042,022	($1,809,367,883)
Net Winner	1,790,699	95,052	$1,583,177,635
TOTAL	10,987,607	2,137,074	($226,190,248)

VI. SUMMARY OF ANALYSIS AND FINDINGS

44. I have assessed the methodologies set forth in the Martin Report relative to: (i) the Aggregation Algorithm that is used as part of the Aggregation Process; and (ii) the Net Equity Calculation for Participants, and specifically Net Winners. While I address each of the methodologies separately below, both the Aggregation Process and the Net Equity Calculation contained flawed assumptions that render Mr. Martin's, and ultimately the Trustee's, conclusion

[39] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 10-11; Huron Analysis of Damages at 34; and Telexfree Net Equity Calculation and Aggregation.pdf at 2.

[40] Huron Analysis of Damages at 34; and Telexfree Net Equity Calculation and Aggregation.pdf at 2. (I am not aware of any analysis performed by Mr. Martin relative why a $20,000 threshold was selected.)

[41] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 10-11.

[42] See **Exhibit 3**.

regarding the Net Equity amounts supposedly won or lost by a given TelexFree Participant to be unreliable.

45. Relative to the Aggregation Process, key considerations include:

- Mr. Martin provides no analysis supporting why deterministic record linkage, a supposedly "generally accepted" methodology for linking disparate patient/medical information as part of research studies, is appropriate for aggregating the Potential Identifiers of User Accounts in an alleged Ponzi/pyramid scheme, especially given the possible differences in both the data quality and relative impact of missing or false links.

- The source literature cited in the Martin Report suggests probabilistic record linkage may be a more appropriate methodology when direct identifiers do not exist and/or the data is of poor quality – both of which are conditions Mr. Martin admit are true.

- I have not seen any meaningful evidence that TelexFree had internal controls in place during the registration of a User Account that would have prevented a Participant from entering different variations of their name or a name other than their own, such as spouse's name, family member name, business name, colleague/employer name, or some fictitious, erroneous, or arbitrary combination of numbers, letters, or punctuation. As a result, utilizing an Aggregation Process that as a "starting point" requires Participants to have accurately entered the same or similar names in every instance across all their User Accounts (which could number in the thousands) is unreliable and inconsistent with the facts of the case.

- Mr. Martin's conclusion that the ePOC data provides "the strongest evidence that the Trustee's Aggregation Process yielded reliable results" is unfounded. The Martin Report fails to account for relevant consideration such as instances where ePOC users were functionally aggregating supposedly separate Participants through the submission of multiple claims or account information sets, and also provides no analysis to suggest that a relatively small percentage of users (i.e., less than 5%) are sufficiently representative of the entirety of TelexFree Participants, including the Net Winner Class. This is especially true given that the majority of user claims in ePOC are for relatively small-dollar Net Losers, which can differ greatly in structure and complexity from that of larger Net Winners.

- Manual review of the top Net Loser and Net Winner Account Aggregations suggests numerous instances of potential under aggregation. This is highly problematic as even a single missed link has the ability to (and often did) dramatically impact the Net Equity associated Participants.

46. Relative to the calculation of Net Equity, key considerations include:

- There is no discussion in the Martin Report regarding why Triangular Transactions are included within the determination of Net Equity, but Credit Transfers are excluded

when both appear to be substantively similar from a real-world (i.e., cash based) perspective.

- To the extent any uncertainty exists regarding if, or how much, cash was exchanged as part of Credit Transfers, then such uncertainly necessarily extends to Triangular Transfers as well given that the SIG database does not provide evidence of cash payments for either transaction type. This internal inconsistency within the Net Equity Calculation between the treatment of Credit Transfers and Triangular Transactions results in a concluded Net Equity value that is not reliable

- A foundational assumption of the Net Equity Calculation is that the Credit values recorded within SIG for Triangular Transactions (and/or Credit Transfers) are "typically" reflective of the amount of cash that was exchanged between Participants. However, the facts in this case suggest that in practice Participants transacted at amounts that could vary greatly, and sometimes at significant discounts, from the amount of Credits shown in SIG due to factors such as specific business arrangements, competition between Participants or collection issues from new recruits. I understand that certain Participants also worked as part of larger teams with profit-sharing arrangement. None of this real-world activity is reflected in the SIG data on which Mr. Martin and the Trustee relies, and as a result the Net Equity Calculation can differ greatly from the net amount actually invested and received by Participants.

A. The Aggregation Process is Unreliable

i. Mr. Martin's Application of Deterministic Record Linkage to the Aggregation Algorithm Does to Provide Reliable Results

47. For the Trustee to determine whether a given Participant was a Net Winner or Net Loser, it was first necessary for the Net Equity associated with the one or more underlying User Accounts created and controlled by a Participant to be combined into a single "Account Aggregation" in a complete and accurate manner.

48. To accomplish this task, Mr. Martin relied upon a Deterministic Record Linkage, which he describes as "generat[ing] links based on the number of individual identifiers that match among the available data sets…[t]wo records are considered a match under a deterministic record linkage procedure if all, or a predetermined number of identifiers are identical."[43]

49. To support this methodology as "generally accepted practice," Mr. Martin cites to certain academic literature, which largely relates to data linkage in the context of healthcare information

[43] Martin Report at 14-15.

and related research.[44] Absent from Mr. Martin's report is any discussion regarding if, or to what extent, a methodology that may be generally accepted for linking disparate patient files or medical information for research studies is appropriate for the purposes of aggregating the Potential Identifiers of User Accounts in a an alleged Ponzi/pyramid scheme.

50. The importance of considering the linkage methodology relative to the required application is even noted in the Martin Report, stating:[45]

> Because different industries require different degrees of accuracy and because the complexity and diversity of data sources vary by company, businesses need to decide which data matching strategy best suits their needs prior to implementation. Each method has its strengths and weaknesses, so businesses should not be concerned about which is better in an absolute sense, but rather what the practical application will be and their tolerance for errors.

51. Further, relative to the complexity and diversity of the data in this case, Mr. Martin states that deterministic linking methods are recommended "[i]n information-rich scenarios where direct identifiers are available and of good quality" and that "[i]n these scenarios, deterministic methods are easy to implement, easy to interpret, and effective.[46] However, such a scenario is clearly not present in the SIG data the forms the basis of the Aggregation Process. Mr. Martin admits this fact, stating:[47]

> Participants often provided inaccurate, inconsistent, or incomplete information when opening accounts, which complicated the task of aggregating multiple accounts held by a single Participant.
>
> …
>
> [T]he information provided by Participants was often incomplete, inaccurate, or clearly incorrect. For example, we noted phone numbers comprised entirely of letters, email addresses missing the "@" symbol and instances where only a period or an ellipsis was entered instead of other requested information.

52. In contrast to scenarios where deterministic linking methods are recommended, the Statistical Data Warehouse Design Manual cited by Mr. Martin states:[48]

[44] Martin Report at 14 and Exhibit B.

[45] Martin Report at 14.

[46] Martin Report at 14.

[47] Martin Report at 10 and 16.

[48] Kroeze, Harold. "Methodology: Data Linkage." European Commission, February 2017.

> In scenarios that are information poor (where direct identifiers are unavailable) and/or the data are of poor quality, probabilistic methods consistently outperform deterministic methods and thus merit the extra time and resources required to implement them.

53. However, the Martin Report neither provides a substantive discussion regarding why probabilistic record linkage did not merit the extra time and resources, nor why he believed deterministic record linkage to be the more appropriate approach given the poor quality of Potential Identifiers in the data in this instance.

54. One of the areas where the importance of data quality is most prominent is the necessity for the name utilized by Participants to be both consistent and accurate across their User Accounts. Specifically, the Aggregation Algorithm required that two User Accounts, or groups of User Accounts, share the same or similar name in order to be combined. Mr. Martin states "[b]ecause Participants could receive Direct Receipts from TelexFree, it is reasonable to assume that a Participant would include an accurate name when registering a User Account."[49] However, Mr. Martin does not expand on this point to explain why a name must be accurate or specific to the individual in order for the Participant to receive a Direct Receipt, nor does he appear to consider situations where a single Participant may have created multiple User Accounts using a range of different names still directly associated with the Participant, such his/her own name, spouse's name, family member's name, or business name.[50] This could still potentially allow for the same person to receive a Direct Receipt from two different Account Aggregations under Mr. Martin's assumption.

55. Further, Direct Receipts only accounted for about $153 million, or about 5%, of the total combined Direct Receipts, Triangular Receipts and Chargebacks in TelexFree.[51] In terms of number of Participants, only around 46,400 of the approximately 2,100,000 TelexFree Participants had Direct Receipts, or roughly 2%.[52] As such, suggesting that Participants necessarily entered

[49] Martin Report at 16.

[50] See e.g. Affidavit of Frantz Balan, July 29, 2020 ("F. Balan Affidavit"); Affidavit of Lusette Balan ("L. Balan Affidavit").

[51] Calculated as $153 million in Direct Receipts divided by $2.9 billion (i.e., total of Direct Receipts, Triangular Receipts and Chargebacks). See **Exhibit 4**.

[52] Based on queries of Huron Data Tables.

consistent and accurate names across all their User Accounts just so they receive Direct Receipts (to the extent that was even a requirement) is unfounded.

56. I am also not aware of any internal controls that took place during the registration of a User Account that would have validated or confirmed a name, or prevented a Participant from entering a fictious, erroneous or arbitrary combination of numbers, letters, or punctuation in place of a name.[53] In fact, there are approximately 2,400 unique names (i.e., full_name or rep_nome) associated with more than 200,000 separate User Accounts where the name entered by the user is less than 4 characters.[54] Of these, the vast majority (i.e., over 2,300 names) are aggregate Net Losers, the losses of which total more than $11.4 million.[55] Examples include the following:

full_name	# of User Accounts	Aggregate Net Loss
1 1	40,687	($2,063,193)
. .	36,498	($1,986,986)
a a	30,363	($1,551,269)
.	8,329	($535,314)
q q	5,339	($278,741)
w w	3,191	($161,427)
c c	3,009	($151,197)
m m	2,077	($108,882)
x x	2,166	($108,732)
d d	2,021	($102,095)
l l	1,966	($98,403)
Jo	1,849	($92,354)
a b	1,770	($89,421)
0 0	1,667	($84,082)
k k	1,512	($76,497)
s s	1,455	($72,954)
j	1,328	($68,650)
j j	1,298	($65,619)
p p	1,242	($62,076)
f f	1,171	($59,481)

57. It is also important to note that while I have combined User Accounts with the same name for illustrative purposes above, each of these User Accounts is treated as a separate Participant

[53] See e.g., F. Balan Affidavit at 4-5.

[54] Based on queries of Huron Data Tables.

[55] Based on queries of Huron Data Tables.

under the Aggregation Algorithm since each is less than four characters in length and thus excluded by Mr. Martin from the Aggregation Process.[56] I also found numerous user names that appear to include Chinese characters. It is unclear how these User Accounts are treated within the Aggregation Algorithm, but given that they were generally two or three characters in length they likely remained unaggregated even though they may contain complete and valid names.

58. Additionally, based on my review of the SIG data, I have also seen countless usernames with more than four more characters that are as equally inaccurate, such as the name ".....", which was used for seven different User Accounts. The examples described above are demonstrative of the poor quality of the name entered by the Participant on which Mr. Martin relies heavily in his Aggregation Process.

59. Assuming that a Participant entered the same or similar name across every one of his/her User Accounts when (i) TelexFree did not appear to have any substantive verification or internal controls process as part registering a new User Account and (ii) Participants had little, if any, incentive to enter names (or any Potential Identifiers) accurately, is not a reliable basis or "starting point" for the Aggregation Process. This is especially true when the names provided in many instances were clearly fictitious.

ii. The ePOC Data Does Not Validate the Aggregation Process

60. I understand that the Trustee and his advisors developed an electronic Proof of Claim process (i.e., ePOC) portal to enable Participants to electronically file a claim. The ePOC portal required Participants to provide at least four Potential Identifiers associated with any of their User Accounts ("Account Information Set"). The Participant was then shown the User Account(s) that the Aggregation Process associated with that Participant based on the Account Information Set(s) provided. I understand that the Participant could then (i) dispute the User Account(s), (ii) provide information regarding additional accounts, or (iii) dispute any transactions occurring in any account associated with that Participant identified by the Aggregation Algorithm.[57]

[56] Martin Report, Addendum A at 1.

[57] Martin Report at 3-4.

61. The Martin Report appears to rely solely on the results of ePOC to validate the results of the Aggregation Process, but such reliance is unfounded. Specifically, Mr. Martin states:[58]

> The Aggregation Process was validated by the results of the ePOC process. There were approximately 130,700 timely filed claims, of which approximately 118,000 provided at least four Common Identifiers. Of the Participants who provided the minimum required Common Identifiers, more than 95% accepted the User Accounts identified by the Aggregation Algorithm. The remaining Participants either manually added additional User Accounts, disputed User Accounts, manually modified their claim without referring to a User Account or some combination of these.

62. In order to test the reliability of Mr. Martin's conclusion that the ePOC data validates the Trustee's Aggregation Process, I first attempted to confirm that the ePOC database I received was complete, and that I could reconcile to the overall claim figures indicated in the Martin Report. In total, I identified approximately 132,100 unique Claim IDs (as compared to 130,700 in the Martin Report), of which approximately 119,400 (as compared to 118,000 in the Martin Report) provided at least four common identifiers as part of Account Information Sets that were "matched"[59] to a TelexFree User Account and ultimately a Participant. While my results are similar to Mr. Martin's with respect to total claims, it is unclear whether the variance is a result of the criteria used to count the claims or the underlying data itself.

63. It is also unclear what precisely Mr. Martin means by "[o]f the Participants who provided the minimum required Common Identifiers, more than 95% accepted the User Accounts identified by the Aggregation Algorithm."[60] This statement lacks transparency based upon the structure of the ePOC database itself, as well as the nature of the information provided by users. Specifically, a given user had the ability, and frequently did, submit more than one claim. Furthermore, each claim could consist of one or more Account Information Sets that matched (or did not match) to a given User Account or Account Aggregation.

64. As a result, it is not clear precisely how Mr. Martin arrived at his 95% figure or what it is intended to represent. Despite this fact, I have attempted to emulate the results by: (i) counting the number of accepted claims, which I have assumed to be instances where the field HasAccepted=1

[58] Martin Report at 4.

[59] Absent additional information, "matched" in this case is assumed to be instances where the field MatchedAccount=1 in the AccountInfo table within ePOC.

[60] Martin Report at 4.

and HasAgreed=1 in the ePOC Claims table;[61] and (ii) more broadly where for a given User the AmountOfClaim is equal to the NetEqiutyPerTelex in the ePOC ClaimSummary table, presumably indicating that the user did not make any changes to the Net Equity amounts, and thus the underlying aggregation(s) or associated transactions. To the extent Mr. Martin's calculation was based on one of these approaches, and attempting to replicate other relevant criteria, I was able to recalculate potentially comparable figures in the 91%-93% range.[62] However, regardless of the approach, the issue with such an analysis is that it fails to appropriately consider how Participants practically claimed Account Aggregations through the ePOC portal, as well as to what extent those claims are representative of the 2.1 million TelexFree Participants, and specifically the Net Winner Class

65. Based on the relationship between users, claims, and Account Information Sets within ePOC there are a number of issues that arise, including: (i) a user can have multiple agreed and accepted claims, each linked to a different Participant; and/or (ii) a user can have a single agreed and accepted claim with multiple Account Information Sets, each linked to a different Participant. In both instances, the real-world outcome is that a user is actually suggesting the Trustee has *incorrectly* aggregated his/her accounts through missed links, despite the individual claim(s) being agreed/accepted and the Account Information Set(s) matched.

66. In fact, of the approximately 85,600 users (i.e., unique UserIDs) that submitted at least one accepted and agreed claim with at least one matched Account Information Set, approximately 12,800, or 15%, were functionally aggregating multiple supposedly separate Participants together by entering multiple claims or Account Information Sets.[63] Further, this number grow to 22,000 users, or 25% of the 85,600 users, when you consider instances where the user in at least one of

[61] I understand that the "HasAccepted" field in the ePOC Claims table relates to whether a user accepted all the transactions associated with his/her account(s) after they were reviewed and potentially edited. I further understand that the "HasAgreed" field in the ePOC Claims table relates to whether a user agreed that all his/her accounts had been inputted by him/her.

[62] If Mr. Martin's definition of Participant acceptance was based on claims where the field HasAccepted=1 and HasAgreed=1 in the ePOC Claims table, then this occurred for approximately 111,500 of the 119,400 claims that had at least one Account Information Set that "matched" (i.e., MatchedAccount=1 in the ePOC AccountInfo table) to a TelexFree Participant, or 93% on the time. However, if Mr. Martin's definition of Participant acceptance is any user for which the total AmountOfClaim is equal to the total NetEqiutyPerTelex,in the ePOC ClaimSummary table, then this occurred for approximately 78,200 of the 85,600 users that submitted an accepted and agreed claim with at least one matched Account Information Set, or 91% of the time. (Based on queries of ePOC Data Tables.)

[63] Based on queries of ePOC Data Tables.

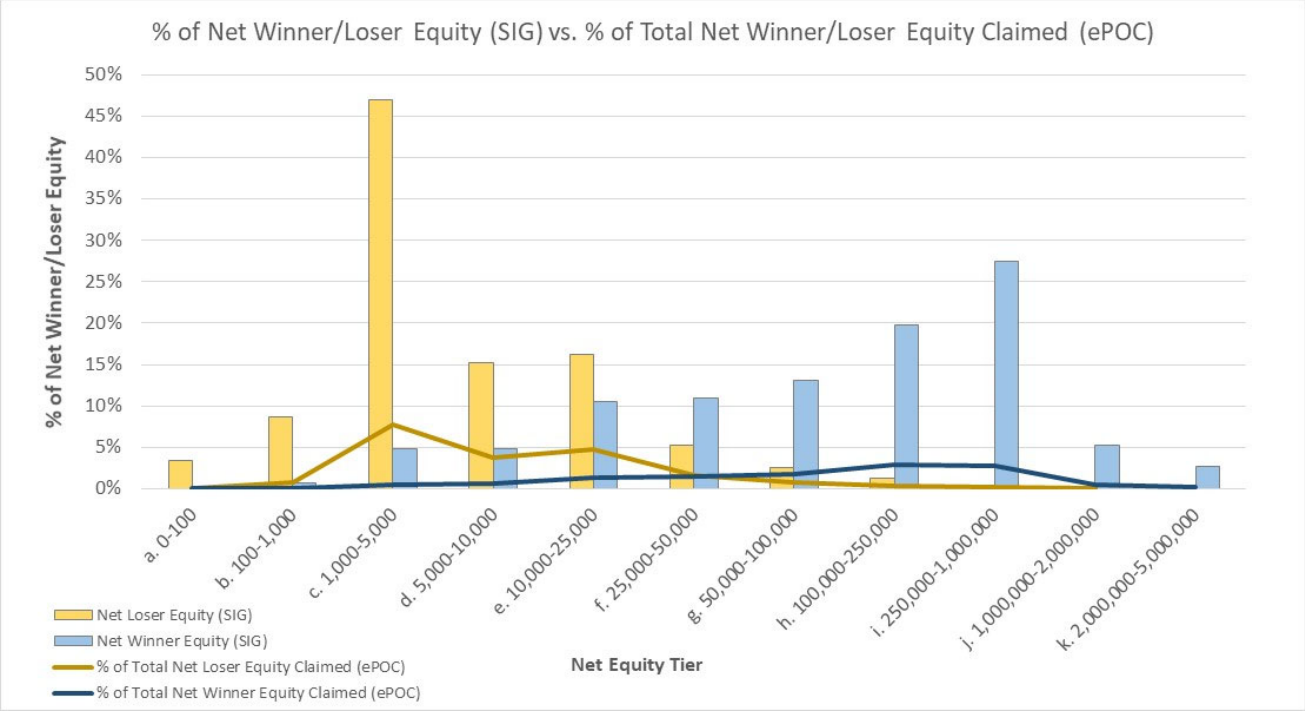
their claims either disputed transaction(s), entered an Account Information Set that didn't match to Potential Identifiers (and for which that was a claim amount), didn't accept or agree to a claim, manually added User Accounts, or made some other adjustment to Net Equity. Such results do not validate the Aggregation Methodology as Mr. Martin suggests.

67. Further, the Martin Report does not appear to consider if, or to what extent, a relatively small percentage of users are sufficiently representative of the entirety of TelexFree Participants, including the Net Winner Class. This is especially important given that the majority of claims made by users through the ePOC portal were for smaller-dollar Net Losers, which can differ greatly in structure and complexity from that of larger Net Winners.

68. Specifically, the 123,300 unique Account Aggregations claimed within ePOC[64] only represent about 6% of the total number of TelexFree Account Aggregations within SIG.[65] Furthermore, the aggregate value of claims within ePOC is also relatively small in comparison to total TelexFree Net Equity, particularly with respect to the Net Winner Class, as shown in the chart below.

[64] Includes Participant Account Aggregations where at least one underlying User Account matched to an Account Information Set (i.e., Matched Account=1 in the AccountInfo table), and which was included as part of an accepted claim within ePOC (i.e., HasAccepted=1 in the ePOC Claims table).

[65] See **Exhibit 5**.



69. There are also significant differences between the limited Account Aggregations claimed in ePOC and the Net Winners Class. For example, across the approximately 123,300 unique Account Aggregations claimed within ePOC (of which less than 10% were Net Winners), there were on average 16 underlying User Accounts.[66] By Comparison, almost 80% of the total Net Equity for Net Winners is comprised of Account Aggregations that averaged 50 or more underlying User Accounts. The chance of the Aggregation Process generating either a missed link or false link would likely be greater for Participants with more underlying User Accounts, as each subsequent User Account created by the Participant would provide additional opportunity for the Participant to change the name or other Potential Identifiers entered during the registration process.

70. There is also the consideration that users may simply have accepted the Account Aggregation shown to them even if the aggregation was not complete or accurate given that the majority of claims were relatively small in value (i.e., less than $5,000).[67] To successfully dispute the Account Aggregation could require additional information that may not have been available to the Participant or otherwise add uncertainty to the claims process.

[66] See **Exhibit 5**; Based on queries of ePOC Data Tables.
[67] See **Exhibit 5**.

71. The Martin Report suggests that the percentage of people who supposedly submitted and accepted claims through the ePOC portal is the "strongest evidence that the Trustee's Aggregation Process yielded reliable results."[68] However, as discussed above, the ePOC data fails to support such a conclusion, and I am not aware of any other independent analysis performed by Mr. Martin that would confirm or even provide a potential error rate associated with the Aggregation Process.

iii. Aggregation Errors Can Result in a Significant Overstatement of Participant Net Equity

72. With respect to the impact of potential linkage errors, the Statistical Data Warehouse Design Manual relied upon by Mr. Martin states:[69]

> Poor quality (eg if variables are missing, indecipherable, inaccurate, incomplete, inconstant, inconsistent) could lead to records not being linked – missed links – or being linked to wrong records – false links. The impact of these two types of errors may not be equal (eg a missed link may be more harmful than a false link), so this needs to be taken into account when designing a data linkage strategy, especially if the linking has legal or healthcare implications.

73. The significant impact of a single missed link on a Participant's concluded Net Equity is particularly apparent in this case when comparing the Trustee's largest Net Loser and Net Winner Account Aggregations. For example, the top 100 Net Losers in TelexFree comprise a total loss of approximately $22 million, and have individual losses ranging from approximately $126,200 on the low end to approximately $1,000,000 on the high end.[70] Comparably, the top 100 Net Winners in TelexFree comprise a total gain of approximately $148 million, and have individual winners ranging from $933,000 on the low end to $4,097,000 on the high end.[71]

74. Given the structure of TelexFree, there is a lower likelihood that a Participant would incur significant net losses on the scale reflected by the top Net Losers (i.e., repeatedly buy into TelexFree either directly or through Triangular Transactions yet receive minimal or no gains).[72] One alternative rationale for the existence of some of these large Net Losers is the Trustee's failure to appropriately consolidate certain Net Loser Account Aggregations with corresponding Net

[68] Martin Report at 20.

[69] Kroeze, Harold. "Methodology: Data Linkage." European Commission, February 2017.

[70] See **Exhibit 7**.

[71] See **Exhibit 8**.

[72] See e.g., **Exhibit 5**.

Winners. To test this hypothesis, I manually reviewed the Potential Identifiers for each of the top Net Losers with losses over $200,000, which totaled 31 Account Aggregations, and subsequently attempted to find matching Potential Identifiers among Net Winners. Of the 31 aggregations I reviewed, I found one or more Potential Identifiers that matched a Net Winner aggregation for 30 of them. This evidence, which is summarized in **Exhibit 6** and detailed **Exhibits 9.1** to **9.31**, includes instances where the same or similar values for Potential Identifiers, such as address, phone number, login, or email, appeared across both a Net Loser and Net Winner Account Aggregations.

75. For each of the potential matches I found, I also queried the Huron Data Tables to determine whether certain transactions, such as Credit Transfers or Triangular Purchases/Receipts, occurred between the matched Account Aggregations. Based on my review of the SIG data in this matter, I found that there was often a high degree of activity between User Accounts held by the same Participant, a factor not considered by Mr. Martin as part of the Aggregation Process. As such, confirming that transactions occurred between the matched Account Aggregations, which it did in most instances, further strengthens the potential connections.

76. By way of example, and as reflected in **Exhibit 9.1**, the single largest Net Loser in TelexFree based on the Trustee's calculation of Net Equity is Parent ID 13043527, Du Painting DBA ("Du Painting"), which lost $1,010,727 and comprised over 1,200 User Accounts. Across these User Accounts, Potential Identifiers for Du Painting included logins structured as "visionusa[##]", email addresses such as davidsonteixeira@hotmail.com, physical addresses such as 33 Vernal St., Everett MA, and phone numbers such as 781-484-7474. A separate Participant under the Trustee's Aggregation Process, Parent ID 4032900, Davidson R. Teixeira, is shown to be a Net Winner of $1,544,904 comprised of 80 different User Accounts. Across Mr. Teixeira's User Accounts, every one of the noted Potential Identifiers for Du Painting can be found, which would provide strong indicia of both Account Aggregations being the same Participant. As a result of what appears to be a missed link by the Aggregation Algorithm, Mr. Teixeira's Net Equity is overstated by more than $1 million, or approximately 189%, a very significant difference.

77. In total, and simply based off the manual review of this sample population, I identified approximately $11 million in Net Loser aggregations that could potentially reduce the $25 million

allegedly owed by the matched Net Winners, resulting in a reduction of approximately 45%.[73] Even considering just those instances where there were also transactions between the matched Account Aggregations, this would still reflect approximately $8.4 million in Net Loser aggregations that could potentially reduce $19 million alleged owed by the matched Net Winner aggregations (i.e., also a reduction of approximately 45%).

78. As discussed above, one of the factors likely driving what appears to be under-aggregation in certain instances, is the disconnect between the Aggregation Algorithm's requirement that the names be the same or similar across every Participant User Account and the reality that Participants could create User Accounts using a variety of names, such as business name, personal name, or even fictional names. This reliance on consistent naming fails to appropriately account for instances where other Potential Identifiers provide strong indicia that the User Accounts potentially belong to a single Participant.

79. For example, based upon my manual review of top Net Losers, I found that Shall Group (SIG parent ID 12650741 and Net Loser of $708,133) shared many of the same or similar Potential Identifiers with Dolarex (SIG parent ID 10992563 and Net Winner of $190,381). In fact, this linkage was confirmed by the claims submitted through the ePOC portal, as I can see that a single user (i.e., Saryas Jaff, UserId 26627) submitted two claims, one for Shall Group (ClaimId 42877) and one for Dolarex (ClaimId 42906). Once this under-aggregation is appropriately considered and accounted for, instead of Dolarex being a $190,000 Net Winner, it is actually a $518,000 Net Loser under the Net Equity Calculation.[74] This further demonstrates the significant impact a single aggregation error can have on a Participant's Net Equity.

80. It also important to consider the impact of linkage errors in the appropriate context. While a given missed or false may not represent a significant portion of TelexFree Net Equity in total, it can dramatically change the results for a Participant.

[73] See **Exhibit 6**.

[74] See **Exhibit 9.4**.

B. The Calculation of Net Equity is Unreliable

i. The Net Equity Calculation Fails to Account for Participant Investments and Receipts Related to Transfers of Credit

81. Relative to what constitutes Net Equity, I understand the Court in this matter stated the following:[75]

> (i) In determining the amount of claims of Participants, any claim or portion of claim based upon accumulated credits in a Participants' User Accounts as of the Petition Date shall be disallowed;
>
> (ii) The claims amounts of Participants shall be determined on a Net Equity basis, which shall be defined as follows: the amount invested by a Participant into the Debtors, including amounts paid by the Participant pursuant to the Triangular Transactions, less amounts received by a Participant from the Debtors' scheme, including amounts received by the Participant pursuant to the Triangular Transactions;
>
> (iii) In determining the amount of a claim of a Participant who has more than one User Account, the activity in all of the Participants' User Accounts shall be aggregated and netted against one another[.]

82. Importantly, while this order notes that claims based on accumulated Credits are disallowed, Net Equity should consider the amount *invested*[76] by a Participant in TelexFree less amounts *received* by a Participant from the TelexFree scheme. Based on how the Trustee and Mr. Martin described their determination of Net Equity, the distinction between accumulated Credits and amounts invested/received appears to be the exchange of actual monetary consideration, such as cash. For example, simply earning Credits through the placement of advertisements on classified websites would not entitle the Participant to claim the value of those Credits, but if that same Participant gave cash to another Participant as part of a Triangular Transfer, then the cash paid would indeed be considered as part of the Net Equity Calculation.

83. However, the determination of Net Equity as executed by the Trustee and Mr. Martin does not account for instances where a Participant uses cash to purchase Credits from another Participant for purposes such as setting up additional User Accounts or performing subsequent Triangular Transactions. In both instances, a Participant is investing actual cash into the TelexFree

[75] Supplemental Order at 2-3.

[76] For purposes of this report, I am using the term "invest" or "invested" to mean the payment of cash into the TelexFree scheme, such as through Direct Payments, Triangular Payments, Credit Transfers In, or Purchased Credits. No other definition or legal meaning is intended or implied.

scheme, yet the Net Equity Calculation treats these situations inconsistently by including Triangular Transactions and excluding Credit Transfers.

84. This dynamic between Triangular Transactions and Credit Transfers is also evident within the SIG data. For example, Mr. Balan, together with the accounts that I understand were incorrectly aggregated to his wife,[77] is alleged to be a $1.3 million Net Winner based upon the Net Equity Calculation performed by Mr. Martin, as shown in **Exhibit 10**. Most of those supposed winnings came in the form of Triangular Receipts, which totaled almost $1.5 million. However, in order to help generate the Credits necessary for this volume of Triangular Receipts, Mr. Balan also was a substantial net importer of approximately $848,000 in Credits, which I understand were not provided to Mr. Balan free of cost as Mr. Martin and the Trustee assumes.[78] By including Triangular Transactions and excluding Credit Transfers, Mr. Martin is artificially increasing the supposed net amounts received by Participants and not properly considering the investments that were necessary in order to generate those supposed receipts.

85. There is no discussion in the Martin Report regarding why Triangular Transactions are considered within the determination of Net Equity but Credit Transfers are excluded when both appear to be substantively similar from a real-world (i.e., cash based) perspective. Additionally, to the extent any uncertainty exists regarding if, or how much, cash was exchanged as part of Credit Transfers, then such uncertainly necessarily extends to Triangular Transfers as well given that the SIG database does not provide evidence of cash payments for either transaction type. This internal inconsistency within the Net Equity Calculation between the treatment of Credit Transfers and Triangular Transactions results in a concluded Net Equity value that can differ greatly from the net amount actually invested and received by Participants.

86. Credit Transfers between accounts and Credit Transfers as part of Triangular Transactions are also stored together within SIG from a data perspective. I understand that the "Transfer Table" contains information about transfers of Credits between User Accounts and includes a "tipo" that delineates "ten different transfer types, including requests to withdraw funds, payments of invoices

[77] L. Balan Affidavit; F. Balan Affidavit at 3.

[78] F. Balan Affidavit at 6-7.

with credits, and transfers of credits between accounts. The remaining types relate to chargebacks and other adjustments."[79]

87. Further, the transfer of Credits in exchange for cash is not just a hypothetical scenario that occurs infrequently; rather, the SIG data reflects a vast transfer of Credits both between distinct Participants and internally within a Participant's User Accounts. This is especially true for many of the Participants that Mr. Martin has determined to be large Net Winners and Net Losers, as shown in **Exhibit 7** and **Exhibit 8**. For example, the top 100 Net Winners made over $210 million in Credit Transfers Out and over $250 million in Transfers In (i.e., significant net importers of Credits). By comparison, the top Net Losers made $31 million in Credit Transfers Out and $18 million in Transfers In (i.e., significant net exporters of Credits).[80]

88. The inclusion of Credit Transfers is highly impactful to the determination of a Participant's Net Equity as in many cases it can change Net Winners to Net Losers or Net Losers to Net Winners, a sample of which is shown below:

Participant Type	**Parent ID**	**Net Equity Excluding Transfers In/Out**	***6 - Credit Transfers In***	***7 - Credit Transfers Out***	**Net Equity Including Transfers In/Out**
Net Loser	13043527	**$(1,010,727)**	*$(576,124)*	*$831,217*	**$(755,634)**
Net Loser	11849449	**$(622,342)**	*$(1,904,355)*	*$2,254,352*	**$(272,344)**
Net Loser	12242105	**$(481,228)**	*$(621,865)*	*$1,008,813*	**$(94,279)**
Net Loser	2876246	**$(406,591)**	*$(54,842)*	*$1,243,613*	**$782,181**
Net Loser	1020407	**$(198,009)**	*$(5,143,557)*	*$6,146,96*	**$805,396**
Net Winner	132785	**$4,096,689**	*$(7,287,436)*	*$6,617,041*	**$3,426,294**
Net Winner	2522900	**$4,016,153**	*$(7,128,607)*	*$3,921,949*	**$809,495**
Net Winner	323967	**$3,159,865**	$(5,155,263)	$3,260,978	**$1,265,580**
Net Winner	10853999	**$1,308,377**	($1,581,410)	$176,389	**($96,644)**
Net Winner	1826886	**$1,163,543**	($7,882,760)	$6,592,369	**($126,847)**

89. It defies economic logic that one Participant would simply give Credits to another Participant without financial consideration in return (particularly in these large sums) given that Credits could be converted into cash through Direct Receipts or, much more frequently, Triangular Transactions. It also worth noting that TelexFree appears to have charged the recipient of Credit Transfers a $3 fee per transaction, demonstrating that such activities were not without a cost.

[79] Huron Analysis of Damages at 29.

[80] See **Exhibit 7** and **Exhibit 8**.

90. There is also direct evidence of Credits being transferred in exchange for cash based on descriptions of events noted by users through the ePOC portal, such as:

- Ana Patricia Salmeron de Pereira (ePOC ClaimId 68594)
    - "Through the account of patriciaycarlos **I was buying the credit for other participants with real cash, so i could accumulate a large amount of credit in this account, to allow me to easily create new accounts for new members**. As this was the fastest method. Therefore when Telexfree got shutdown, all of this accumulated money ($90,038) had been distributed and ultimately given to Telexfree through accounts and Voips. So I never actually had this amount of money, therefore all I am owed for this account is the original $1425 that I paid."[81] (emphasis added)
    - The transactions reflected in SIG confirm that Ms. Salmeron de Pereira had $100,237 in total Credit Transfers In, of which $83,507 came from his login "patriciayalex" or "patriciaycarlos."[82]
    - If Credit Transfers In and Out were considered as part of Net Equity, Ms. Salmeron de Pereira's Net Winnings of $115,981 would be reduced to $42,839, a decrease of ($73,142), or 63%.
- Marcel McFarren (ePOC ClaimId4282)
    - "**I gave money to a friend of mine Diego Alonso Lizama Zegarra on several opportunities who transfered telexfree credits to my accounts in order to purchase new accounts and VOIP plans**. I currently don't remember or have records of all the transfers I received or the dates nor individual amounts. But **in the time I participated in telexfree I paid in cash [$30,383.40] US dollars that where transferred to my account from other user accounts**. Wich [$26,193.60] were used in opening user accounts and voip accounts, [$597] US dolars in user account maintenance and [$3,592.80] US dolars in voip renewals."[83] (emphasis added)
    - The transactions reflected in SIG confirm that Mr. McFarren had $30,889 of Credit Transfers In from User accounts not aggregated to Mr. McFarren.[84]

[81] ePOC database: UnmatchedAccount table, UnmatchedAccountId=6961, Description field.

[82] Huron Data Tables: Rep_ID_Summary and Claim_Statement tables, C13=12902527.

[83] ePOC database: UnmatchedAccount table, UnmatchedAccountId=23413, Description field.

[84] Huron Data Tables: Rep_ID_Summary and Claim_Statement tables, C13=13094257.

- If Credit Transfers In and Out were considered as part of Net Equity, Mr. McFarren's Net Winnings of $2,736 would be reduced to a Net Loss of ($21,924), a decrease of ($24,660) or 900%.

- Jose Norford (ePOC ClaimId 150603)
  - "This claim indicates that I owe approximately [$202,698.80] however I should be receiving a reimbursement of approximately [$32,775.00]."[85]
  - The transactions reflected in SIG confirm the determined Net Equity to be $202,698.80. However, when Credit Transfers In and Out are considered, Mr. Norford's Net Equity is reduced to $39,118.12.[86]
  - If Credit Transfers In and Out were considered as part of Net Equity, Mr. Norford's Net Winnings of $202,699 would be reduced to Net Winnings of $39,118, a decrease of ($163,581) or 81%.

91. This exclusion of Credit Transfers fails to appropriately consider the "amount invested" or "amounts received" by Participants within the determination of Net Equity, and, as a result, the Net Equity determined by Mr. Martin is not reliable for the almost 900,000 Account Aggregations that took part in Credit Transfer transactions, of which approximately 196,000 are currently Net Winners.[87]

ii. The Amount of Cash Invested/Received by Participants as Part of Triangular Transactions or Credits Transfers Can Differ from Amount of Credits Recorded in SIG

92. With respect to Triangular Transactions, Mr. Martin describes the process as follows:[88]

> To redeem credits for a recruited Participant a new User Account was opened for the recruited Participant and the recruiting Participant satisfied the resulting membership fee with credits held by the recruiting Participant (a "Triangular Transaction").
>
> Based upon interviews with Participants and others, the recruited Participant **typically** paid the membership fee in cash to the recruiting Participant who, in turn, satisfied the recruited Participant's membership fee with their own credits. (emphasis added)

[85] ePOC database: UnmatchedAccount table, UnmatchedAccountId=13886, Description field.

[86] Huron Data Tables: Rep_ID_Summary and Claim_Statement tables, C13=10789133.

[87] Based on queries of Huron Data Tables.

[88] Martin Report at 7.

93. This description is exemplary of why the Net Equity Calculation is not reliable. Mr. Martin and the Trustee are assuming that the amount of Credits transferred as part of a Triangular Transaction is equivalent to the amount of cash invested or received by Participants. However, the assumption that cash was exchanged is based on interviews with what could only be a very small percentage of the total 2.1 million TelexFree Participants, many of whom are outside of the U.S., and even then, cash exchanges only "typically" occurred.[89]

94. Mr. Martin does not provide any specific guidance regarding what "typically" means in this context. For example, was cash exchanged between Participants as part of Triangular Transactions 90% of the time, or less than 75%? In those instances where cash was exchanged, was it always the exact amount of, and equivalent to, the amount of Credits recorded within SIG, the data source which forms the basis for the Trustee's determination of Net Equity? Given the massive size and volume of Triangular Transactions that occurred on a global scale (i.e., more than $2.7 billion dollars and 13.3 million transactions for Triangular Receipts alone),[90] much of which I understand was transacted in physical cash,[91] the assumption that Credits recorded in SIG are necessarily equivalent to investments or receipts by Participants is unsupported.[92]

95. In direct contradiction to this assumption, I understand that Frantz Balan, a supposed Net Winner according the Trustee,[93] affirmed the following:[94]

> It is a mistake to assume that 100% of the face amount of invoices were collected 100% of the time. In practice, promoters like me rarely received the full amount in cash. I would estimate for me and the promoters I worked with or spoke to that at best this happened only 10% of the time.
>
> Most often, promoters were only able to ask for steeply discounted payments, as low as $250 for an AdCentral Family membership. Most commonly, the amount requested ranged from 50% to 75% of the invoiced amount.

96. Additionally, I understand that Participants were often part of larger teams working alongside or employing other Participants. In these instances, I understand each team, and each

[89] I understand that Counsel requested the notes from these interviews, but was informed that notes were not compiled.

[90] See **Exhibit 4**

[91] See e.g., F. Balan Affidavit at 5.

[92] F. Balan Affidavit at 5.

[93] Based on queries of the Huron Data Tables.

[94] F. Balan Affidavit at 5.

pair of Participants within a team, had their own compensation arrangements and there was no universal payment practice.[95]

97. As one approach to teaming, I understand that it was the role of the lower-level recruiting Participants of the team to identify and signup new members; however, the responsibility of satisfying the resulting invoices with Credits generally fell to the team lead higher up in the TelexFree pyramid structure. The recruiting Participants would also collect the cash from the new recruits (frequently at a discount relative to the amount of Credits, as noted above), and then distribute a percentage of the cash collected to the lead Participant and keep the remainder as a fee.[96] In this way, Participants were actually sharing compensation in a manner not reflected in the SIG data.

98. The information within SIG would not reflect this real-world dynamic, making it a potentially unreliable source for purposes of determining Net Equity. As a result, the Trustee's concluded Net Equity for the lead Participant would be overstated relative to cash, and the Net Equity for the recruiting Participant would be understated.

99. Also, as it relates to teaming, I'm not aware of anything that would have prevented to two people from potentially sharing User Accounts and splitting any cash investments and receipts associated therewith. There is evidence that this may have occurred based upon logins such as in the "patriciayalex" or "patriciaycarlos."[97]

100. Further, as part of my manual review of top Net Winners and Net Losers (discussed above), I explored certain of the underlying transaction detail to help me better understand the potential connections between top Net Losers and Net Winners. For example, the second largest supposed Net Loser, "Ming Hua," is alleged to have lost over $840,000 (parent ID 21036267).[98] This account aggregation is comprised of 16,909 underlying User Accounts, every one of which is a VOIP Phone Plan with a value of $49.90 that was purchased through a Triangular Payment. The chance that a single person actually bought 16,900 VOIP Phone Plans for over $800,000 during

[95] F. Balan Affidavit at 4.

[96] F. Balan Affidavit at 6.

[97] Huron Data Tables: Rep_ID_Summary and Claim_Statement tables, C13=12902527.

[98] See **Exhibit 7** and **Exhibit 9.2**.

the approximately one-month period of March 20, 2014 to April 13, 2014[99] (a process that would likely have taken one person months of full-time work to accomplish since each VOIP Phone Plan required the creation of a new User Account) does not stand up to common sense logic.

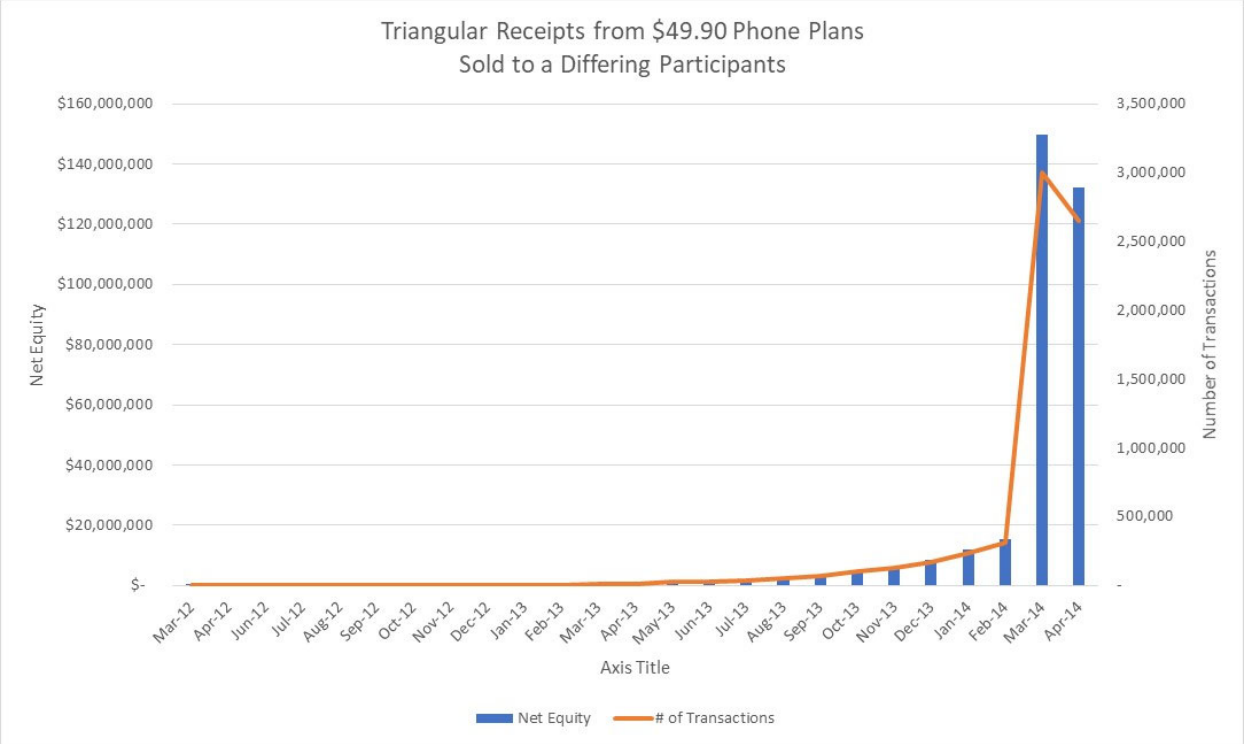
101. An alternative explanation lies in the timing of the first User Account in "Ming Hua's" aggregation (i.e., March 20, 2014). I understand that on or around March 9, 2014 TelexFree changed its compensation plan such that Participants were required to sell at least one VOIP Phone Plan, which many did through Triangular Transactions .[100] It is appears that "Ming Hua" is the result of this modification to TelexFree's policy. Whether the creation of these 16,909 User Accounts were done by a service company, an automated software product, or by some other means is not clear from the data, but it does introduce significant uncertainty regarding whether "Ming Hua" truly made over 16,000 cash payments of $49.90 to what the SIG data shows to be over 1,000 unique counterparty Participants in the span of a month.[101] To the extent that such amounts were not actually paid by "Ming Hua" to the counterparty Participants, then the Net Equity for those Participants is necessarily overstated. This is due to the fact that each supposed $49.90 Triangular Payment by "Ming Hua" would have a corresponding Triangular Receipt associated with the counterparty Participant in the SIG data, which is a component of the Net Equity Calculation.

102. In fact, when all TelexFree Triangular Receipts of $49.90 are charted over time a demonstrable spike beginning in March 2014 becomes clearly apparent, as shown in the table below:

[99] Based on queries of the Huron Data Tables.

[100] www.sec.gov/litigation/complaints/2014/comp-pr2014-79.pdf; www.justice.gov/usao-ma/pr/former-president-telexfree-sentenced-billion-dollar-pyramid-scheme; queries of the Huron Data Tables.

[101] Based on queries of the Huron Data Tables.



103. During just the post-March 9, 2014 period, there are over 5.6 million Triangular Receipts for $49.90 between User Accounts supposedly belonging to different Participants, resulting in supposed net winnings of more than $270 million.[102] As noted above relative to "Ming Hua," there is significant uncertainty (and no support in the SIG data) regarding whether Participants actually received the associated cash despite the increase to their Net Equity.

104. Additionally, it is not clear to what extent the Net Equity calculation considers the potential impact of foreign exchange rates. Mr. Martin identifies over 1.6 million Net Losers outside the United States across more than 200 different countries.[103] However, I understand that the Credit amounts within SIG are intended to reflect U.S. dollars (i.e., USD). As such, when foreign Participants exchange cash presumably in local currency, it is neither clear what exchange rate may have been used at the time, nor how changes in those exchange rates between the time of the transactions and today may disproportionately impact certain Participants.

[102] Based on queries of the Huron Data Tables.

[103] Huron Analysis of Damages at 18.

105. For example, Mr. Martin indicates that there were approximately 950,000 victims in Brazil.[104] The exchange rate today is approximately 5 Brazilian Reais (i.e., BRL) to 1 USD. By comparison, the exchange rate in 2014 was approximately 2.4 BRL to 1 USD.[105] Therefore, depending on whether you are a Net Winner or Net Loser, the amount of your claim or liability could differ from the original net transaction amount invested or received by as much as 100%.

106. The potential inconsistency between the actual value of the cash exchanged and the amount of Credits recorded for Triangular Transactions within SIG is equally true with respect to Credit Transfers. Given this uncertainty regarding a foundational assumption of the Net Equity calculation, which is that the Credit amounts reflected in SIG are a reliable and accurate proxy for actual cash investments and receipts, the resulting Net Equity amounts concluded by the Trustee are equally unreliable in this instance.

VII. CONCLUSION

I have assessed the methodologies set forth in the Martin Report relative to: (i) the Aggregation Algorithm that is used as part of the Aggregation Process; and (ii) the Net Equity Calculation for Participants, and specifically Net Winners. As addressed in detail above, both the Aggregation Process and the Net Equity Calculation contained flawed assumptions that render Mr. Martin's, and ultimately the Trustee's, conclusion regarding the Net Equity amounts supposedly won or lost by a given TelexFree Participant to be unreliable.

[104] Huron Analysis of Damages at 18.

[105] www.ofx.com/en-us/forex-news/historical-exchange-rates/yearly-average-rates.

EXHIBITS



Joshua Dennis

CVA

Partner

T: +1 617 570 3789
E: jdennis@stoneturn.com

Boston
75 State Street
Suite 902
Boston, MA 02109

Joshua Dennis, a Partner with StoneTurn, has more than 15 years of experience working with clients and counsel on issues requiring complex data analysis, including investigations, economic damages, intellectual property, enforcement, forensic accounting, compliance monitoring, and valuation engagements.

Through his expertise, which includes financial modeling and data analysis, Josh has assisted companies and their counsel across a wide range of investigative and dispute-related matters. He not only has significant experience analyzing accounting, economic and business data, as well as issues related thereto, but also applying his expertise to the creation of dynamic visual models that provide clarity and key insights.

Josh has also worked extensively in matters involving data analysis, mining, anomaly detection, and reporting of transactional and trading data utilized across a range of industries, including banking, financial services, manufacturing, pharmaceuticals, e-commerce and retail, among others.

Josh also has experience in evaluating and quantifying economic damages related to a variety of issues, including contractual, purchase price and employment disputes, as well as class action matters. He has also quantified and assessed reasonable royalty and lost profits damages related to patent, trademark, trade secret and copyright infringement claims, as well as the valuation of intellectual property outside of litigation.

Josh is a Certified Valuation Analyst, and earned a Certificate in Intellectual Property Law from Suffolk University.

Education

B.S., Management and Business, Skidmore College

Minor, Computer Science, Skidmore College

Practice Areas

Data Analytics

Intellectual Property

Litigation

SELECT PROFESSIONAL EXPERIENCE

Financial Modeling and Complex Data Analysis

Josh has designed, created and analyzed complex databases to assist clients in matters involving asset misappropriation, breach of contract, trading irregularities, regulatory inquiries and accounting issues. These data models, including graphic visualizations and dashboards, allowed users to dynamically explore large and disparate data sets to help confirm fact patterns, identify trends and anomalies, and inform decision making. Representative case experience includes:

- Retained by counsel for Plaintiff, an individual allegedly defrauded out of millions in capital contributed to a foreign investment project, as part of an effort to help identify the use and current location of such funds for purposes of recovering an adjudicated award. The analysis first required the creation of a database from over 500 hardcopy bank statements across a network of related entities around the world that was compiled using a combination of optical character recognition (OCR), programmatic data restructuring and manual input. The resulting database, which contained more than 35,000 transactions, was then analyzed to identify potential transactions of interest based on “fuzzy” keyword searches, as well as the creation of a dynamic dashboard that allowed users to quickly traverse transactions across entities and time periods to help identify the source and use of funds.
- Engaged by counsel for defendant, a global brokerage firm, in connection with allegations of a rogue trader manipulating gains and losses through the trading of metals futures. StoneTurn worked closely with the company and counsel to gather the information and trading data necessary to understand and quantify the magnitude of issue, as well as confirm the scope of which accounts were impacted. StoneTurn communicated the results of its findings, which included a review and assessment of internal controls, to the company’s independent auditor with respect its determination of any potential material weaknesses or significant deficiencies.
- Outside counsel retained StoneTurn to assist in the defense of a DOJ investigation into certain anti-trust and price fixing allegations in connection with agency bonds. Specifically, StoneTurn analyzed over 2 million lines of trading data provided by both the investment company and FINRA to help identify trends in the primary and secondary market sales. We also created dynamic visualizations in Tableau to identify any potential trading patterns and trends relative to certain prices contemplated by bond syndicate members.
- Outside counsel for a pharmacy company engaged StoneTurn to assist in a class action matter involving the quantification of potential wages due. Specifically, plaintiffs alleged certain supervisors were due wages and overtime (if applicable) for time spent on premises during unpaid meal breaks where they were the only managerial staff present. StoneTurn analyzed over 20 million lines of employee time punch and scheduling data across various systems to determine the duration of such meal breaks, as well as the additional wages allegedly due.

- Assisted counsel in a class-action matter to rebut the damages asserted by plaintiff related to allegations of market timing within the defendant's mutual funds over a period of more than a decade. Analyses included the creation of a dynamic model that could generate multiple scenarios based upon user-selected inputs, and which encompassed over six million transactions across more than 30 mutual funds. This model was also used to determine the proportionate settlement amount that was, ultimately, paid to the class members.
- Engaged by a public registrant to reconstruct its accounting records following an investigation by the Audit Committee into revenue recognition issues. We analyzed the underlying systems, processes, controls, and accounting records to assess thousands of revenue transactions under individual contracts in the technology services space. As part of this analysis, StoneTurn requested and received the company's entire Great Plains accounting system in native format, which contained over 1,000 disparate tables. This data was recompiled and used to generate targeted reports related to key accounts and time periods, as well as create charts and graphs used to help visually identify patterns in the company's historic revenue recognition practices, such as those surrounding quarter and year end.
- Engaged to quantify the economic damages related to pricing used to determine reimbursement pursuant to a federal drug rebate program. Specifically, the allegations related to improperly reporting prices based upon the bundled sale of certain types of products. As part of the data analysis, a model was created in SQL to determine the "unbundled" price of products under various scenarios, the result of which were applied in the determination of the appropriate rebate calculated under the program. StoneTurn also assisted in preparation for trial, including the identification and subsequent modeling of damages scenarios based upon specific hospitals or groups of hospitals, as well as other potential legal outcomes. The case was ultimately settled by the parties for an amount consistent with the total damages conclusion generated by the economic model.
- Assisted various companies on matters related to the alleged misappropriation of assets or payment of potential bribes as part of civil litigation, compliance issues and regulatory inquiries. As part of the analysis, StoneTurn frequently worked directly with the subject-company and counsel to help extract the necessary financial and accounting data from systems such as QuickBooks, Great Plains and SAP. This data, once validated, normalized and uploaded to StoneTurn's secure SQL server, was then typically subjected to a number of standard tests (e.g., round dollar, duplicate vendors, Benford's law, etc.), as well as tests based on the facts and circumstances specific to each case. This may involve the identification of relevant transactions based on searches for keywords or individuals, as well as the creation of a dynamic and visual interface that allows the user to quickly explore the dataset to identify trends and anomalies.

Complex Business Disputes

Josh has worked on a wide-variety of dispute matters in litigation and arbitration settings. He has significant experience in quantifying economic damages related to business interruptions, non-interference and non-compete agreements, supply agreements, franchise agreements and other contractual disputes. Representative case experience includes:

- Retained by a pharmaceutical technology company to perform a contract compliance audit in connection with a profit sharing agreement. StoneTurn performed a detailed analysis and recreation of the profit sharing model, as well as sample of the underlying revenue and costs. The compliance audit culminated in a report and supporting exhibits identifying significant underpayments of profits based largely on the inclusion of direct and indirect expenditures that fell outside the scope of contract.
- Engaged by defendant, a leading manufacturer and distributor of clinical laboratory instrumentation for in vitro diagnostic application, to determine economic damages related to defendant's counterclaim involving an alleged breach of contract. The breach related to plaintiff's interference with the attempted auction of a number of defendant's subsidiaries, and the solicitation of key employees.
- Retained by defendant, a national owner and franchisor of travel centers, to evaluate the alleged damages suffered by plaintiff, a global manufacturer of pharmaceutical and its insurance carrier, resulting from a stolen truckload of pharmaceutical drugs.
- Engaged by defendant, a global chemical manufacturer to assist in quantifying damages in a litigation that related to an alleged breach of contract. The contract included a "most favored customer" pricing provision and the parties disputed the interpretation and historical application of the provision.

Intellectual Property

Josh has performed quantitative and qualitative analyses to determine reasonable royalty and lost profit damages for leading software / hardware providers, retailers and financial institutions. He has also provided in-depth examinations of the relevant industries, markets and technologies, as well as each party's financial performance, to assist in the determination of value. Representative case experience includes:

- Performed a fair market valuation of a company that specialized in the synthesis of keratin, including its patent portfolio that contained 28 patents related to the applications of keratin in a range of cosmetic and medical products.
- Retained by defendant, a worldwide manufacturer of mobile devices and electronics, to assess damages and provide rebuttal analyses in connection with an alleged patent infringement. The patent related to the resizing and formatting of digital images prior to being sent via MMS message.
- Engaged by plaintiff to provide both industry and damages expertise in connection with a breach of a software licensing agreement between the licensor and licensee. Analyses included the development of several complex

models in order to compute damages under a number of different assumptions based on the language of the agreement.

- Retained by defendants in a multi-defendant litigation that included OEM computer manufacturers, as well as certain operating system and computer chip suppliers. The lawsuit, brought by a non-practicing entity, asserted alleged infringement against the defendants with respect to four patents that were purported to provide power savings in desktop and laptop computers. Analyses included the determination of a reasonable royalty, as well as a rebuttal of the damages asserted by plaintiff.

Business Valuation

Josh has assisted in the preparation of business valuation analyses and reports for various matters, including sale negotiations, contract disputes and for tax purposes. These reports required the use of both the income and market valuation approaches, and included analyses of discounted cash flows, as well as comparable companies and sale transactions. Representative case experience includes:

- Retained to calculate the value of a minority interest in a privately held international manufacturer of fluid sealing products. The subject company was also a defendant in a significant number of asbestos-related lawsuits that greatly impacted the fair market value, which had to be accounted for within the valuation. The final report was submitted to the IRS for estate tax purposes.
- Retained to perform a valuation of a franchise area development agreement for a company in the heating and air conditioning industry. Analyses included the capitalization method of valuation based upon the historical financial performance of the company and a determination of the appropriate risk-adjusted discount rate.
- Engaged by counsel for plaintiff, an individual stakeholder in an oil and gas distribution company, to assess and critique certain valuation reports that had been used in connection with a potential buy-out of the plaintiffs' interest in the subject company.

Non-Profit

- Josh assisted a local non-profit organization on a pro-bono basis by providing economic analysis related to foster children entering into a pre-college program. The analysis focused on helping the organization quantitatively demonstrate the incremental economic value to the State associated with individuals that hold a college degree.
- Josh assisted a local non-profit organization on a pro-bono basis by providing an analysis of the organization's overhead costs, and how those costs compare to other similar organizations in the industry. The results of this analysis were presented to the Board of Directors.

Exhibit 2: Documents Considered

Legal Filings

Motion by Chapter 11 Trustee for Entry of Order Finding that Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, October 7, 2015 (Doc. #623)
Memorandum in Support of Proposed Supplemental Order Respecting Motion by Chapter 11 Trustee for Entry of Order Finding that Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, January 19, 2016 (Doc. #677)
Supplemental Order Respecting Motion by Chapter 11 Trustee for Entry of Order Finding that Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, January 26, 2016 (Doc. #687)
Amended Complaint, Adversary Proceeding No. 16-04006, April 14, 2016 (Doc. #113)
Government's Trial Memorandum, No. 14-CR-40028-TSH, September 28, 2016 (Doc. #299)
Class Certification Order and Approval of Class Counsel, Adversary Proceeding No. 16-4006, October 6, 2016 (Doc. #194)
Government's Sentencing Memorandum, No. 14-CR-40028-TSH, March 16, 2017 (Doc. #332)
Class Certification Order and Approval of Class Counsel, Adversary Proceeding No. 16-4007, August 3, 2017 (Doc. #479)
TelexFree Analysis of Damages, No. 14-CR-40028-TSH, March 16, 2017 (Doc. #332-1)
Affidavit of Lusette Balan, July 29, 2020
Affidavit of Frantz Balan, July 29, 2020
Affidavit of Sandro Paulo Freitas, July 31, 2020

Expert Reports

Expert Report of Timothy J. Martin, April 13, 2018

Huron Documents

ePOC Data Tables, as provided by Huron
Huron Data Tables, as provided by Huron
Domestic and Foreign Net Winners v2.xlsx, as provided by Huron
TelexFree - User Account Detail.xlsx, as provided by Huron
Telexfree Net Equity Calculation and Aggregation.pdf, as provided by Huron

Other

Christen, Peter; Churches, Tim. "Febrl: Freely extensible biomedical record linkage, Release 0.3". Australian National University, 2002-2005.
Dusetzina SB, Tyree S, Meyer AM, Meyer A, Green L, Carpenter WR. Linking Data for Health Services Research: A Framework and Instructional Guide. (Prepared by the University of North Carolina at Chapel Hill under Contract No. 290-2010-000141.) AHRQ Publication No. 14-EHC033-EF. Rockville, MD: Agency for Healthcare Research and Quality; September 2014. (www.effectivehealthcare.ahrq.gov/reports/final.cfm.)
Kroeze, Harold. "Methodology: Data Linkage." European Commission, February 2017.
Schumacher, Scott. “Probabilistic Versus Deterministic Data Matching : Making an Accurate Decision.” 2010. (https://pdfs.semanticscholar.org/743a/6f53aa08eb0b89b3c4108c2ab36765658998.pdf)
Soko Setoguchi, MD, DrPH; Ying Zhu, MPH; Jessica J. Jalbert, PhD; Lauren A. Williams, BA; Chih-Ying Chen, PhD. Methods Paper "Validity of Deterministic Record Linkage Using Multiple Indirect Personal Identifiers: Linking a Large Registry to Claims Data." Circ Cardiovasc Qual Outcomes, May 2014. (http://ahajournals.org)
Transmittal letter from Andrew Lizotte to Ilyas Rona, January 5, 2017
www.ofx.com/en-us/forex-news/historical-exchange-rates/yearly-average-rates
www.sec.gov/litigation/complaints/2014/comp-pr2014-79.pdf
www.justice.gov/usao-ma/pr/former-president-telexfree-sentenced-billion-dollar-pyramid-scheme.

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 3: User Accounts and Participants by Entity and Type

SUMMARY BY ENTITY (ALL):

Entity	User Accounts	Participants	Total Net Equity
Hybrid	138	113	$8,279,788
TelexFree	10,987,607	2,137,074	($226,190,248) *[A]*
Ympactus	4,006,249	1,234,149	($425,439,930)
Total	**14,993,994**	**3,371,336**	**($643,350,389)**
Unique Count	14,993,994	3,327,992	($643,350,389)

SUMMARY BY NET WINNER AND NET LOSER (ALL):

Participant Type	User Accounts	Participants	Total Equity
Net Loser	9,196,908	2,042,022	($1,809,367,883)
Net Winner	1,790,699	95,052	$1,583,177,635
Total	10,987,607	2,137,074	($226,190,248) *[A]*

Source(s): Based on queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 4: Summary of TelexFree Net Equity by Tier and Transaction Type

				Included in Net Equity						*Excluded from Net Equity*		
Participant Type	**Tiers**	**Total Equity, Including Transfers**	**Total Net Equity**	**1 - Direct Payment**	**2 - Direct Receipt**	**3 - Triangular Payment**	**4 - Triangular Receipt**	**5 - Purchase of Credits**	**8 - Chargebacks**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**5 - Purchase of Credits (Excluded)**
All	a. 0-100	$ (55,883,729)	**$ (61,723,699)**	$ (6,350,040)	$ 1,771,231	$ (87,483,965)	$ 28,860,552	$ (72,529)	$ 1,551,051	$ (22,021,085)	$ 27,861,055	$ (45,500)
All	b. 100-1,000	$ (85,204,419)	**$ (147,224,989)**	$ (29,396,673)	$ 13,221,676	$ (261,771,166)	$ 129,250,382	$ (448,168)	$ 1,918,960	$ (86,034,990)	$ 148,055,560	$ -
All	c. 1,000-5,000	$ (322,041,264)	**$ (776,074,847)**	$ (149,438,063)	$ 33,008,696	$ (1,048,338,884)	$ 384,575,001	$ (1,953,492)	$ 6,071,896	$ (354,649,674)	$ 808,683,257	$ (32,184,929)
All	d. 5,000-10,000	$ (112,840,158)	**$ (197,135,876)**	$ (53,831,028)	$ 20,589,709	$ (405,641,383)	$ 239,935,281	$ (1,718,329)	$ 3,529,873	$ (252,761,851)	$ 337,057,570	$ (1,550,948)
All	e. 10,000-25,000	$ 123,694,778	**$ (125,779,939)**	$ (63,862,445)	$ 26,980,870	$ (455,389,382)	$ 366,849,804	$ (3,670,979)	$ 3,312,194	$ (413,379,390)	$ 662,854,107	$ (237,321)
All	f. 25,000-50,000	$ 42,326,042	**$ 77,905,194**	$ (23,646,642)	$ 18,121,810	$ (186,143,877)	$ 271,167,415	$ (2,967,847)	$ 1,374,335	$ (312,261,698)	$ 276,682,546	$ (2,946,374)
All	g. 50,000-100,000	$ 87,101,808	**$ 162,929,893**	$ (16,085,950)	$ 13,965,856	$ (114,955,642)	$ 282,418,175	$ (3,394,710)	$ 982,165	$ (337,639,337)	$ 261,811,252	$ (14,326,425)
All	h. 100,000-250,000	$ 120,253,309	**$ 289,990,232**	$ (9,034,426)	$ 10,572,174	$ (84,739,775)	$ 378,188,218	$ (5,588,973)	$ 593,014	$ (429,514,497)	$ 259,777,574	$ (7,650,026)
All	i. 250,000-1,000,000	$ 211,394,198	**$ 427,578,150**	$ (6,529,505)	$ 11,200,997	$ (53,645,439)	$ 482,274,368	$ (6,023,635)	$ 301,365	$ (622,120,110)	$ 405,936,158	$ (29,957,640)
All	j. 1,000,000-2,000,000	$ 48,968,394	**$ 81,200,318**	$ (1,349,415)	$ 2,608,874	$ (9,849,404)	$ 92,296,947	$ (2,517,278)	$ 10,595	$ (141,665,881)	$ 109,433,957	$ (17,317,435)
All	k. 2,000,000-5,000,000	$ 19,248,445	**$ 42,145,315**	$ (260,818)	$ 540,114	$ (5,719,817)	$ 49,405,425	$ (1,827,871)	$ 8,283	$ (64,010,771)	$ 41,113,901	$ (93,375)
All	l. 5,000,000+	$ -	**$ -**	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Total All		**$ 77,017,406**	**$ (226,190,248)**	**$ (359,785,006)**	**$ 152,582,007**	**$ (2,713,678,735)**	**$ 2,705,221,567**	**$ (30,183,810)**	**$ 19,653,730**	**$ (3,036,059,284)**	**$ 3,339,266,938**	**$ (106,309,974)**
	Number of Transactions			**440,969**	**126,599**	**13,333,462**	**13,320,684**	**5,281**	**5,064**	**4,461,778**	**4,461,866**	**747**

				Included in Net Equity						*Excluded from Net Equity*		
Participant Type	**Tiers**	**Total Net Equity, Including Transfers**	**Total Net Equity**	**1 - Direct Payment**	**2 - Direct Receipt**	**3 - Triangular Payment**	**4 - Triangular Receipt**	**5 - Purchase of Credits**	**8 - Chargebacks**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**5 - Purchase of Credits (Excluded)**
Net Loser	a. 0-100	$ (59,177,305)	**$ (62,133,114)**	$ (3,731,914)	$ 824,395	$ (76,423,539)	$ 16,781,894	$ (45,325)	$ 461,374	$ (12,932,413)	$ 15,888,222	$ -
Net Loser	b. 100-1,000	$ (108,488,471)	**$ (157,629,250)**	$ (20,492,949)	$ 6,559,807	$ (215,180,908)	$ 71,001,627	$ (264,470)	$ 747,644	$ (41,868,816)	$ 91,009,594	$ -
Net Loser	c. 1,000-5,000	$ (487,578,348)	**$ (851,345,539)**	$ (129,357,455)	$ 13,410,325	$ (939,344,760)	$ 202,855,660	$ (954,985)	$ 2,045,676	$ (186,812,864)	$ 550,580,055	$ (32,184,929)
Net Loser	d. 5,000-10,000	$ (167,756,527)	**$ (274,527,424)**	$ (42,171,409)	$ 6,659,952	$ (338,497,888)	$ 99,662,327	$ (830,344)	$ 649,938	$ (111,859,512)	$ 218,630,409	$ (30,407)
Net Loser	e. 10,000-25,000	$ 30,792,460	**$ (292,252,263)**	$ (47,989,538)	$ 5,567,997	$ (352,476,595)	$ 104,116,395	$ (1,882,920)	$ 412,398	$ (142,218,490)	$ 465,263,213	$ (51,719)
Net Loser	f. 25,000-50,000	$ (50,272,007)	**$ (94,801,934)**	$ (13,452,862)	$ 1,842,080	$ (115,070,596)	$ 32,900,855	$ (1,071,633)	$ 50,223	$ (54,352,327)	$ 98,882,254	$ (101,800)
Net Loser	g. 50,000-100,000	$ (14,601,445)	**$ (45,237,836)**	$ (7,721,437)	$ 1,012,269	$ (55,633,454)	$ 18,424,444	$ (1,364,233)	$ 44,575	$ (32,823,309)	$ 63,459,700	$ (5,360,627)
Net Loser	h. 100,000-250,000	$ (9,640,360)	**$ (22,906,058)**	$ (2,114,009)	$ 227,488	$ (28,399,952)	$ 8,053,996	$ (673,580)	$ -	$ (19,047,886)	$ 32,313,584	$ -
Net Loser	i. 250,000-1,000,000	$ (2,133,904)	**$ (7,523,739)**	$ (808,729)	$ 56,583	$ (7,898,568)	$ 1,553,800	$ (426,825)	$ -	$ (5,288,800)	$ 10,678,635	$ -
Net Loser	j. 1,000,000-2,000,000	$ (755,634)	**$ (1,010,727)**	$ -	$ 1,462	$ (1,457,939)	$ 445,750	$ -	$ -	$ (576,124)	$ 831,217	$ -
Net Loser	k. 2,000,000-5,000,000	$ -	**$ -**	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Net Loser	l. 5,000,000+	$ -	**$ -**	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Total Net Losers		**$ (869,611,541)**	**$ (1,809,367,883)**	**$ (267,840,302)**	**$ 36,162,357**	**$ (2,130,384,199)**	**$ 555,796,749**	**$ (7,514,316)**	**$ 4,411,828**	**$ (607,780,542)**	**$ 1,547,536,883**	**$ (37,729,482)**

				Included in Net Equity						*Excluded from Net Equity*		
Participant Type	**Tiers**	**Total Net Equity, Including Transfers**	**Total Net Equity**	**1 - Direct Payment**	**2 - Direct Receipt**	**3 - Triangular Payment**	**4 - Triangular Receipt**	**5 - Purchase of Credits**	**8 - Chargebacks**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**5 - Purchase of Credits (Excluded)**
Net Winner	a. 0-100	$ 3,293,577	**$ 409,415**	$ (2,618,127)	$ 946,837	$ (11,060,427)	$ 12,078,658	$ (27,204)	$ 1,089,677	$ (9,088,671)	$ 11,972,833	$ (45,500)
Net Winner	b. 100-1,000	$ 23,284,052	**$ 10,404,261**	$ (8,903,724)	$ 6,661,869	$ (46,590,258)	$ 58,248,755	$ (183,697)	$ 1,171,316	$ (44,166,174)	$ 57,045,966	$ -
Net Winner	c. 1,000-5,000	$ 165,537,084	**$ 75,270,692**	$ (20,080,608)	$ 19,598,370	$ (108,994,124)	$ 181,719,341	$ (998,507)	$ 4,026,220	$ (167,836,810)	$ 258,103,202	$ -
Net Winner	d. 5,000-10,000	$ 54,916,370	**$ 77,391,548**	$ (11,659,619)	$ 13,929,757	$ (67,143,495)	$ 140,272,954	$ (887,984)	$ 2,879,935	$ (140,902,339)	$ 118,427,161	$ (1,520,542)
Net Winner	e. 10,000-25,000	$ 92,902,319	**$ 166,472,324**	$ (15,872,907)	$ 21,412,873	$ (102,912,788)	$ 262,733,409	$ (1,788,059)	$ 2,899,796	$ (271,160,900)	$ 197,590,894	$ (185,602)
Net Winner	f. 25,000-50,000	$ 92,598,049	**$ 172,707,128**	$ (10,193,780)	$ 16,279,731	$ (71,073,281)	$ 238,266,561	$ (1,896,214)	$ 1,324,112	$ (257,909,371)	$ 177,800,292	$ (2,844,574)
Net Winner	g. 50,000-100,000	$ 101,703,254	**$ 208,167,729**	$ (8,364,513)	$ 12,953,587	$ (59,322,188)	$ 263,993,731	$ (2,030,477)	$ 937,589	$ (304,816,028)	$ 198,351,552	$ (8,965,798)
Net Winner	h. 100,000-250,000	$ 129,893,669	**$ 312,896,290**	$ (6,920,417)	$ 10,344,686	$ (56,339,822)	$ 370,134,222	$ (4,915,393)	$ 593,014	$ (410,466,611)	$ 227,463,990	$ (7,650,026)
Net Winner	i. 250,000-1,000,000	$ 213,528,102	**$ 435,101,889**	$ (5,720,776)	$ 11,144,414	$ (45,746,872)	$ 480,720,568	$ (5,596,810)	$ 301,365	$ (616,831,310)	$ 395,257,523	$ (29,957,640)
Net Winner	j. 1,000,000-2,000,000	$ 49,724,028	**$ 82,211,045**	$ (1,349,415)	$ 2,607,412	$ (8,391,465)	$ 91,851,197	$ (2,517,278)	$ 10,595	$ (141,089,757)	$ 108,602,740	$ (17,317,435)
Net Winner	k. 2,000,000-5,000,000	$ 19,248,445	**$ 42,145,315**	$ (260,818)	$ 540,114	$ (5,719,817)	$ 49,405,425	$ (1,827,871)	$ 8,283	$ (64,010,771)	$ 41,113,901	$ (93,375)
Net Winner	l. 5,000,000+	$ -	**$ -**	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Total Net Winners		**$ 946,628,948**	**$ 1,583,177,635**	**$ (91,944,704)**	**$ 116,419,649**	**$ (583,294,536)**	**$ 2,149,424,818**	**$ (22,669,494)**	**$ 15,241,902**	**$ (2,428,278,743)**	**$ 1,791,730,055**	**$ (68,580,492)**

Source(s): Based on queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 5: Summary of TelexFree User Accounts and Participants by Tier

		Huron Data Tables (SIG)								*ePOC Data Tables (Claims)*		
Participant Type	**Tiers**	**User Accounts**	**Avg. User Accounts per Aggregation**	**Avg. Net Equity per User Account**	**Participant Account Aggregations**	**Participant %**	**Avg. Net Equity per Participant**	**Total Equity**	**Total Equity %**	**Cumulative Total Equity %**	**# of Account Aggregations Claimed** [1]	**% of Total Account Aggregations Claimed**	**TelexFree Net Equity Claimed** [1]
All	a. 0-100	**1,270,024**	1	n/a	**1,184,220**	55.4%	n/a	**$ (61,723,699)**	n/a	n/a	6,697	0.31%	n/a
All	b. 100-1,000	**1,734,497**	5	n/a	**383,695**	18.0%	n/a	**$ (147,224,989)**	n/a	n/a	25,476	1.19%	n/a
All	c. 1,000-5,000	**3,296,955**	7	n/a	**472,790**	22.1%	n/a	**$ (776,074,847)**	n/a	n/a	69,839	3.27%	n/a
All	d. 5,000-10,000	**1,517,999**	30	n/a	**50,862**	2.4%	n/a	**$ (197,135,876)**	n/a	n/a	11,146	0.52%	n/a
All	e. 10,000-25,000	**1,638,142**	53	n/a	**30,777**	1.4%	n/a	**$ (125,779,939)**	n/a	n/a	7,301	0.34%	n/a
All	f. 25,000-50,000	**648,612**	84	n/a	**7,758**	0.4%	n/a	**$ 77,905,194**	n/a	n/a	1,602	0.07%	n/a
All	g. 50,000-100,000	**384,286**	105	n/a	**3,674**	0.2%	n/a	**$ 162,929,893**	n/a	n/a	616	0.03%	n/a
All	h. 100,000-250,000	**304,835**	138	n/a	**2,211**	0.1%	n/a	**$ 289,990,232**	n/a	n/a	341	0.02%	n/a
All	i. 250,000-1,000,000	**163,147**	162	n/a	**1,010**	0.0%	n/a	**$ 427,578,150**	n/a	n/a	114	0.01%	n/a
All	j. 1,000,000-2,000,000	**20,662**	339	n/a	**61**	0.0%	n/a	**$ 81,200,318**	n/a	n/a	6	0.00%	n/a
All	k. 2,000,000-5,000,000	**8,453**	528	n/a	**16**	0.0%	n/a	**$ 42,145,315**	n/a	n/a	2	0.00%	n/a
All	l. 5,000,000+	**-**	-	n/a	**-**	0.0%	n/a	**$ -**	n/a	n/a	-	0.00%	n/a
Total All		**10,987,612**	**5**	n/a	**2,137,074**	**100.0%**	n/a	**$ (226,190,248)**	n/a	n/a	**123,140**	**5.76%**	n/a

Total Number of Unique User with Claims [1] **85,616**
Est. Percentage of Participants that Submitted Claims **4.01%**

		Huron Data Tables (SIG)								*ePOC Data Tables (Claims)*		
Participant Type	**Tiers**	**User Accounts**	**Avg. User Accounts per Aggregation**	**Avg. Net Equity per User Account**	**Participant Account Aggregations**	**Participant %**	**Avg. Net Equity per Participant**	**Total Equity**	**Total Equity %**	**Cumulative Total Equity %**	**# of Account Aggregations Claimed** [1]	**% of Total Account Aggregations Claimed**	**TelexFree Net Equity Claimed** [1]
Net Loser	a. 0-100	**1,243,515**	1	$ (50)	**1,175,480**	57.6%	$ (53)	**$ (62,133,114)**	3.4%	3.4%	6,031	0.28%	$ (286,630)
Net Loser	b. 100-1,000	**1,608,799**	4	$ (98)	**360,150**	17.6%	$ (438)	**$ (157,629,250)**	8.7%	12.1%	23,430	1.10%	$ (12,906,466)
Net Loser	c. 1,000-5,000	**2,986,680**	7	$ (285)	**442,457**	21.7%	$ (1,924)	**$ (851,345,539)**	47.1%	59.2%	66,628	3.12%	$ (140,408,781)
Net Loser	d. 5,000-10,000	**1,305,076**	33	$ (210)	**39,983**	2.0%	$ (6,866)	**$ (274,527,424)**	15.2%	74.4%	9,822	0.46%	$ (68,014,478)
Net Loser	e. 10,000-25,000	**1,302,038**	64	$ (224)	**20,233**	1.0%	$ (14,444)	**$ (292,252,263)**	16.2%	90.5%	5,948	0.28%	$ (86,456,229)
Net Loser	f. 25,000-50,000	**401,127**	141	$ (236)	**2,853**	0.1%	$ (33,229)	**$ (94,801,934)**	5.2%	95.8%	921	0.04%	$ (30,528,785)
Net Loser	g. 50,000-100,000	**192,764**	283	$ (235)	**682**	0.0%	$ (66,331)	**$ (45,237,836)**	2.5%	98.3%	223	0.01%	$ (14,791,439)
Net Loser	h. 100,000-250,000	**117,007**	705	$ (196)	**166**	0.0%	$ (137,988)	**$ (22,906,058)**	1.3%	99.5%	44	0.00%	$ (6,058,446)
Net Loser	i. 250,000-1,000,000	**38,671**	2,275	$ (195)	**17**	0.0%	$ (442,573)	**$ (7,523,739)**	0.4%	99.9%	6	0.00%	$ (2,454,033)
Net Loser	j. 1,000,000-2,000,000	**1,232**	1,232	$ (820)	**1**	0.0%	$ (1,010,727)	**$ (1,010,727)**	0.1%	100.0%	-	0.00%	$ -
Net Loser	k. 2,000,000-5,000,000	**-**	-	$ -	**-**	0.0%	$ -	**$ -**	0.0%	100.0%	-	0.00%	$ -
Net Loser	l. 5,000,000+	**-**	-	$ -	**-**	0.0%	$ -	**$ -**	0.0%	100.0%	-	0.00%	$ -
Total Net Losers		**9,196,909**	**5**	$ (197)	**2,042,022**	**100.0%**	$ (886)	**$ (1,809,367,883)**	**100.0%**	**100.0%**	**113,053**	**5.29%**	**$ (361,905,287)**

		Huron Data Tables (SIG)								*ePOC Data Tables (Claims)*		
Participant Type	**Tiers**	**User Accounts**	**Avg. User Accounts per Aggregation**	**Avg. Net Equity per User Account**	**Participant Account Aggregations**	**Participant %**	**Avg. Net Equity per Participant**	**Total Equity**	**Total Equity %**	**Cumulative Total Equity %**	**# of Account Aggregations Claimed** [1]	**% of Total Account Aggregations Claimed**	**TelexFree Net Equity Claimed** [1]
Net Winner	a. 0-100	**26,509**	3	$ 15	**8,740**	9.2%	$ 47	**$ 409,415**	0.0%	0.0%	666	0.03%	$ 34,594
Net Winner	b. 100-1,000	**125,698**	5	$ 83	**23,545**	24.8%	$ 442	**$ 10,404,261**	0.7%	0.7%	2,046	0.10%	$ 944,918
Net Winner	c. 1,000-5,000	**310,275**	10	$ 243	**30,333**	31.9%	$ 2,481	**$ 75,270,692**	4.8%	5.4%	3,211	0.15%	$ 8,051,163
Net Winner	d. 5,000-10,000	**212,923**	20	$ 363	**10,879**	11.4%	$ 7,114	**$ 77,391,548**	4.9%	10.3%	1,324	0.06%	$ 9,472,582
Net Winner	e. 10,000-25,000	**336,104**	32	$ 495	**10,544**	11.1%	$ 15,788	**$ 166,472,324**	10.5%	20.8%	1,353	0.06%	$ 21,391,857
Net Winner	f. 25,000-50,000	**247,485**	50	$ 698	**4,905**	5.2%	$ 35,210	**$ 172,707,128**	10.9%	31.7%	681	0.03%	$ 23,711,122
Net Winner	g. 50,000-100,000	**191,522**	64	$ 1,087	**2,992**	3.1%	$ 69,575	**$ 208,167,729**	13.1%	44.9%	393	0.02%	$ 26,847,712
Net Winner	h. 100,000-250,000	**187,828**	92	$ 1,666	**2,045**	2.2%	$ 153,006	**$ 312,896,290**	19.8%	64.7%	297	0.01%	$ 45,959,512
Net Winner	i. 250,000-1,000,000	**124,476**	125	$ 3,495	**993**	1.0%	$ 438,169	**$ 435,101,889**	27.5%	92.1%	108	0.01%	$ 43,464,757
Net Winner	j. 1,000,000-2,000,000	**19,430**	324	$ 4,231	**60**	0.1%	$ 1,370,184	**$ 82,211,045**	5.2%	97.3%	6	0.00%	$ 8,316,422
Net Winner	k. 2,000,000-5,000,000	**8,453**	528	$ 4,986	**16**	0.0%	$ 2,634,082	**$ 42,145,315**	2.7%	100.0%	2	0.00%	$ 2,034,063
Net Winner	l. 5,000,000+	**-**	-	$ -	**-**	0.0%	$ -	**$ -**	0.0%	100.0%	-	0.00%	$ -
Total Net Winners		**1,790,703**	**19**	$ 884	**95,052**	**100.0%**	$ 16,656	**$ 1,583,177,635**	**100.0%**	**100.0%**	**10,087**	**0.47%**	**$ 190,228,702**

[1] Includes Participant Account Aggregations where at least one underlying User Account matched to an Account Information Set, and which was included as part of an accepted and agreed claim within ePOC.

Source(s): Based on queries of the Huron Data Tables and ePOC Data Tables.

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 6: Summary of TelexFree Top Net Loser Matched to Net Winners

Top Net Loser Detail (Net Losses Over $200,000)								*Matched Net Winner(s)* [1]					*Similar/Matching Values*			
Num	**Top Net Loser Parent ID**	**Was Claimed?**	**Name**	**Login**	**Email**	**Number of Telex Accounts**	**Net Loser Equity**	**Identified Net Winner Parent Id**	**Was Claimed?**	**Net Winner Equity**	**Adjusted Total Net Equity** [2]	**Transacted Together?**	**Login**	**Email**	**Address**	**Phone**
1	13043527	No	Du Painting	bancotelex1	pontualsucesso7@gmail.com	1,232	**$ (1,010,727)**	4032900	No	**$ 1,544,904**	**$ 534,177**	X	X	X	X	X
2	21036267	No	ming hua	liuppp	15900554272@163.com	16,909	**$ (843,759)**	17401199	No	**$ 46,519**	**$ (797,240)**			X	X	
3	15152377	No	fan fan	fan16801	kwan1987311@gmail.com	665	**$ (721,474)**	4452796	No	**$ 192,955**	**$ (528,519)**			X	X	
4	12650741	***Yes***	Shall Group LLC	shall1	ajaff@shalsd.com	509	**$ (708,133)**	10992563	***Yes***	**$ 190,381**	**$ (517,752)**	X		X	X	X
5	11849449	No	DL1 INC	dlld1	davidbeeba@gmail.com	928	**$ (622,342)**	707775	No	**$ 1,216,914**	**$ 594,573**	X	X	X	X	X
6	12242105	No	Jun Zhang	art124	telexusa127@gmail.com	517	**$ (481,228)**	4987334	No	**$ 947,923**	**$ 466,695**	X	X			X
7	12072673	***Yes***	telex boston inc	bboston33	edstorck@hotmail.com	378	**$ (438,338)**	1922040	No	**$ 1,299,826**	**$ 1,754,724**	X	X	X	X	X
9	2876246	No	edilene	storck11	edstorck@hotmail.com	1,142	**$ (406,591)**									
8	19298561	No	yabin zhang	sjzgdm	vition@vitionst.com	8,163	**$ (407,384)**	18296273	No	**$ 43,208**	**$ (364,176)**	X		X		
10	15747789	No	qun zheng	usqun001	py900531@gmail.com	371	**$ (390,125)**	16655191	No	**$ 171,287**	**$ 1,162,602**	X		X	X	X
								10844649	No	**$ 1,381,440**		X		X	X	
11	3392494	***Yes***	Youqin Wang	md16888	yuc6688@gmail.com	351	**$ (377,696)**	4486096	No	**$ 28,223**		X	X	X	X	X
								4987334	No	**$ 947,923**	**$ 985,287**	X	X			
								3780002	No	**$ 386,838**		X	X	X		
12	15084253	***Yes***	bryan matt	quincy68	grupoboston@icloud.com	266	**$ (360,775)**	632789	No	**$ 968,101**	**$ 607,327**	X	X	X		X
13	10630965	No	Genaro Olortegui Carrera	mainnetwork2	olorteguiomar@live.com	431	**$ (349,610)**	312457	No	**$ 259,814**	**$ 121,076**	X	X	X		
								4993498	No	**$ 210,873**		X	X	X	X	X
14	15772887	***Yes***	Lai Yee FOOK	fook1	rizatelex@gmail.com	6,759	**$ (307,009)**	*n/a*	No		***n/a***	***na/***	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
15	7127624	No	Imad Thweib	imadthweib	zthweib@gmail.com	319	**$ (291,301)**	1183234	No	**$ 40,899**	**$ 563,672**	X	X	X		
								1309022	No	**$ 814,074**		X		X		
16	117321	No	renato simao honorio dos santos	sirenato1	renato_simao@hotmail.com	738	**$ (278,073)**	11034097	No	**$ 444,520**	**$ 183,746**	X		X		
								12294889	No	**$ 17,300**		X	X	X	X	X
17	227597	No	Cleonice de Campos leite Ribeiro	becca	nicecr@hotmail.com	12	**$ (277,789)**	10525055	No	**$ 34,124**	**$ (243,665)**		X	X	X	X
18	13531273	***Yes***	Gilson Da Silva	sirleneduarte14	sirleneduartetelex@gmail.com	213	**$ (262,114)**	966001	No	**$ 819,345**	**$ 557,232**	X		X		
19	10811451	No	Oscar Sosa	oscarsosa	sosaos@live.com	1,027	**$ (241,723)**	10838191	No	**$ 295,498**	**$ 53,776**	X	X	X	X	X
20	29758080	No	MGS Marketing Inc.	mgsteam	mgsmarketinginc@gmail.com	2,705	**$ (240,141)**	10539115	No	**$ 1,677,542**	**$ 1,437,402**	X			X	X
21	17439659	No	cincin zhao	npnp007	telexchina668@gmail.com	165	**$ (235,125)**	17248041	No	**$ 128,244**		X	X	X		X
								5108450	No	**$ 120,730**	**$ 89,586**					X
								17957689	No	**$ 75,737**		X				X
22	11487123	***Yes***	BARRY MILLER	barry02	nichelexrosa@gmail.com	383	**$ (230,922)**	11482181	No	**$ 53,340**	**$ 47,111**			X		
								10594737	No	**$ 224,693**				X	X	
23	14473091	No	Jônotas Lucas	i9inove	arthurbrazsoares@gmail.com	213	**$ (226,402)**	13557817	No	**$ 368,173**	**$ 194,327**	X		X	X	X
								13559825	No	**$ 52,556**		X		X	X	X
24	15486069	No	EDIVALDO REIS	aide001	baiakclean@yahoo.com	170	**$ (224,723)**	11583717	No	**$ 3,628**	**$ (221,095)**	X		X		
25	12112907	No	Guixiang Sun	art888	mingmz876@gmail.com	264	**$ (221,880)**	4987334	No	**$ 947,923**	**$ 726,044**	X		X	X	X
26	10572419	No	Brahim Boulbahaiem	bb333999002	bb33399@hotmail.com	169	**$ (214,739)**	10572045	No	**$ 106,857**	**$ (107,882)**	X	X		X	X
27	14745823	***Yes***	anne wang	1anne	zhonglinzhang@yahoo.com	193	**$ (210,408)**	10396601	No	**$ 288,064**	**$ 77,656**	X		X		
28	12242931	***Yes***	Vacerly S. Petty	vpetty	rhadassausa@hotmail.com	204	**$ (209,781)**	2004122	***Yes***	**$ 199,241**	**$ 389,365**	X	X	X		
								5510906	No	**$ 399,904**		X		X		
29	17443697	No	grupo aguia usa corp	aguiaus1	foxbrasil10@gmail.com	938	**$ (205,032)**	147979	No	**$ 2,405,821**	**$ 2,200,790**	X	X	X	X	X
30	17495171	No	Calvary Assembly of God	nlm01	zieffj@yahoo.com	162	**$ (204,723)**	149697	No	**$ 1,228,447**	**$ 5,120,413**			X		
								132785	No	**$ 4,096,689**				X		
31	17369623	No	francisca Alarcon	paquia	mauricioalarconmartinez@gma	286	**$ (202,659)**	2615682	No	**$ 327,837**	**$ 125,178**	X		X		
	Total					**46,782**	**$ (11,402,723)**			**$ 25,008,315**	**$ 15,212,427**					
							Total Where Net Loser and Net Winner Transacted Together	**$ (8,410,456)**		**$ 19,010,819**	**$ 11,614,328**					

[1] See Exhibits 9.1 to 9.31.

[2] Calculated as the sum of the Net Loser Equity and matched New Winner(s) Equity.

Source(s): Based upon queries of the Huron Data Tables and ePOC Data Tables.

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 7: Top 100 TelexFree Net Losers

								Included in Net Equity						*Excluded from Net Equity*				
Num	**Rep. Id.**	**Name**	**Login**	**Email**	**Number of Telex Accounts**	**Was Claimed?**	**Total Net Equity**	**1 - Direct Payment**	**2 - Direct Receipt**	**3 - Triangular Payment**	**4 - Triangular Receipt**	**5 - Purchase of Credits**	**8 - Chargebacks**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**5 - Purchase of Credits (Excluded)**	**Total Net Equity, Including Transfers**	***Net Equity Variance From Transfers***
1	13043527	Du Painting	bancotelex1	pontualsucesso7@gmail.com	1,232	No	**$ (1,010,727)**		$ 1,462	$ (1,457,939)	$ 445,750	$ -	$ -	$ (576,124)	$ 831,217	**$ -**	**$ (755,634)**	*$ 255,093*
2	21036267	ming hua	liuppp	15900554272@163.com	16,909	No	**$ (843,759)**			$ (843,759)		$ -	$ -	$ -	$ -	**$ -**	**$ (843,759)**	*$ -*
3	15152377	fan fan	fan16801	kwan1987311@gmail.com	665	No	**$ (721,474)**			$ (743,631)	$ 22,157	$ -	$ -	$ (751,636)	$ 1,501,823	**$ -**	**$ 28,712**	*$ 750,187*
4	12650741	Shall Group LLC	shall1	ajaff@shalsd.com	509	***Yes***	**$ (708,133)**			$ (710,509)	$ 2,376	$ -	$ -	$ -	$ 469,662	**$ -**	**$ (238,471)**	*$ 469,662*
5	11849449	DL1 INC	dlld1	davidbeeba@gmail.com	928	No	**$ (622,342)**	$ (481,770)		$ (837,034)	$ 696,463	$ -	$ -	$ (1,904,355)	$ 2,254,352	**$ -**	**$ (272,344)**	*$ 349,997*
6	12242105	Jun Zhang	art124	telexusa127@gmail.com	517	No	**$ (481,228)**			$ (726,956)	$ 245,729	$ -	$ -	$ (621,865)	$ 1,008,813	**$ -**	**$ (94,279)**	*$ 386,949*
7	12072673	telex boston inc	bboston33	edstorck@hotmail.com	378	***Yes***	**$ (438,338)**			$ (463,110)	$ 24,772	$ -	$ -	$ (178,774)	$ 317,416	**$ -**	**$ (299,695)**	*$ 138,643*
8	19298561	yabin zhang	sjzgdm	vition@vitionst.com	8,163	No	**$ (407,384)**			$ (407,384)		$ -	$ -	$ -	$ -	**$ -**	**$ (407,384)**	*$ -*
9	2876246	edilene	storck11	edstorck@hotmail.com	1,142	No	**$ (406,591)**			$ (424,675)	$ 18,084	$ -	$ -	$ (54,842)	$ 1,243,613	**$ -**	**$ 782,181**	*$ 1,188,771*
10	15747789	qun zheng	usqun001	py900531@gmail.com	371	No	**$ (390,125)**			$ (393,000)	$ 2,875	$ -	$ -	$ (138,705)	$ 285,110	**$ -**	**$ (243,720)**	*$ 146,405*
11	3392494	Youqin Wang	md16888	yuc6688@gmail.com	351	***Yes***	**$ (377,696)**			$ (408,783)	$ 31,087	$ -	$ -	$ (468,158)	$ 748,238	**$ -**	**$ (97,616)**	*$ 280,080*
12	15084253	bryan matt	quincy68	grupoboston@icloud.com	266	***Yes***	**$ (360,775)**			$ (361,822)	$ 1,048	$ -	$ -	$ -	$ 95,009	**$ -**	**$ (265,766)**	*$ 95,009*
13	10630965	Genaro Olortegui Carrera	mainnetwork2	olorteguiomar@live.com	431	No	**$ (349,610)**			$ (362,309)	$ 12,698	$ -	$ -	$ (36,782)	$ 413,935	**$ -**	**$ 27,543**	*$ 377,153*
14	15772887	Lai Yee FOOK	fook1	rizatelex@gmail.com	6,759	***Yes***	**$ (307,009)**	$ (307,009)	$ 300	$ (262,993)	$ 262,694	$ -	$ -	$ (297)	$ 300	**$ -**	**$ (307,006)**	*$ 3*
15	7127624	Imad Thweib	imadthweib	zthweib@gmail.com	319	No	**$ (291,301)**	$ (14,250)	$ 56,283	$ (366,070)	$ 35,236	$ (2,500)	$ -	$ (131,332)	$ 491,316	**$ -**	**$ 68,682**	*$ 359,983*
16	117321	renato simao honorio dos santos	sirenato1	renato_simao@hotmail.com	738	No	**$ (278,073)**			$ (320,077)	$ 42,004	$ -	$ -	$ (112,429)	$ 215,698	**$ -**	**$ (174,804)**	*$ 103,269*
17	227597	Cleonice de Campos leite Ribeiro	becca	nicecr@hotmail.com	12	No	**$ (277,789)**			$ (7,966)	$ 154,502	$ (424,325)	$ -	$ (755,701)	$ 1,387,346	**$ -**	**$ 353,856**	*$ 631,645*
18	13531273	Gilson Da Silva	sirleneduarte14	sirleneduartetelex@gmail.com	213	***Yes***	**$ (262,114)**	$ (5,700)		$ (258,490)	$ 2,076	$ -	$ -	$ (133,925)	$ 246,003	**$ -**	**$ (150,036)**	*$ 112,078*
19	10811451	Oscar Sosa	oscarsosa	sosaos@live.com	1,027	No	**$ (241,723)**	$ (14,736)	$ 600	$ (306,175)	$ 78,589	$ -	$ -	$ (66,228)	$ 235,447	**$ -**	**$ (72,504)**	*$ 169,219*
20	29758080	MGS Marketing Inc.	mgsteam	mgsmarketinginc@gmail.com	2,705	No	**$ (240,141)**			$ (240,141)		$ -	$ -	$ -	$ -	**$ -**	**$ (240,141)**	*$ -*
21	17439659	cincin zhao	npnp007	telexchina668@gmail.com	165	No	**$ (235,125)**			$ (235,125)		$ -	$ -	$ -	$ 27,643	**$ -**	**$ (207,482)**	*$ 27,643*
22	11487123	BARRY MILLER	barry02	nichelexrosa@gmail.com	383	***Yes***	**$ (230,922)**			$ (300,324)	$ 69,402	$ -	$ -	$ (54,229)	$ 74,593	**$ -**	**$ (210,558)**	*$ 20,364*
23	14473091	Jônotas Lucas	i9inove	arthurbrazsoares@gmail.com	213	No	**$ (226,402)**			$ (289,626)	$ 63,223	$ -	$ -	$ (16,976)	$ 112,397	**$ -**	**$ (130,981)**	*$ 95,421*
24	15486069	EDIVALDO REIS	aide001	baiakclean@yahoo.com	170	No	**$ (224,723)**			$ (227,673)	$ 2,950	$ -	$ -	$ (10,144)	$ 31,031	**$ -**	**$ (203,836)**	*$ 20,887*
25	12112907	Guixiang Sun	art888	mingmz876@gmail.com	264	No	**$ (221,880)**			$ (357,570)	$ 135,690	$ -	$ -	$ (556,490)	$ 936,884	**$ -**	**$ 158,515**	*$ 380,394*
26	10572419	Brahim Boulbahaiem	bb333999002	bb33399@hotmail.com	169	No	**$ (214,739)**	$ (115,425)	$ 16,600	$ (116,264)	$ 349	$ -	$ -	$ (22,262)	$ 74,500	**$ -**	**$ (162,501)**	*$ 52,238*
27	14745823	anne wang	1anne	zhonglinzhang@yahoo.com	193	***Yes***	**$ (210,408)**			$ (217,565)	$ 7,157	$ -	$ -	$ -	$ 111,701	**$ -**	**$ (98,707)**	*$ 111,701*
28	12242931	Vacerly S. Petty	vpetty	rhadassausa@hotmail.com	204	***Yes***	**$ (209,781)**			$ (226,562)	$ 16,782	$ -	$ -	$ (41,702)	$ 186,162	**$ -**	**$ (65,320)**	*$ 144,460*
29	17443697	grupo aguia usa corp	aguiaus1	foxbrasil10@gmail.com	938	No	**$ (205,032)**			$ (205,032)		$ -	$ -	$ -	$ 9,401	**$ -**	**$ (195,631)**	*$ 9,401*
30	17495171	Calvary Assembly of God	nlm01	zieffj@yahoo.com	162	No	**$ (204,723)**			$ (204,723)		$ -	$ -	$ -	$ 11,929	**$ -**	**$ (192,794)**	*$ 11,929*
31	17369623	francisca Alarcon	paquia	mauricioalarconmartinez@gmail	286	No	**$ (202,659)**			$ (206,173)	$ 3,514	$ -	$ -	$ (1,288)	$ 85,162	**$ -**	**$ (118,785)**	*$ 83,874*
32	865905	RENATO DE SOUSA BARROS	renatobarrosrn	renatodsbarros@hotmail.com	265	No	**$ (199,778)**			$ (218,390)	$ 18,612	$ -	$ -	$ (31,684)	$ 147,032	**$ -**	**$ (84,431)**	*$ 115,347*
33	1020407	Norberto Rey	norber19526	reytrucking@yahoo.com	1,649	No	**$ (198,009)**	$ (22,392)	$ 12,566	$ (1,439,438)	$ 1,314,281	$ (63,026)	$ -	$ (5,143,557)	$ 6,146,961	**$ -**	**$ 805,396**	*$ 1,003,405*
34	11725985	Valdemir Campos	rima	valdemircampos25@gmail.com	15	No	**$ (197,936)**			$ (17,819)	$ 134,113	$ (314,230)	$ -	$ (635,000)	$ 843,900	**$ -**	**$ 10,964**	*$ 208,900*
35	11731233	Luana Polini	ashleyp1	pedropolini123@hotmail.com	494	No	**$ (189,445)**			$ (192,100)	$ 2,656	$ -	$ -	$ -	$ 100,154	**$ -**	**$ (89,290)**	*$ 100,154*
36	3995992	julio paz	mirandausaclien	jcp68@msn.com	348	No	**$ (187,774)**			$ (457,240)	$ 286,366	$ (16,900)	$ -	$ (443,714)	$ 574,511	**$ -**	**$ (56,977)**	*$ 130,797*
37	12535321	Rafael Cubero	rexyboy	cuberor6@gmail.com	196	***Yes***	**$ (186,576)**			$ (214,799)	$ 28,224	$ -	$ -	$ (139,604)	$ 256,874	**$ -**	**$ (69,306)**	*$ 117,270*
38	11583925	JOSUE JACOB MADRID DERAS	jderas01	jjderas83@hotmail.com	1,826	No	**$ (186,222)**			$ (370,054)	$ 183,832	$ -	$ -	$ (142,660)	$ 163,166	**$ -**	**$ (165,716)**	*$ 20,506*
39	16262197	Sarafina Pinto Gomes	sonnen1	ck.telexfree@gmail.com	147	***Yes***	**$ (178,821)**	$ (19,950)	$ 21,078	$ (180,249)	$ 299	$ -	$ -	$ (18,845)	$ 27,493	**$ -**	**$ (170,173)**	*$ 8,648*
40	13478333	Black Diamond LLP	blackdi001	blackdiamondboston@gmail.com	172	No	**$ (178,769)**			$ (225,781)	$ 47,012	$ -	$ -	$ (226,630)	$ 424,291	**$ -**	**$ 18,892**	*$ 197,660*
41	16580483	daniel covelo	onicial01	uniondois@yahoo.com	197	No	**$ (178,265)**	$ (34,200)		$ (200,560)	$ 56,495	$ -	$ -	$ (104,060)	$ 155,970	**$ -**	**$ (126,355)**	*$ 51,910*
42	11943027	Paulo Goncalves	skymaster1	Paulopefg@gmail.com	684	No	**$ (177,799)**			$ (184,587)	$ 6,788	$ -	$ -	$ (139,317)	$ 272,517	**$ -**	**$ (44,599)**	*$ 133,200*
43	15008947	Sergio Souza	sergusa1	sergiosouza07@hotmail.com	1,456	No	**$ (177,622)**			$ (179,218)	$ 1,597	$ -	$ -	$ (87,878)	$ 149,030	**$ -**	**$ (116,470)**	*$ 61,152*
44	16698911	Jack Zheng	xml016c1	james3599@gmail.com	3,545	No	**$ (177,494)**			$ (177,494)		$ -	$ -	$ -	$ -	**$ -**	**$ (177,494)**	*$ -*
45	12401359	leonel dias	cdias1	njas0210@gmail.com	203	***Yes***	**$ (175,982)**	$ (179,550)		$ (105,949)	$ 109,517	$ -	$ -	$ (144,847)	$ 177,366	**$ -**	**$ (143,463)**	*$ 32,519*
46	13296817	shuang zhao	zs01	zhaoshuang76@163.com	516	***Yes***	**$ (175,825)**	$ (123,975)		$ (190,517)	$ 138,668	$ -	$ -	$ (213,169)	$ 276,099	**$ -**	**$ (112,895)**	*$ 62,930*
47	15564537	Otilio Prado	otiliotelex	fredgladiator@gmail.com	176	No	**$ (173,794)**			$ (174,942)	$ 1,148	$ -	$ -	$ (58,816)	$ 145,851	**$ -**	**$ (86,759)**	*$ 87,035*
48	4511520	miguel angel polinar arias	mapa1804	chizas84@gmail.com	219	No	**$ (173,727)**	$ (2,950)		$ (188,527)	$ 124,625	$ (106,875)	$ -	$ (21,024)	$ 120,390	**$ -**	**$ (74,361)**	*$ 99,366*
49	12812693	Pik Ha Wai	rw88	rebeccawai38@gmail.com	1,764	***Yes***	**$ (172,525)**	$ (153,900)		$ (185,390)	$ 166,765	$ -	$ -	$ (46,266)	$ 45,241	**$ -**	**$ (173,550)**	*$ (1,025)*
50	12429741	jose enrique leça graterol	enriquetb	enrique.lg@hotmail.com	157	No	**$ (172,126)**	$ (94,050)		$ (102,594)	$ 24,519	$ -	$ -	$ (67,518)	$ 99,125	**$ -**	**$ (140,518)**	*$ 31,607*
51	11332681	ying pan	ying168	kwan1987311@gmail.com	544	No	**$ (170,542)**		$ 300	$ (223,050)	$ 52,208	$ -	$ -	$ (372,095)	$ 497,039	**$ -**	**$ (45,598)**	*$ 124,944*
52	16850333	Neildes beatriz Martins	robertoit	jefferson@mundonetwork.com.b	179	No	**$ (169,241)**			$ (175,459)	$ 6,218	$ -	$ -	$ (44,834)	$ 43,546	**$ -**	**$ (170,529)**	*$ (1,288)*
53	656321	jancarla Bellido	cliente74	snemes27@hotmail.com	142	No	**$ (169,202)**			$ (176,168)	$ 6,965	$ -	$ -	$ (43,519)	$ 295,073	**$ -**	**$ 82,352**	*$ 251,554*
54	14184447	william araujo	equipecia	williams.williams@hotmail.com	201	***Yes***	**$ (168,871)**			$ (196,170)	$ 27,299	$ -	$ -	$ (71,223)	$ 93,460	**$ -**	**$ (146,634)**	*$ 22,237*

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 7: Top 100 TelexFree Net Losers

								Included in Net Equity						*Excluded from Net Equity*				
Num	**Rep. Id.**	**Name**	**Login**	**Email**	**Number of Telex Accounts**	**Was Claimed?**	**Total Net Equity**	**1 - Direct Payment**	**2 - Direct Receipt**	**3 - Triangular Payment**	**4 - Triangular Receipt**	**5 - Purchase of Credits**	**8 - Chargebacks**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**5 - Purchase of Credits (Excluded)**	**Total Net Equity, Including Transfers**	***Net Equity Variance From Transfers***
55	13443573	Carmen Luiza Manfredini Nascime	nicatera	carmenluizam@gmail.com	617	No	**$ (168,199)**			$ (176,759)	$ 8,561	$ -	$ -	$ (194,598)	$ 28,310	**$ -**	**$ (334,487)**	*$ (166,288)*
56	1568618	JoÃ£o Batista Lima	joaobatistalima	joaolima@sidermetal.ind.br	693	No	**$ (167,527)**			$ (313,455)	$ 145,928	$ -	$ -	$ (184,654)	$ 329,116	**$ -**	**$ (23,065)**	*$ 144,462*
57	12672343	Esther Placido Abreu	3com1	telexruddy@gmail.com	131	No	**$ (166,362)**	$ (9,975)		$ (163,528)	$ 7,141	$ -	$ -	$ (6,676)	$ 200,567	**$ -**	**$ 27,529**	*$ 193,891*
58	186465	Jackson Ranielly Gomes de Freitas	jacson	waubiamichelly@hotmail.com	245	***Yes***	**$ (166,334)**	$ (21,375)		$ (222,437)	$ 77,478	$ -	$ -	$ (89,293)	$ 88,680	**$ -**	**$ (166,947)**	*$ (613)*
59	12408411	Maria Isabel Calaça	pfuturo1	pfuturocontas@gmail.com	225	No	**$ (163,123)**	$ (37,050)		$ (189,229)	$ 63,156	$ -	$ -	$ (11,960)	$ 111,753	**$ -**	**$ (63,330)**	*$ 99,793*
60	11965107	Chantal Marie Antoinette Campo	mavie	csjpc2013@gmail.com	218	***Yes***	**$ (162,908)**			$ (291,399)	$ 128,491	$ -	$ -	$ (159,126)	$ 183,924	**$ -**	**$ (138,110)**	*$ 24,798*
61	130915	EUNICE GIUSTI THIBES DE CA	eunice	eunilsegiusti@yahoo.com.br	189	No	**$ (162,694)**			$ (247,773)	$ 85,079	$ -	$ -	$ (75,839)	$ 48,293	**$ -**	**$ (190,240)**	*$ (27,546)*
62	6331948	I KONSULTANTS CORP	ribatto20	miriamo123@yahoo.com	165	No	**$ (161,567)**	$ (14,250)		$ (181,245)	$ 33,928	$ -	$ -	$ (102,164)	$ 233,480	**$ -**	**$ (30,252)**	*$ 131,316*
63	3620372	neuza pires	neuzap1	neuza1088@yahoo.com	218	No	**$ (160,396)**			$ (238,841)	$ 78,445	$ -	$ -	$ (244,157)	$ 377,879	**$ -**	**$ (26,674)**	*$ 133,721*
64	3780878	JHONATAN JAVIER ARIAS LOZ	registroenperu1	ingenierojarias@gmail.com	250	***Yes***	**$ (159,218)**	$ (5,750)		$ (177,263)	$ 23,795	$ -	$ -	$ (11,155)	$ 182,404	**$ -**	**$ 12,031**	*$ 171,249*
65	11823829	Mai I I Darwish	ethweib	zthweib@gmail.com	110	No	**$ (156,750)**			$ (156,750)		$ -	$ -	$ -	$ 267,172	**$ -**	**$ 110,422**	*$ 267,172*
66	13036573	Margaret Kumukunda	psalm121	kumukundam@gmail.com	147	No	**$ (149,725)**	$ (126,825)		$ (27,793)	$ 4,893	$ -	$ -	$ (113,988)	$ 230,472	**$ -**	**$ (33,241)**	*$ 116,484*
67	16970533	Mary Elda Cossio Camacho	tres50	pjulio601@gmail.com	146	No	**$ (149,150)**			$ (149,590)	$ 439	$ -	$ -	$ (37,204)	$ 29,924	**$ -**	**$ (156,431)**	*$ (7,280)*
68	21236029	erjia gu	gej3699	shiying369369@163.com	2,581	No	**$ (147,777)**	$ (25,277)		$ (171,015)	$ 48,515	$ -	$ -	$ (15,622)	$ -	**$ -**	**$ (163,399)**	*$ (15,622)*
69	2628692	willian valerio borges	willianvborges	willianvborges@terra.com.br	228	No	**$ (146,054)**		$ 300	$ (180,190)	$ 33,836	$ -	$ -	$ (55,512)	$ 116,954	**$ -**	**$ (84,612)**	*$ 61,442*
70	16914909	talita massa	redemassa001	amarildoorestes@gmail.com	129	No	**$ (145,727)**			$ (190,332)	$ 44,605	$ -	$ -	$ (26,269)	$ 81,321	**$ -**	**$ (90,674)**	*$ 55,053*
71	15807311	Evelyn Vasco	barnes1	luise244@msn.com	162	No	**$ (144,636)**			$ (146,482)	$ 1,846	$ -	$ -	$ (27,010)	$ 114,423	**$ -**	**$ (57,222)**	*$ 87,414*
72	16074139	Silverio Quintana	w3t4lp01	ultoh51bx@yahoo.com.pe	2,480	***Yes***	**$ (144,364)**			$ (193,350)	$ 48,985	$ -	$ -	$ -	$ 8,877	**$ -**	**$ (135,487)**	*$ 8,877*
73	20705069	nuoshen mo	qdwy016001	13722192484@qq.com	2,893	No	**$ (144,361)**			$ (144,361)		$ -	$ -	$ -	$ -	**$ -**	**$ (144,361)**	*$ -*
74	16659731	Marli Santos	sasasasa0	carlos2046@msn.com	104	No	**$ (143,925)**			$ (144,075)	$ 150	$ -	$ -	$ (889)	$ -	**$ -**	**$ (144,814)**	*$ (889)*
75	11150755	Erasmo Barroso	eraso	erasmotelex@yahoo.com	776	***Yes***	**$ (143,819)**		$ 35,575	$ (336,571)	$ 187,102	$ (29,925)	$ -	$ (351,536)	$ 383,170	**$ -**	**$ (112,185)**	*$ 31,633*
76	13562959	namir v dias	somira	somiradias@msn.com	314	***Yes***	**$ (143,621)**		$ 18,600	$ (186,764)	$ 24,543	$ -	$ -	$ (143,304)	$ 194,887	**$ -**	**$ (92,039)**	*$ 51,582*
77	16471435	Victor Blanco	vbmarte01	victorblanco1@gmail.com	255	***Yes***	**$ (142,377)**		$ 350	$ (222,319)	$ 79,592	$ -	$ -	$ (164,133)	$ 120,794	**$ -**	**$ (185,716)**	*$ (43,339)*
78	11070751	manuel armindo	amanuel1	manuelalmada2@gmail.com	141	No	**$ (141,971)**			$ (161,668)	$ 19,697	$ -	$ -	$ -	$ 290,627	**$ -**	**$ 148,656**	*$ 290,627*
79	15857549	Aldo Roberto Licata	aldoleo	aldolicata1@gmail.com	150	***Yes***	**$ (141,903)**	$ (142,500)		$ (23,253)	$ 23,851	$ -	$ -	$ (52,355)	$ 58,056	**$ -**	**$ (136,202)**	*$ 5,701*
80	13588995	paulo cerguio barradas santos barra	paulocerguio	ribs19.RS@gmail.com	120	No	**$ (141,275)**			$ (158,774)	$ 17,499	$ -	$ -	$ (23,435)	$ 69,400	**$ -**	**$ (95,310)**	*$ 45,965*
81	12745495	xufang wu	xufang1	zhonglinzhang@yahoo.com	648	No	**$ (138,356)**			$ (143,736)	$ 5,380	$ -	$ -	$ -	$ 46,011	**$ -**	**$ (92,345)**	*$ 46,011*
82	12389887	Adais Bento DaSilva	adais	adaissilva@hotmail.com	328	No	**$ (137,330)**			$ (178,658)	$ 41,328	$ -	$ -	$ (173,336)	$ 304,507	**$ -**	**$ (6,159)**	*$ 131,171*
83	444339	america beauty services	ameriglobal1	ruddy0821@gmail.com	132	No	**$ (137,128)**	$ (32,595)		$ (115,179)	$ 77,522	$ (66,875)	$ -	$ (96,882)	$ 599,269	**$ -**	**$ 365,259**	*$ 502,387*
84	1397716	webxx limited	stb100	webxxlimited@yahoo.com	109	No	**$ (137,045)**			$ (147,474)	$ 10,429	$ -	$ -	$ -	$ 551,935	**$ -**	**$ 414,890**	*$ 551,935*
85	6639486	Ariel Ruben Cufre	arielruben	telexricardofree@gmail.com	559	No	**$ (135,747)**			$ (181,910)	$ 46,163	$ -	$ -	$ (10,831)	$ 20,795	**$ -**	**$ (125,783)**	*$ 9,964*
86	13292001	HGP LLC	hgp01	luismejia4@hotmail.com	95	***Yes***	**$ (135,325)**	$ (5,700)		$ (129,675)	$ 50	$ -	$ -	$ -	$ 87,947	**$ -**	**$ (47,378)**	*$ 87,947*
87	7141148	Mai Darwish	rlinets	zthweib@gmail.com	120	No	**$ (132,910)**	$ (14,250)	$ 300	$ (146,329)	$ 29,869	$ (2,500)	$ -	$ (85,769)	$ 232,646	**$ -**	**$ 13,967**	*$ 146,877*
88	4174224	Maria Fatima da Silva Goncalves	silvafatima	fernando-gonzalez@live.com.pt	108	No	**$ (131,205)**	$ (8,550)		$ (133,216)	$ 10,561	$ -	$ -	$ (11,646)	$ 230,144	**$ -**	**$ 87,293**	*$ 218,498*
89	13750081	Chi Hung Wong	grace7151	stephen93553637@gmail.com	122	***Yes***	**$ (129,836)**	$ (143,925)		$ (7,398)	$ 21,487	$ -	$ -	$ -	$ 1,129	**$ -**	**$ (128,707)**	*$ 1,129*
90	9375740	Luiz Carlos Ferreira	lfsucesso	ferreira2927@gmail.com	140	No	**$ (129,610)**			$ (129,610)		$ -	$ -	$ -	$ 41,361	**$ -**	**$ (88,248)**	*$ 41,361*
91	8509036	Varicon LLC	vllclogin2	variconllc@gmail.com	235	No	**$ (129,154)**	$ (22,850)		$ (120,021)	$ 13,717	$ -	$ -	$ -	$ 181,489	**$ -**	**$ 52,335**	*$ 181,489*
92	12133859	jorge alberto fuentes romaña	oslo01	fuentes0007@hotmail.com	183	No	**$ (128,999)**			$ (204,396)	$ 75,397	$ -	$ -	$ (270,072)	$ 394,657	**$ -**	**$ (4,414)**	*$ 124,585*
93	21879157	silvia fernandez sampedro	despiste	lys.cars@gmail.com	1,675	***Yes***	**$ (127,907)**			$ (127,907)		$ -	$ -	$ (694)	$ 300	**$ -**	**$ (128,301)**	*$ (394)*
94	16534753	Santiago Piris	gutosantos33	patypostnet@gmail.com	147	***Yes***	**$ (127,593)**			$ (138,061)	$ 10,468	$ -	$ -	$ (32,724)	$ 81,146	**$ -**	**$ (79,170)**	*$ 48,422*
95	3369788	Jesus Ignacia Garcia Lizana de Rios	akemi18	jgarcia-31@hotmail.com	211	No	**$ (127,511)**			$ (218,413)	$ 90,902	$ -	$ -	$ (94,576)	$ 215,948	**$ -**	**$ (6,139)**	*$ 121,372*
96	13167147	John Hur	bobby06	john.hur@hurchem.com	578	No	**$ (127,073)**		$ 400	$ (129,959)	$ 2,485	$ -	$ -	$ -	$ 39,290	**$ -**	**$ (87,783)**	*$ 39,290*
97	23575094	Luz Silva	ochoa42	estreliciasvip@outlook.pt	1,662	No	**$ (126,880)**			$ (128,127)	$ 1,248	$ -	$ -	$ (23,999)	$ -	**$ -**	**$ (150,879)**	*$ (23,999)*
98	15995953	Gertrude Pangalile Rwakatare	millionaire	jrwenza@gmail.com	703	***Yes***	**$ (126,720)**	$ (122,550)		$ (42,593)	$ 38,423	$ -	$ -	$ (40,326)	$ 38,706	**$ -**	**$ (128,340)**	*$ (1,620)*
99	5556338	Michelle Machado	dinizcel131	waldiniz@genrent.com.br	105	No	**$ (126,325)**			$ (137,433)	$ 11,109	$ -	$ -	$ (101,490)	$ 143,538	**$ -**	**$ (84,276)**	*$ 42,049*
100	14722395	luciano ziani	recebe1	xxlyp2705@gmail.com	110	No	**$ (126,197)**		$ 900	$ (129,719)	$ 2,623	$ -	$ -	$ (32,100)	$ 69,473	**$ -**	**$ (88,823)**	*$ 37,373*
Total					**83,534**		**$ (22,123,674)**	**$ (2,303,279)**	**$ 165,613**	**$ (25,750,977)**	**$ 6,792,125**	**$ (1,027,156)**	**$ -**	**$ (17,895,827)**	**$ 31,492,564**	**$ -**	**$ (8,526,937)**	***$ 13,596,737***

Source(s): Based upon queries of the Huron Data Tables and ePOC Data Tables.

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 8: Top 100 TelexFree Net Winners

								Included in Net Equity						*Excluded from Net Equity*				
Num	**Rep. Id.**	**Name**	**Login**	**Email**	**Number of Telex Accounts**	**Was Claimed?**	**Total Net Equity**	**1 - Direct Payment**	**2 - Direct Receipt**	**3 - Triangular Payment**	**4 - Triangular Receipt**	**5 - Purchase of Credits**	**8 - Chargebacks**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**5 - Purchase of Credits (Excluded)**	**Total Net Equity, Including Transfers**	***Net Equity Variance From Transfers***
1	132785	Maria Teresa Milagres Neves	ttmilagres	ttmilagres@gmail.com	431	No	**$ 4,096,688.59**	$ (9,705.00)		$ (660,396.53)	$ 4,956,440.12	$ (189,650.00)	$ -	$ (7,287,435.94)	$ 6,617,041.09	**$ -**	$ 3,426,293.74	*$ (670,394.85)*
2	2522900	Benjamin Argueta	mayasol	benjamin_gauchao@yahoo.com	2010	No	**$ 4,016,152.70**	$ (21,106.00)	$ 4,540.00	$ (886,980.20)	$ 5,322,248.90	$ (402,550.00)	$ -	$ (7,128,606.82)	$ 3,921,948.74	**$ -**	$ 809,494.62	*$ (3,206,658.08)*
3	323967	Marcos Vinicios Goncalves de Lana	marcoslana	marcosvinicius7782@hotmail.cor	588	No	**$ 3,159,864.80**			$ (753,917.90)	$ 3,938,007.70	$ (24,225.00)	$ -	$ (5,155,262.67)	$ 3,260,977.83	**$ -**	$ 1,265,579.96	*$ (1,894,284.84)*
4	1944460	Alexandro Rocha	telexboston	sandrorio@comcast.net	763	No	**$ 3,051,870.61**	$ (2,731.00)	$ 158,429.81	$ (630,746.10)	$ 3,552,337.90	$ (25,420.00)	$ -	$ (5,261,552.05)	$ 3,091,807.35	**$ -**	$ 882,125.91	*$ (2,169,744.70)*
5	100889	Jose Neto	josearantes	josearantesbr@gmail.com	151	No	**$ 2,728,731.10**	$ (10,591.80)	$ 47,700.00	$ (245,173.24)	$ 3,434,292.14	$ (497,496.00)	$ -	$ (6,875,062.64)	$ 6,638,626.15	**$ 65,995.58**	$ 2,492,294.61	*$ (236,436.49)*
6	178585	LUIZ NUNES BARROS	luiznb	l.honddin@hotmail.com	1063	No	**$ 2,707,829.10**	$ (1,400.00)	$ 50,500.00	$ (478,221.50)	$ 3,154,000.60	$ (17,050.00)	$ -	$ (3,815,802.48)	$ 1,633,971.82	**$ -**	$ 525,998.44	*$ (2,181,830.66)*
7	190916	Paola Zollo Alecci	paolazollo	paolapaola433@hotmail.com	39	No	**$ 2,530,423.70**	$ (348.90)	$ 61,350.00	$ (12,144.10)	$ 2,481,566.70	$ -	$ -	$ (1,736,344.88)	$ 680,677.80	**$ -**	$ 1,474,756.62	*$ (1,055,667.08)*
8	147979	Eduardo da Silva	mutum	eduardomutum@gmail.com	542	No	**$ 2,405,821.37**	$ (15,174.90)	$ 40,790.00	$ (253,190.67)	$ 2,815,123.67	$ (181,726.73)	$ -	$ (3,432,871.94)	$ 3,102,336.14	**$ 27,379.90**	$ 2,075,285.57	*$ (330,535.80)*
9	1220280	bruno graziani	bgraziani	graziani70@hotmail.com	1057	No	**$ 2,400,427.49**		$ 88,669.09	$ (707,109.00)	$ 3,190,627.40	$ (171,760.00)	$ -	$ (3,575,862.72)	$ 1,829,642.92	**$ -**	$ 654,207.69	*$ (1,746,219.80)*
10	218612	Maria Ricardina Sousa Dos Santos Delgado	ricardina	ricatelex@gmail.com	168	No	**$ 2,294,213.80**	$ (6,656.00)	$ 27,300.00	$ (200,505.73)	$ 2,475,500.53	$ (1,425.00)	$ -	$ (3,765,870.20)	$ 3,490,495.65	**$ -**	$ 2,018,839.25	*$ (275,374.55)*
11	407579	SÃ³nia Maria Nunes Viveiros	soniaviv	soniaviv38@gmail.com	271	No	**$ 2,255,981.41**	$ (4,125.00)	$ 1,324.91	$ (52,792.80)	$ 2,311,574.30	$ -	$ -	$ (2,218,853.93)	$ 958,685.07	**$ -**	$ 995,812.55	*$ (1,260,168.86)*
12	10526589	julio c paz	jcpaz	pjulio601@gmail.com	3	No	**$ 2,241,668.60**			$ (4,275.00)	$ 2,268,765.60	$ (22,822.00)	$ -	$ (2,903,419.62)	$ 1,439,167.01	**$ -**	$ 777,415.99	*$ (1,464,252.61)*
13	600489	Michel Cristiano santolin de Arruda	michelarruda	michelcristiano22@hotmail.com	134	No	**$ 2,106,522.59**		$ 19,109.89	$ (138,429.30)	$ 2,225,842.00	$ -	$ -	$ (2,612,277.95)	$ 692,270.96	**$ -**	$ 186,515.60	*$ (1,920,006.99)*
14	1891490	julian nunez	jnunez1	carolinapc95@hotmail.com	461	***Yes***	**$ 2,085,323.85**	$ (184,314.83)	$ 40,000.00	$ (71,429.10)	$ 2,301,067.78	$ -	$ -	$ (2,448,520.32)	$ 1,382,749.80	**$ -**	$ 1,019,553.33	*$ (1,065,770.52)*
15	1061715	francisca lindiharla martins marques	lindiharla01	lindiharla@hotmail.com	670	***Yes***	**$ 2,034,063.01**			$ (598,558.40)	$ 2,926,367.20	$ (293,745.79)	$ -	$ (3,788,693.98)	$ 2,192,564.93	**$ -**	$ 437,933.96	*$ (1,596,129.05)*
16	10591645	Alexander N. Aurio	alexxsheree	eworld_homeoffice@yahoo.com	102	No	**$ 2,029,732.10**	$ (4,664.50)	$ 400.00	$ (25,947.90)	$ 2,051,662.00	$ -	$ 8,282.50	$ (2,004,332.70)	$ 180,937.52	**$ -**	$ 206,336.92	*$ (1,823,395.18)*
17	948619	Fernando da Silva Gonzalez	fernando456	fernando-gonzalez@live.com.pt	431	No	**$ 1,906,250.75**		$ 28,900.00	$ (124,784.45)	$ 2,002,135.20	$ -	$ -	$ (2,287,838.64)	$ 613,615.67	**$ -**	$ 232,027.78	*$ (1,674,222.97)*
18	1735008	Ana Paula Melim Vital	pvital	pvital@sapo.pt	11	No	**$ 1,827,560.30**		$ 10,800.00	$ (28,685.39)	$ 1,845,445.69	$ -	$ -	$ (1,846,694.17)	$ 293,623.45	**$ -**	$ 274,489.58	*$ (1,553,070.72)*
19	3301426	Melba Estevez	melba1	melbasantana@gmail.com	1	No	**$ 1,812,165.60**	$ (339.00)		$ (289.00)	$ 1,812,793.60	$ -	$ -	$ (1,273,206.02)	$ 133,339.53	**$ -**	$ 672,299.11	*$ (1,139,866.49)*
20	593987	amilcar lopez	lopesa	lopesasvs@gmail.com	760	No	**$ 1,762,485.81**	$ (4,274.90)		$ (621,237.52)	$ 2,553,373.23	$ (165,375.00)	$ -	$ (2,257,912.74)	$ 858,178.73	**$ 29,894.71**	$ 362,751.80	*$ (1,399,734.01)*
21	7459370	Sofia Jesus NUGENT OLSEN	suscripcion	amoroso100pre@hotmail.com	2	No	**$ 1,750,832.90**			$ (388.90)	$ 1,751,221.80	$ -	$ -	$ (1,857,445.72)	$ 384,754.43	**$ -**	$ 278,141.61	*$ (1,472,691.29)*
22	131817	Danilo SaÃºde Cordeiro de Moura	dcsaude	cyclonepublicidade@yahoo.com.l	86	No	**$ 1,733,312.40**			$ (50,068.80)	$ 1,783,381.20	$ -	$ -	$ (7,547,285.73)	$ 5,842,828.82	**$ -**	$ 28,855.49	*$ (1,704,456.91)*
23	24501788	Renato Sacramento	turbina	albertoestaine@hotmail.com	5	No	**$ 1,731,637.40**			$ (828.90)	$ 1,732,466.30	$ -	$ -	$ -	$ -	**$ -**	$ 1,731,637.40	*$ -*
24	10926119	Renato Alves	recalves	rcanet2@gmail.com	1	No	**$ 1,704,263.60**	$ (339.00)	$ 8,500.00		$ 1,696,491.60	$ (389.00)	$ -	$ (26,140.88)	$ 15,238,326.12	**$ 16,074,261.60**	$ 16,916,448.84	*$ 15,212,185.24*
25	10461263	yuxian liu	yxian8881	xinling1197@126.com	42	***Yes***	**$ 1,692,764.40**	$ (4,006.00)		$ (20,554.90)	$ 1,717,325.30	$ -	$ -	$ (1,781,571.19)	$ 501,300.01	**$ -**	$ 412,493.22	*$ (1,280,271.18)*
26	10539115	euzebio sudre neto	mariasudre5	mariasudre@yahoo.com	144	No	**$ 1,677,542.30**		$ 2,000.00	$ (192,698.90)	$ 1,987,141.20	$ (118,900.00)	$ -	$ (4,627,521.36)	$ 3,422,430.33	**$ -**	$ 472,451.27	*$ (1,205,091.03)*
27	8166194	FRANKLIN BEZERRA DA FONSECA	fbf000	motoqueiro.2007@globo.com	247	No	**$ 1,648,513.14**			$ (160,880.90)	$ 1,809,513.80	$ (119.76)	$ -	$ (2,122,481.42)	$ 1,346,527.88	**$ -**	$ 872,559.60	*$ (775,953.54)*
28	389305	Sandro Paulo Freitas	sandrobond	sandrofisioclinic@gmail.com	24	No	**$ 1,647,862.95**	$ (1,375.00)	$ 49,251.05	$ (13,445.90)	$ 1,613,432.80	$ -	$ -	$ (1,838,992.29)	$ 523,337.26	**$ -**	$ 332,207.92	*$ (1,315,655.03)*
29	5325978	TARCISIO FIABANE	tarcyy	tarcy@webmyway5.net	451	No	**$ 1,617,668.30**			$ (81,167.00)	$ 1,698,835.30	$ -	$ -	$ (1,829,007.10)	$ 461,105.47	**$ -**	$ 249,766.67	*$ (1,367,901.63)*
30	1619724	JULIO SILVA	jcboston	JCBOSTON01@GMAIL.COM	115	No	**$ 1,579,488.48**	$ (99.80)	$ 31,182.28	$ (51,155.00)	$ 1,723,575.00	$ (124,014.00)	$ -	$ (6,295,245.98)	$ 5,326,906.19	**$ -**	$ 611,148.69	*$ (968,339.79)*
31	1554886	Pedro Taveras	moisestv1	ptc59@hotmail.com	7	No	**$ 1,559,131.57**	$ (216,809.63)	$ 11,819.00	$ (6,049.90)	$ 1,770,172.10	$ -	$ -	$ (780,816.39)	$ 631,497.36	**$ -**	$ 1,409,812.54	*$ (149,319.03)*
32	4032900	Davidson Teixeira	postecliente3	davidsonteixeira@hotmail.com	80	No	**$ 1,544,903.50**	$ (1,425.00)	$ 9,092.59	$ (55,963.10)	$ 1,628,906.60	$ (35,707.59)	$ -	$ (2,990,175.10)	$ 1,757,869.68	**$ -**	$ 312,598.08	*$ (1,232,305.42)*
33	10935163	ana costa	equipebhusa	michelle-msc@ig.com.br	1	No	**$ 1,539,350.40**	$ (339.00)			$ 1,576,189.40	$ (36,500.00)	$ -	$ (304,081.47)	$ 62,878.26	**$ 1,148,392.00**	$ 1,298,147.19	*$ (241,203.21)*
34	106223	FABIO FERNANDES	telexfreegbi	suportetitanium@gmail.com	407	***Yes***	**$ 1,524,467.73**		$ 17,793.73	$ (44,657.90)	$ 1,551,331.90	$ -	$ -	$ (2,326,823.85)	$ 1,202,194.74	**$ 14,936.98**	$ 399,838.62	*$ (1,124,629.11)*
35	1385480	Jose Carlos Maciel	jcarlosm	jcmkgb@hotmail.com	1100	No	**$ 1,508,395.00**		$ 12,000.00	$ (967,714.00)	$ 2,484,509.00	$ (20,400.00)	$ -	$ (3,208,431.82)	$ 1,539,575.80	**$ -**	$ (160,461.02)	*$ (1,668,856.02)*
36	182532	MARIA DO SOCORRO MELO DOS SANTOS	socorromelo	msmsantostelexfree@hotmail.con	3574	***Yes***	**$ 1,498,259.83**	$ (203,568.00)	$ 186,583.05	$ (1,050,719.70)	$ 2,807,856.40	$ (241,891.92)	$ -	$ (3,117,346.34)	$ 1,863,569.18	**$ -**	$ 244,482.67	*$ (1,253,777.16)*
37	4179258	yuncheng wu	mdou888	mdwu158@163.com	73	No	**$ 1,488,075.30**	$ (4,424.70)	$ 9,597.00	$ (7,989.70)	$ 1,490,892.70	$ -	$ -	$ (1,868,726.16)	$ 542,143.66	**$ -**	$ 161,492.80	*$ (1,326,582.50)*
38	215595	jancarlo oliveira borborema	liagoalmeida	ciacristajba@gmail.com	12	No	**$ 1,478,557.22**			$ (2,493.60)	$ 1,481,330.50	$ (279.68)	$ -	$ (1,584,973.53)	$ 682,012.23	**$ -**	$ 575,595.92	*$ (902,961.30)*
39	7355804	Fredy Ernesto Prialé Palomino	rpriale	iglesiafiladelfia_cuerpodecristo@	68	No	**$ 1,458,261.20**		$ 300.00	$ (44,291.10)	$ 1,547,252.30	$ (45,000.00)	$ -	$ (1,982,650.74)	$ 696,895.67	**$ -**	$ 172,506.13	*$ (1,285,755.07)*
40	3681066	joel de jesus javiel	joeldejesus	lic.joel_dejesus@hotmail.com	52	***Yes***	**$ 1,446,111.90**	$ (50.00)	$ 41,665.00	$ (13,001.50)	$ 1,417,498.40	$ -	$ -	$ (1,247,138.17)	$ 328,860.49	**$ -**	$ 527,834.22	*$ (918,277.68)*
41	10113909	yanmin liu	lym6789	telexfree3396815@gmail.com	1	No	**$ 1,425,656.70**	$ (1,425.00)	$ 300.00	$ (49.90)	$ 1,426,831.60	$ -	$ -	$ (1,517,350.11)	$ 896,561.69	**$ -**	$ 804,868.28	*$ (620,788.42)*
42	10418533	Jesus Osuna	allintowin	jivanosuna@gmail.com	27	No	**$ 1,398,169.95**	$ (388.90)		$ (21,874.70)	$ 1,420,433.55	$ -	$ -	$ (2,614,703.03)	$ 1,602,789.98	**$ -**	$ 386,256.90	*$ (1,011,913.05)*
43	10844649	chai hock ng	ng823	panbingjian@yahoo.com	50	No	**$ 1,381,439.70**	$ (2,850.00)		$ (45,998.90)	$ 1,430,288.60	$ -	$ -	$ (2,228,455.23)	$ 1,138,084.44	**$ -**	$ 291,068.91	*$ (1,090,370.79)*
44	965375	Ana R Ramos	orionteam	classicaorboston@yahoo.com	339	No	**$ 1,363,501.09**	$ (8,550.00)	$ 25,799.39	$ (98,454.20)	$ 1,584,355.90	$ (139,650.00)	$ -	$ (2,035,613.49)	$ 1,197,697.32	**$ -**	$ 525,584.92	*$ (837,916.17)*
45	1220088	Carla Freitas	reisandrade	csfreitas@gmail.com	197	No	**$ 1,349,121.34**	$ (5,550.00)	$ 10,132.34	$ (81,943.20)	$ 1,426,482.20	$ -	$ -	$ (2,021,223.35)	$ 824,670.34	**$ -**	$ 152,568.33	*$ (1,196,553.01)*
46	144642	Hugo Alvarado	hugo2	hugo_alvarado@live.com	664	No	**$ 1,340,480.70**	$ (15,175.00)	$ 10,800.00	$ (510,857.56)	$ 1,882,088.26	$ (26,375.00)	$ -	$ (3,624,096.56)	$ 3,022,346.39	**$ -**	$ 738,730.53	*$ (601,750.17)*
47	10853999	gabriela hijar medina	romanhijar	acrelux4990@yahoo.es	8	No	**$ 1,308,377.10**			$ (11,749.30)	$ 1,320,126.40	$ -	$ -	$ (1,581,410.12)	$ 176,389.41	**$ -**	$ (96,643.61)	*$ (1,405,020.71)*
48	1922040	Edilene Storck navarro	storcknavarro	edstorck@hotmail.com	181	No	**$ 1,299,826.40**		$ 58,500.00	$ (227,469.30)	$ 1,468,795.70	$ -	$ -	$ (2,782,977.22)	$ 1,738,694.17	**$ -**	$ 255,543.35	*$ (1,044,283.05)*
49	175156	Ligia Maria de Freitas Andrade	ligiamfa	ligia.andrade@madeiratecnopolo.	96	No	**$ 1,277,870.94**	$ (2,830.00)	$ 21,150.00	$ (50,954.64)	$ 1,310,505.58	$ -	$ -	$ (1,105,685.77)	$ 310,766.40	**$ -**	$ 482,951.57	*$ (794,919.37)*
50	1252670	Marcelino Salazar Bacilio	msalazar	marcelino.sb@outlook.com	82	No	**$ 1,275,403.96**		$ 363,140.42	$ (345,919.37)	$ 1,286,682.91	$ (28,500.00)	$ -	$ (556,834.05)	$ 1,222,107.06	**$ -**	$ 1,940,676.97	*$ 665,273.01*
51	522685	gelalin3377	gt3	gelalin3377llc@gmail.com	1	No	**$ 1,268,113.51**		$ 1,005.00	$ (26,299.76)	$ 1,293,458.27	$ (50.00)	$ -	$ (3,323,989.33)	$ 2,161,483.93	**$ 49,950.00**	$ 105,608.11	*$ (1,162,505.40)*
52	5043050	Bayron Emilio Garcia Amaya	bega1971	moshe107@hotmail.com	3492	No	**$ 1,264,023.41**	$ (8,142.00)	$ 1,504.82	$ (590,174.30)	$ 1,861,510.80	$ (675.91)	$ -	$ (1,713,222.97)	$ 1,210,281.62	**$ -**	$ 761,082.06	*$ (502,941.35)*
53	620265	Marco AurÃ©lio Martins de Almeida Filho	marcoalmeida	marcoalmeidausa@yahoo.com	190	No	**$ 1,240,376.69**	$ (15,386.50)	$ 62,637.89	$ (76,193.90)	$ 1,448,594.20	$ (179,275.00)	$ -	$ (3,754,287.03)	$ 4,408,119.41	**$ -**	$ 1,894,209.07	*$ 653,832.38*
54	149697	Helio Maltaro Barbosa	galodoido	hbmaltaro@gmail.com	831	No	**$ 1,228,447.40**	$ (2,889.90)		$ (884,161.30)	$ 2,115,498.60	$ -	$ -	$(18,590,214.98)	$ 19,553,754.00	**$ -**	$ 2,191,986.42	*$ 963,539.02*
55	8791020	j f	mytree	linyu209@yahoo.com	1	No	**$ 1,227,124.60**			$ (1,425.00)	$ 1,228,549.60	$ -	$ -	$ (2,094,281.29)	$ 1,033,271.39	**$ -**	$ 166,114.70	*$ (1,061,009.90)*
56	279835	ruddy abreu	ruddy	ruddy0821@gmail.com	118	No	**$ 1,226,943.88**		$ 105,470.00	$ (124,981.41)	$ 1,534,850.29	$ (288,395.00)	$ -	$ (1,992,617.82)	$ 1,079,347.40	**$ -**	$ 313,673.46	*$ (913,270.42)*

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 8: Top 100 TelexFree Net Winners

								Included in Net Equity						*Excluded from Net Equity*				
Num	**Rep. Id.**	**Name**	**Login**	**Email**	**Number of Telex Accounts**	**Was Claimed?**	**Total Net Equity**	**1 - Direct Payment**	**2 - Direct Receipt**	**3 - Triangular Payment**	**4 - Triangular Receipt**	**5 - Purchase of Credits**	**8 - Chargebacks**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**5 - Purchase of Credits (Excluded)**	**Total Net Equity, Including Transfers**	***Net Equity Variance From Transfers***
57	707775	Linda Suzanne Hackett	dl1	davidbeeba@gmail.com	1224	No	**$ 1,216,914.18**	$ (92,971.64)	$ 1,178,646.87	$ (497,546.90)	$ 884,647.50	$ (255,861.65)	$ -	$ (992,920.54)	$ 1,045,022.59	$ -	$ 1,269,016.23	*$ 52,102.05*
58	10492319	Soraya Ferreira	ajfer	lferreiratf@gmail.com	23	No	**$ 1,198,069.50**		$ 1,000.00	$ (6,530.70)	$ 1,203,600.20	$ -	$ -	$ (1,937,729.38)	$ 1,459,826.39	$ -	$ 720,166.51	*$ (477,902.99)*
59	5626576	silvestre duranona martinez	silverduranona	silvestreduranona81@gmail.com	128	No	**$ 1,171,096.90**			$ (26,508.80)	$ 1,197,605.70	$ -	$ -	$ (1,403,216.24)	$ 304,654.78	$ -	$ 72,535.44	*$ (1,098,561.46)*
60	1826886	Tommy Lin	talkfreedom	telexfree579@gmail.com	37	No	**$ 1,163,543.48**	$ (25,649.90)		$ (16,207.40)	$ 1,205,400.78	$ -	$ -	$ (7,882,759.50)	$ 6,592,369.04	$ -	$ (126,846.98)	*$ (1,290,390.46)*
61	5843812	Abraham PÃ©rez DomÃ-ngez	onubagades	onubagadesnetwork@gmail.com	23	No	**$ 1,151,054.72**	$ (49.90)	$ 81,976.32	$ (3,491.10)	$ 1,072,619.40	$ -	$ -	$ (346,230.47)	$ 344,725.34	$ -	$ 1,149,549.59	*$ (1,505.13)*
62	156921	Edson F Souza	edsonsouza1	edsoujr@hotmail.com	132	No	**$ 1,142,732.18**	$ (15,224.80)	$ 30,806.25	$ (82,499.23)	$ 1,221,649.96	$ (12,000.00)	$ -	$ (1,863,976.57)	$ 1,002,690.84	$ -	$ 281,446.45	*$ (861,285.73)*
63	3267782	Vaming Services	vamingservices	vamingx@gmail.com	339	No	**$ 1,139,569.73**	$ (44,813.90)	$ 8,641.73	$ (45,580.50)	$ 1,221,322.40	$ -	$ -	$ (676,761.07)	$ 168,799.67	$ -	$ 631,608.33	*$ (507,961.40)*
64	868675	Jorge Antonio Mejia Sequeira	jmejia1	jmejiasequeira@gmail.com	2174	No	**$ 1,131,656.61**	$ (339.00)	$ 1,409.14	$ (330,515.90)	$ 1,895,952.37	$ (434,850.00)	$ -	$ (1,566,939.39)	$ 884,500.19	$ -	$ 449,217.41	*$ (682,439.20)*
65	10419383	yong wu	wuyong88	wuyong679@163.com	14	No	**$ 1,130,924.30**	$ (1,425.00)	$ 313.00	$ (17,002.90)	$ 1,149,039.20	$ -	$ -	$ (1,063,099.86)	$ 361,744.71	$ -	$ 429,569.15	*$ (701,355.15)*
66	254771	gustavo raniele fontes silva	mania2012	raniele21@gmail.com	144	No	**$ 1,124,882.50**			$ (57,441.20)	$ 1,182,323.70	$ -	$ -	$ (1,206,906.33)	$ 228,431.20	$ -	$ 146,407.37	*$ (978,475.13)*
67	3431510	Smart Magara Protus	cranes	sutorps1@gmail.com	33	No	**$ 1,114,648.60**	$ (2,681.00)		$ (5,577.20)	$ 1,301,937.80	$ (179,031.00)	$ -	$ (483,972.85)	$ 300,575.18	$ -	$ 931,250.93	*$ (183,397.67)*
68	947289	jonathan vargas	bonhavo1	bonhavo@hotmail.com	72	***Yes***	**$ 1,103,227.74**	$ (99,553.16)	$ 10,000.00	$ (43,275.80)	$ 1,236,056.70	$ -	$ -	$ (1,518,332.90)	$ 925,113.90	$ -	$ 510,008.74	*$ (593,219.00)*
69	1007487	Rodrigo Castro	billionaire1	rodrigocastro21@yahoo.com	41	No	**$ 1,087,835.88**	$ (99,918.50)		$ (32,747.60)	$ 1,209,906.98	$ -	$ 10,595.00	$ (1,726,397.53)	$ 1,218,509.56	$ -	$ 579,947.91	*$ (507,887.97)*
70	10384329	gildete barbosa	liderexecultivo	wagnernainglaterra@hotmail.com	9	No	**$ 1,073,748.46**	$ (388.90)		$ (2,214.70)	$ 1,082,499.90	$ (6,147.84)	$ -	$ (1,384,642.49)	$ 706,920.97	$ -	$ 396,026.94	*$ (677,721.52)*
71	945611	carlos o pina	copa22	telleriasjenny@gmail.com	1	No	**$ 1,072,143.74**	$ (114,781.76)	$ 800.00		$ 1,186,125.50	$ -	$ -	$ (935,470.56)	$ 332,494.49	$ -	$ 469,167.67	*$ (602,976.07)*
72	2858866	oscar alfredo caceres haro	oscarc	curiosodeti@hotmail.com	22	***Yes***	**$ 1,066,527.50**			$ (23,504.00)	$ 1,090,031.50	$ -	$ -	$ (1,230,673.24)	$ 288,260.65	$ -	$ 124,114.91	*$ (942,412.59)*
73	188401	Roberto Araujp	araujoroberto	robertoaraujo89@gmail.com	317	No	**$ 1,043,260.30**	$ (11,221.00)	$ 24,777.40	$ (126,030.57)	$ 1,155,734.47	$ -	$ -	$ (1,188,305.89)	$ 409,368.55	$ -	$ 264,322.96	*$ (778,937.34)*
74	2590958	david reis	teamusadavidr	teamusadavidr@gmail.com	215	No	**$ 1,037,715.50**			$ (145,694.60)	$ 1,289,082.10	$ (105,672.00)	$ -	$ (3,115,588.11)	$ 2,471,376.42	$ -	$ 393,503.81	*$ (644,211.69)*
75	713679	carlos pina	cpina1	empresaspinateixeira@gmail.com	179	No	**$ 1,020,550.83**	$ (329,059.68)	$ 128,313.69	$ (31,382.30)	$ 1,253,472.12	$ (793.00)	$ -	$ (2,085,491.95)	$ 2,516,315.59	$ -	$ 1,451,374.47	*$ 430,823.64*
76	1246474	Rodrigo Montemor	rodmontemor	rodmontemor@gmail.com	331	No	**$ 1,008,202.87**	$ (11,099.80)	$ 60,804.17	$ (260,140.99)	$ 1,290,064.49	$ (71,425.00)	$ -	$ (1,940,873.44)	$ 1,210,904.22	$ -	$ 278,233.65	*$ (729,969.22)*
77	169217	Rui Gouveia	ruigouveia	ruigouveia.mlm@gmail.com	78	No	**$ 999,024.28**	$ (319.00)	$ 29,750.00	$ (40,668.17)	$ 1,010,261.45	$ -	$ -	$ (905,463.20)	$ 186,754.16	$ -	$ 280,315.24	*$ (718,709.04)*
78	5084572	Fausto Fermin	fermint20	faustoft.20@hotmail.com	38	***Yes***	**$ 996,673.40**	$ (2,870.00)	$ 20,350.00	$ (6,223.90)	$ 985,417.30	$ -	$ -	$ (1,415,749.83)	$ 685,159.31	$ -	$ 266,082.88	*$ (730,590.52)*
79	112371	angela aparecida oliveira laskoski	angelaaf	angeloliveira20@hotmail.com	253	***Yes***	**$ 991,257.60**	$ (9,975.00)	$ 4,681.00	$ (84,620.20)	$ 1,081,171.80	$ -	$ -	$ (1,001,369.85)	$ 314,828.76	$ -	$ 304,716.51	*$ (686,541.09)*
80	2509162	DIANA LUCIA MENDEZ	dlmendez	dianaluciamendez@gmail.com	99	No	**$ 991,251.32**	$ (7,174.90)	$ 1,781.92	$ (33,749.70)	$ 1,046,069.00	$ (15,675.00)	$ -	$ (1,113,265.12)	$ 650,518.49	$ -	$ 528,504.69	*$ (462,746.63)*
81	268589	THIAGO REZENDE DANTAS	trd1	thiagorezende7@hotmail.com	248	No	**$ 991,009.29**		$ 806.00	$ (37,488.90)	$ 1,029,470.10	$ (1,777.91)	$ -	$ (1,126,171.08)	$ 355,304.62	$ -	$ 220,142.83	*$ (770,866.46)*
82	10451797	Ana Santos	gpaes	gabypaes_19@hotmail.com	2	No	**$ 985,770.00**			$ (1,077.20)	$ 986,847.20	$ -	$ -	$ (1,150,421.90)	$ 358,111.39	$ -	$ 193,459.49	*$ (792,310.51)*
83	10549549	rosemeire de sousa e silva koch	freevgbo1	fredfkoch@gmail.com	3	No	**$ 983,747.30**	$ (339.00)	$ 300.00	$ (1,764.00)	$ 985,550.30	$ -	$ -	$ (1,178,909.37)	$ 243,254.00	$ -	$ 48,091.93	*$ (935,655.37)*
84	3561088	David Livingstone Kiragga	kdl2012	dlkiragga@gmail.com	147	No	**$ 983,232.90**	$ (2,900.00)		$ (24,647.10)	$ 1,010,780.00	$ -	$ -	$ (937,340.40)	$ 233,112.35	$ -	$ 279,004.85	*$ (704,228.05)*
85	637321	ODILON BARRETO DE MACEDO MONTEIR(	clientelaila	emaildoodilon@gmail.com	82	No	**$ 970,601.70**	$ (2,202.80)		$ (49,369.90)	$ 1,023,330.40	$ (1,156.00)	$ -	$ (1,449,000.80)	$ 902,085.02	$ -	$ 423,685.92	*$ (546,915.78)*
86	10926373	Nathana Santos Reis	frentebr	pelereisneto@gmail.com	57	No	**$ 970,398.72**	$ (1,116.80)		$ (6,626.70)	$ 1,060,955.30	$ (82,813.08)	$ -	$(20,799,049.48)	$ 48,271,390.27	**$ 25,166,753.41**	$ 28,442,739.51	*$ 27,472,340.79*
87	632789	wesley gomes dias	wesleygd	wesley_comcast@yahoo.com	1331	No	**$ 968,101.20**			$ (671,068.10)	$ 1,727,034.30	$ (87,865.00)	$ -	$ (2,146,258.63)	$ 1,777,185.23	**$ 45,675.00**	$ 599,027.80	*$ (369,073.40)*
88	100605	EDUARDO FERREIRA ROSA	negocio	eduardo@grupoeduarts.com.br	73	No	**$ 966,576.20**		$ 300.00	$ (10,703.40)	$ 977,318.60	$ (339.00)	$ -	$ (854,118.96)	$ 336,214.08	$ -	$ 448,671.32	*$ (517,904.88)*
89	2281408	fabio malvestio faria	telexflorida	telexflorida@gmail.com	160	No	**$ 964,344.87**	$ (21,524.70)	$ 14,399.97	$ (24,887.40)	$ 1,004,007.00	$ (7,650.00)	$ -	$ (1,287,689.41)	$ 1,740,188.35	**$ 983,052.00**	$ 1,416,843.81	*$ 452,498.94*
90	216513	Vanderlei Ramalho dos Santos	vanderlei86	fabioramalhodossantos@gmail.co	9	No	**$ 962,042.30**			$ (7,585.90)	$ 969,628.20	$ -	$ -	$ (1,180,956.45)	$ 360,116.94	$ -	$ 141,202.79	*$ (820,839.51)*
91	339909	Rosalba Marysol	marysolparedes	rpp2205@hotmail.com	12	No	**$ 960,045.70**			$ (14,366.60)	$ 974,412.30	$ -	$ -	$ (1,058,576.24)	$ 230,669.29	$ -	$ 132,138.75	*$ (827,906.95)*
92	11284921	Paulo Camacho	befine	pauloajcamacho.telexfree@gmail.	56	No	**$ 956,548.20**			$ (5,130.20)	$ 961,678.40	$ -	$ -	$ (1,019,129.05)	$ 217,605.00	$ -	$ 155,024.15	*$ (801,524.05)*
93	10953203	youlong gu	gyl007	zhzp55@126.com	2	No	**$ 952,605.40**	$ (1,425.00)		$ (2,023.80)	$ 956,054.20	$ -	$ -	$ (1,339,888.17)	$ 555,952.20	$ -	$ 168,669.43	*$ (783,935.97)*
94	4987334	Peng Zhang	dandong	james3593@gmail.com	211	No	**$ 947,923.10**			$ (13,251.30)	$ 961,174.40	$ -	$ -	$ (1,038,550.37)	$ 132,966.90	$ -	$ 42,339.63	*$ (905,583.47)*
95	100099	Clair Berti	telexfreey	portalyatri@gmail.com	1	No	**$ 939,737.45**			$ (1,425.00)	$ 941,162.45	$ -	$ -	$ -	$ -	$ -	$ 939,737.45	*$ -*
96	10663047	Filipe Fernandes	fili100	san1000.telexfree@gmail.com	63	No	**$ 937,731.30**	$ (3,139.00)	$ 24,550.00	$ (21,593.60)	$ 937,913.90	$ -	$ -	$ (795,091.69)	$ 199,078.24	$ -	$ 341,717.85	*$ (596,013.45)*
97	2837590	FELIPE MATEO OLIVARES CAPUTO	olicompu	oli-compu@hotmail.com	136	No	**$ 936,482.90**	$ (9,755.90)	$ 34,000.00	$ (68,310.80)	$ 979,124.60	$ -	$ 1,425.00	$ (999,796.76)	$ 331,871.11	$ -	$ 268,557.25	*$ (667,925.65)*
98	944601	maryellen ayres lemos	melayres	melayres_lemos@hotmail.com	4	No	**$ 936,404.10**			$ (4,376.30)	$ 940,780.40	$ -	$ -	$ (2,492,809.44)	$ 2,184,763.45	$ -	$ 628,358.11	*$ (308,045.99)*
99	10676571	qianxi zhang	cf688	gyzyh0725@126.com	1	No	**$ 934,856.30**	$ (1,425.00)		$ (99.80)	$ 936,381.10	$ -	$ -	$ (828,809.06)	$ 527,085.00	$ -	$ 633,132.24	*$ (301,724.06)*
100	10703017	Juan Jose Tremol Campechano	dionicampeche	dionifree@gmail.com	51	No	**$ 932,773.10**	$ (14,250.00)	$ 2,500.00	$ (33,336.30)	$ 977,859.40	$ -	$ -	$ (1,108,906.07)	$ 355,217.65	$ -	$ 179,084.68	*$ (753,688.42)*
Total					**31,000**		**$ 147,515,498.32**	**$ (1,688,650.30)**	**$ 3,280,944.72**	###########	###########		**$ 20,302.50**	###########		**$ 43,606,291.18**	***$ 106,053,721.46***	**$(41,461,776.86)**

Source(s): Based upon queries of the Huron Data Tables and ePOC Data Tables.

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.1: #1 Top Net Loser - Parent ID 13043527 ($1.0M Net Loser) v. Parent ID 4032900 ($1.5M Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
13043527	$ (1,010,726.99)	1,232

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ 1,461.91	$ (1,457,939.30)	$ 445,750.40	$ -	$ (576,123.91)	$ 831,217.20	$ -	$ (1,010,726.99)
13223189	13043527	**visionusa3**	Du painting Dba	**davidsonteixeira@hotmail.com**	main st	555	Medfort		02155	US	16175016788	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
14037849	13043527	gopats	Du painting Dba	allansouza01@yahoo.com	**Vernal st**	**33**	**Everett**		**02149**	US	16175016788	ADCentral	Promoter Plan			$ (1,425.00)	$ 18,078.60		$ (16,699.00)	$ 1,498.00	$ -	$ 16,653.60
14108655	13043527	10cstfone	Du painting Dba	suportebrteam@gmail.com	main st	**33**	Medford	MA	**02149**	US	16175016788	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
14108713	13043527	11cstfone	Du painting Dba	suportebrteam@gmail.com	main st	**33**	Medford	MA	**02149**	US	16175016788	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
14674471	13043527	robadel40	du painting dba	**davidsonteixeira@hotmail.com**	main st	555	medford	MA	02155	US	6175016788	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 964.91	$ -	$ (1,325.20)
14739215	13043527	dinaousa1	du painting dba	nadiusa44@hotmail.com	hancock st	170	**everett**	MA	**02149**	US	6175016788	ADCentral	Promoter Plan			$ (1,425.00)	$ 598.80		$ -	$ 500.00	$ -	$ (826.20)
15103521	13043527	visionfone4	Du Painting Dba	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
15104757	13043527	**visionusa8**	Du Painting Dba	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (339.00)			$ -	$ -	$ -	$ (339.00)
15105763	13043527	**visionusa9**	Du Painting DBA	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (339.00)			$ -	$ -	$ -	$ (339.00)
15269611	13043527	**visionusa11**	Du Painting DBa	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (339.00)			$ -	$ -	$ -	$ (339.00)
15534787	13043527	flyamerica73	Du painting Dba	**davidsonteixeira@hotmail.com**	main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 500.00	$ -	$ (1,425.00)
16046521	13043527	kleysonusa	Du Painting DBA	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (339.00)			$ -	$ -	$ -	$ (339.00)
16321827	13043527	flyamerica74	Du painting Dba	**davidsonteixeira@hotmail.com**	main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
16321897	13043527	flyamerica75	Du painting Dba	**davidsonteixeira@hotmail.com**	main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
16438965	13043527	bidddusa	Du painting dba	tubara85@hotmail.com	main st	**33**	Medford	MA	464257144	US	16175016788	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17777595	13043527	getsugausa2	Du Painting DBA	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17778181	13043527	getsugausa3	Du Painting DBA	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
18217293	13043527	**visionusa13**	Du Painting DBA	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
18217483	13043527	**visionusa14**	Du Painting DBA	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
18219101	13043527	**visionusa15**	Du Painting DBA	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
18222429	13043527	**visionusa17**	Du Painting DBA	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
24417094	13043527	evelynusa11	Du Painting DBA	**davidsonteixeira@hotmail.com**	Main st	555	Medford	MA	02155	US	16175016788	Associate	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
33770812	13043527	technew1	Du Painting	viniwa@gmail.com	main st	555	**Everett**	MA	02149	US	**7814847474**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
4032900	$ 1,544,903.50	80

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (1,425.00)	$ 9,092.59	$ (55,963.10)	$ 1,628,906.60	$ (35,707.59)	$ (2,990,175.10)	$ 1,757,869.68	$ -	$ 1,544,903.50
4185634	4032900	liderboston	Davidson Teixeira	**davidsonteixeira @hotmail.com**	**vernal st**	**33**	**Everett**	Ma	**02149**	US	**7814847474**	ADCentral	Promoter Plan			$ (1,474.90)	$ 1,674.50		$ (6,449.00)	$ 6,189.82	$ -	$ 199.60
6994526	4032900	postepanela8	Davidson R Teixeira	**davidsonteixeira@hotmail.com**	**Vernal st**	**33**	**Everett**	MA	**02149**	US	**17814847474**	ADCentral	Promoter Plan			$ (1,425.00)	$ 5,849.70		$ (6,192.00)	$ 8,400.28	$ -	$ 4,424.70
7826764	4032900	postepanela9	Davidson R Teixeira	**davidsonteixeira@hotmail.com**	**vernal st**	**33**	**Everett**	MA	**02149**	US	**17814847474**	ADCentral	Promoter Plan			$ (1,425.00)	$ 3,538.30		$ (3,211.96)	$ 5,733.13	$ -	$ 2,113.30
8447492	4032900	postepanela11	Davidson R Teixeira	**davidsonteixeira@hotmail.com**	**Vernal st**	**33**	**Everett**	MA	**02149**	US	**17814847474**	ADCentral	Promoter Plan			$ (1,425.00)	$ 6,049.30		$ (5,523.74)	$ 5,036.42	$ -	$ 4,624.30
8551994	4032900	postecliente9	Davidson R Teixeira	**davidsonteixeira@hotmail.com**	**Vernal st**	**33**	**Everett**	MA	**02149**	US	**17814847474**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
8670864	4032900	postecliente10	Davidson R Teixeira	**davidsonteixeira@hotmail.com**	**Vernal st**	**33**	**Everett**	ma	**02149**	US	**17814847474**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
8672522	4032900	postecliente11	Davidson R Teixeira	**davidsonteixeira@hotmail.com**	**Vernal st**	**33**	**Everett**	ma	**02149**	US	**17814847474**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10485039	4032900	posteusa	Davidson Teixeira	**davidsonteixeira@hotmail.com**	bufford st	4	**Saugus**	ma	01906	US	**17814847474**	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,574.70		$ (4,097.00)	$ 5,422.00	$ -	$ 149.70
10507767	4032900	grupofallriver	Davidson Teixeira	**davidsonteixeira@hotmail.com**	bufford st	4	saugus	ma	01906	US	**17814847474**	ADCentral	Promoter Plan			$ (339.00)	$ 199.60		$ -	$ 320.00	$ -	$ (139.40)
10784211	4032900	posteusa10	Davidson R. Teixeira	**davidsonteixeira@hotmail.com**	**33 vernal st**	**33**	**everett**		**02149**	US	**7814847474**	ADCentral	Promoter Plan			$ (339.00)	$ 1,594,877.50	$ (9,975.00)	$ (2,950,198.87)	$ 1,580,056.10	$ -	$ 1,584,563.50
11753893	4032900	posteusacliente	Davidson R Teixeira	**davidsonteixeira@hotmail.com**	**33 vernal st #1**		**Everett**	MA	**02149**	US	**17814847474**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
12026623	4032900	ussa12	Davidson R Teixeira	**davidsonteixeira@hotmail.com**	**33 vernal st #1**		**Everett**	MA	**02149**	US	**17814847474**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
12224221	4032900	poste16	DAVIDSON TEIXEIR	**DAVIDSONTEIXEIRA@HOTMAIL.COM**	**33 vernal st #1**		**Everett**	MA	**02149**	US	**17814847474**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
12320649	4032900	uusa3	Davidson R Teixeira	**DAVIDSONTEIXEIRA@HOTMAIL.COM**	**33 vernal st #1**		**Everett**	MA	**02149**	US	**17814847474**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
12367405	4032900	ussa4	Davidson Teixeira	**davidsonteixeira@hotmail.com**	**Vernal St #1**	**33**	**Everett**	MA	**02149**	US	**17814847474**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
13117655	4032900	**visionusa02**	davidson teixeira	**davidsonteixeira@hotmail.com**	**vernal st**	**33**	**everett**	MA	**02149**	US	**7814847474**	Partner	Phone Plan / Other			$ (199.60)			$ -	$ -	$ -	$ (199.60)
13281021	4032900	pstusa13	Davidson Teixeira	**davidsonteixeira@hotmail.com**	**vernal st #1**	**33**	**everett**	MA	**02149**	US	**17814847474**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.2: #2 Top Net Loser - Parent ID 21036267 ($844,000 Net Loser) v. Parent ID 17401199 ($47,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
21036267	$ (843,759.10)	16,909

Example of User Accounts:

														$ -	$ -	$ (843,759.10)	$ -	$ -	$ -	$ -	$ (843,759.10)
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
25141338	21036267	1029house1	ming hua	**13958541375@13999.com**	**chuanqiao Road**	**350**	**shanghai**	31	201206	CN	8613958541375	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25140960	21036267	1029housec	ming hua	**13958541375@13999.com**	**chuanqiao Road**	**350**	**shanghai**	31	201206	CN	8613958541375	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25201528	21036267	10hgc00001	ming hua	**13958541375@13999.com**	**chuanqiao Road**	**350**	**shanghai**	31	201206	CN	8613958541375	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25201600	21036267	10hgc00001002	ming hua	**13958541375@13999.com**	**chuanqiao Road**	**350**	**shanghai**	31	201206	CN	8613958541375	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25201640	21036267	10hgc00001003	ming hua	**13958541375@13999.com**	**chuanqiao Road**	**350**	**shanghai**	31	201206	CN	8613958541375	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25201720	21036267	10hgc00001005	ming hua	**13958541375@13999.com**	**chuanqiao Road**	**350**	**shanghai**	31	201206	CN	8613958541375	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32676138	21036267	zhenjj002	ming hua	ydjt948@126.com	**chuanqiao Road**	**350**	zhengzhou	41	450000	CN	8613622849074	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32675430	21036267	zhenjj003	ming hua	ydjt948@126.com	**chuanqiao Road**	**350**	zhengzhou	41	450000	CN	8613622849074	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32675506	21036267	zhenjj004	ming hua	ydjt948@126.com	**chuanqiao Road**	**350**	zhengzhou	41	450000	CN	8613622849074	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32675582	21036267	zhenjj005	ming hua	ydjt948@126.com	**chuanqiao Road**	**350**	zhengzhou	41	450000	CN	8613622849074	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32675662	21036267	zhenjj006	ming hua	ydjt948@126.com	**chuanqiao Road**	**350**	zhengzhou	41	450000	CN	8613622849074	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32675954	21036267	zhenjj007	ming hua	ydjt948@126.com	**chuanqiao Road**	**350**	zhengzhou	41	450000	CN	8613622849074	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32677356	21036267	zhenjj008	ming hua	ydjt948@126.com	**chuanqiao Road**	**350**	zhengzhou	41	450000	CN	8613622849074	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32676380	21036267	zhenjj009	ming hua	ydjt948@126.com	**chuanqiao Road**	**350**	zhengzhou	41	450000	CN	8613622849074	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32676230	21036267	zhenjj010	ming hua	ydjt948@126.com	**chuanqiao Road**	**350**	zhengzhou	41	450000	CN	8613622849074	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32676774	21036267	zhenjj011	ming hua	ydjt948@126.com	**chuanqiao Road**	**350**	zhengzhou	41	450000	CN	8613622849074	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
31397136	21036267	zzw2014b04010	ming hua	gdllwz@gmail.com	**chuanqiao Road**	350	**shanghai**	31	201206	CN	18704601937	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
17401199	$ 46,519.40	38

Example of User Accounts:

														$ -	$ -	$ (3,970.40)	$ -	$ -	$ (34,506.74)	$ -	$ 46,519.40
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
17401199	17401199	yc5888	chuan ye	jxwnyechuan@gmail.com	jiujiang		jiujiang	32	332000	CN	18970275888	ADCentral	Promoter Plan			$ (1,524.80)			$ (34,506.74)	$ -	$ -	$ 48,695.60
32789186	17401199	yzh04001	chuan ye	**13958541375@13999.com**	**chuanqiao Road**	**350**	**shanghai**	36	332300	CN	1897027	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32789246	17401199	yzh04001002	chuan ye	**13958541375@13999.com**	**chuanqiao Road**	**350**	**shanghai**	36	332300	CN	1897027	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32789294	17401199	yzh04001003	chuan ye	**13958541375@13999.com**	**chuanqiao Road**	**350**	**shanghai**	36	332300	CN	1897027	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32789342	17401199	yzh04001004	chuan ye	**13958541375@13999.com**	**chuanqiao Road**	**350**	**shanghai**	36	332300	CN	1897027	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
32789402	17401199	yzh04001005	chuan ye	**13958541375@13999.com**	**chuanqiao Road**	**350**	**shanghai**	36	332300	CN	1897027	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.3: #3 Top Net Loser - Parent ID 15152377 ($721,000 Net Loser) v. Parent ID 4452796 ($193,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
15152377	$ (721,474.30)	665

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (743,631.30)	$ 22,157.00	$ -	$ (751,635.78)	$ 1,501,822.54	$ -	**$ (721,474.30)**
15153733	15152377	usny001	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NC	11355	US	6463215658	ADCentral	Promoter Plan			$ (1,425.00)	$ 4,242.90		$ (285,669.74)	$ 357,484.00	$ -	$ 2,817.90
15173867	15152377	usny102	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	6463215658	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15173891	15152377	usny103	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NC	11355	US	6463215658	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15174149	15152377	usny108	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	6463215658	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15174175	15152377	usny109	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NC	11355	US	6463215658	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15174245	15152377	usny110	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	6463215658	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15175079	15152377	usny127	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NC	11355	US	6463215658	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,996.00		$ (38,443.33)	$ 44,906.25	$ -	$ 571.00
15532235	15152377	usny344	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	6463215658	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15532263	15152377	usny345	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	6463215658	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15532313	15152377	usny346	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	6463215658	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
18369289	15152377	usnew004c8	fan fan	py900531@gmail.com	3	3	**flushing**	NY	11355	US	6463215658	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18369319	15152377	usnew004c9	fan fan	py900531@gmail.com	3	3	**flushing**	NY	11355	US	6463215658	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18369351	15152377	usnew004c10	fan fan	py900531@gmail.com	3	3	**flushing**	NY	11355	US	6463215658	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18494789	15152377	usnew099	fan fan	py900531@gmail.com	**5735 lawrence st.**		**flushing**	NY	11355	US	6463215658	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
18494839	15152377	usnew129	fan fan	py900531@gmail.com	**5735 lawrence st.**		**flushing**	NY	11355	US	6463215658	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
18550117	15152377	usnew126	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	6463215658	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
18550147	15152377	usnew155	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	6463215658	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
18550185	15152377	usnew156	fan fan	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	6463215658	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
23145739	15152377	usny383	fan fan	**kwan1987311@gmail.com**	**5665 136th st. 2nd floor**		**flushing**	NY	11355	US	6463215658	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
4452796	$ 192,955.00	109

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (1,425.00)	$ 306.00	$ (145,932.10)	$ 340,006.10	$ -	$ (781,955.17)	$ 709,563.59	$ -	**$ 192,955.00**
4452796	4452796	kwan311	Kwan Cheuk	**kwan1987311@gmail.com**	**5735 Lawrence ST.**	3F	**Flushing**	NY	11355	US	3479718531	ADCentral	Promoter Plan			$ (1,425.00)	$ 237,726.50		$ (466,099.31)	$ 281,550.18	$ -	$ 236,301.50
10375207	4452796	pan0701	kwan cheuk	**kwan1987311@gmail.com**	**57-35 lawrence st**		**flushing**	ny	11355	US	3013177872	ADCentral	Promoter Plan			$ (1,425.00)	$ 998.00		$ (5,208.63)	$ 9,094.37	$ -	$ (427.00)
10376819	4452796	pan0701a	kwan cheuk	**kwan1987311@gmail.com**	**57-35 lawrence st**		**flushing**	ny	11355	US	3479718531	ADCentral	Promoter Plan			$ (1,425.00)	$ 948.10		$ -	$ 3,264.87	$ -	$ (476.90)
10886765	4452796	terrysam0827a	kwan cheuk	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	ny	11355	US	13479718531	ADCentral	Promoter Plan	$ (1,425.00)			$ 2,197.00		$ -	$ 3,366.36	$ -	$ 772.00
11105815	4452796	joe0911c	kwan cheuk	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	ny	11355	US	3479718531	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,078.40		$ -	$ 2,285.23	$ -	$ (346.60)
11109569	4452796	joe0911h	kwan cheuk	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	ny	11355	US	3479718531	ADCentral	Promoter Plan			$ (1,425.00)	$ 748.50		$ -	$ 2,185.23	$ -	$ (676.50)
11109771	4452796	joe0911i	kwan cheuk	**kwan1987311@gmail.com**	**5735 lawrence st.**		**flushing**	ny	11355	US	3479718531	ADCentral	Promoter Plan			$ (1,425.00)	$ 748.50		$ -	$ 2,330.23	$ -	$ (676.50)
23139439	4452796	us0325	kwan cheuk	**kwan1987311@gmail.com**	**5665 136th st. 2nd floor**		**flushing**	NY	11355	US	3479748531	ADFamily	Promoter Plan			$ (339.60)	$ 49.90		$ (84,896.89)	$ 44,567.00	$ -	$ (289.70)
28950648	4452796	usterry009	kwan cheuk	**kwan1987311@gmail.com**	yuexiu qu beijing lu gao di NULL		guangzhou		44 510000	CN	3479718531	ADFamily	Promoter Plan			$ (339.60)			$ (5,000.00)	$ 4,579.79	$ -	$ (339.60)
28950918	4452796	usterry019	kwan cheuk	**kwan1987311@gmail.com**	yuexiu qu beijing lu gao di NULL		guangzhou		44 510000	CN	3479718531	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
28951298	4452796	usterry039	kwan cheuk	**kwan1987311@gmail.com**	yuexiu qu beijing lu gao di NULL		guangzhou		44 510000	CN	3479718531	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.4: #4 Top Net Loser - Parent ID 12650741 ($708,000 Net Loser) v. Parent ID 10992563 ($190,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
12650741	$ (708,132.90)	509

Example of User Accounts:

														1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (710,508.70)	$ 2,375.80	$ -	$ -	$ 469,662.33	$ -	**$ (708,132.90)**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type									
12652035	12650741	shalclient	Shal Group LLC	**ajaff@shalsd.com**	.	.	.	.	.	US	6199571109	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
13059489	12650741	shalclient1	shall group llc	**ajaff@shalsd.com**	.	.	.	CA	.	US	19142262551	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
13385007	12650741	shal33	shal group LLC	**ajaff@shalsd.com**	**707 Broadway**	suite	**san Diego**	CA	92101	US	**16196626049**	ADCentral	Promoter Plan			$ (1,495.00)			$ -	$ 920.00	$ -	$ (1,495.00)
13558983	12650741	shal60	Shal Group LLC	**ajaff@shalsd.com**	**Broadway**	707	**San Diego**	CA	92101	US	**16196626049**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,200.00	$ -	$ (1,425.00)
15099559	12650741	shal129	Shal Group LLC	**ajaff@shalsd.com**	**707 Broadway**		**San Diego**	CA	92101	US	**6196626049**	ADCentral	Promoter Plan			$ (1,495.00)			$ -	$ 600.00	$ -	$ (1,495.00)
15102561	12650741	shal141	Shal Group LLC	**ajaff@shalsd.com**	**707 Broadway**	**1410**	**San Diego**	CA	92101	US	**6196626049**	ADCentral	Promoter Plan			$ (1,495.00)			$ -	$ 600.00	$ -	$ (1,495.00)
15162083	12650741	shal4c	shal group llc	**ajaff@shalsd.com**	.	.	.	CA	.	US		Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
16942913	12650741	shal436	shal group llc	**claudia@dolarex.com**	**707 broadway**	**1410**	**san diego**	CA	92101	US	**6196626049**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
16948569	12650741	shal117c	shal group llc	**claudia@dolarex.com**	xxxx	xxxxx	xxxxx	CA	xxxx	US		Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17119259	12650741	shal488	shal group llc	**claudia@dolarex.com**	**707 broadway**	**1410**	**san diego**	CA	92101	US	**6196626049**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
10992563	$ 190,380.65	1

User Account:

														1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ 60,605.85	$ (1,425.00)	$ 131,199.80	$ -	$ (547,192.24)	$ 245,118.00	$ -	**$ 190,380.65**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type									
10992563	10992563	dolarex1	Dolarex LLC	**ajaff@dolarex.com**	**707 broadway**	**1410**	**San diego**	**Ca**	92101	US	**6196626049**	ADCentral	Promoter Plan		$ 60,605.85	$ (1,425.00)	$ 131,199.80		$ (547,192.24)	$ 245,118.00	$ -	$ 190,380.65

Source(s): Based upon queries of the Huron Data Tables

Exhibit 9.5: #5 Top Net Loser - Parent ID 11849449 ($622,000 Net Loser) v. Parent ID 707775 ($1.2M Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
11849449	$ (622,341.60)	928

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (481,770.00)	$ -	$ (837,034.20)	$ 696,462.60	$ -	$ (1,904,354.96)	$ 2,254,352.23	$ -	**$ (622,341.60)**
11849449	11849449	**dlld1**	DL1 INC	**davidbeeba@gmail.com**	**97 bellevue Avenue**	.	**Melrose**		02176	US	8573122571	ADCentral	Promoter Plan			$ (1,425.00)	$ 667,160.20		$ (1,160,800.60)	$ 199,425.12	$ -	$ 665,735.20
12957057	11849449	**dlld2**	Dl1 Inc	**davidbeeba@gmail.com**	**97 Bellevue Avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan			$ (1,425.00)	$ 2,675.20		$ (471,991.79)	$ 508,479.31	$ -	$ 1,250.20
12957617	11849449	**dlld4**	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue Avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,427.70		$ -	$ 16,689.93	$ -	$ 2.70
13185577	11849449	ldh18	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan	$ (1,425.00)					$ -	$ 1,424.55	$ -	$ (1,425.00)
13185629	11849449	ldh20	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan	$ (1,425.00)					$ -	$ 1,124.55	$ -	$ (1,425.00)
13282399	11849449	ldh31	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,224.55	$ -	$ (1,425.00)
13700575	11849449	lsh25	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,324.55	$ -	$ (1,425.00)
14300071	11849449	hackett21	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 899.00	$ -	$ (1,425.00)
14300229	11849449	hackett22	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 899.00	$ -	$ (1,425.00)
14367325	11849449	hackett77	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 899.00	$ -	$ (1,425.00)
14367605	11849449	hackett80	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 899.00	$ -	$ (1,425.00)
14405571	11849449	hackett104	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 899.00	$ -	$ (1,425.00)
14405721	11849449	hackett108	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 899.00	$ -	$ (1,425.00)
14405775	11849449	hackett109	Dl1 Inc	**davidbeeba@gmail.com**	**97 bellevue avenue**	.	**Melrose**	MA	02176	US	**7816200168**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 899.00	$ -	$ (1,425.00)
15800715	11849449	dlus32	Dl1 Inc	**davidbeeba@gmail.com**	**97 Bellevue Avenue**	.	**Melrose**	MA	02176	US	**7816201441**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15800767	11849449	dlus33	Dl1 Inc	**davidbeeba@gmail.com**	**97 Bellevue Avenue**	.	**Melrose**	MA	02176	US	**7816201441**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
707775	$ 1,216,914.18	1,224

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (92,971.64)	$ 1,178,646.87	$ (497,546.90)	$ 884,647.50	$ (255,861.65)	$ (992,920.54)	$ 1,045,022.59	$ -	**$ 1,216,914.18**
707775	707775	**dl1**	Linda Suzanne Hackett	**davidbeeba@gmail.com**	**97 bellevue ave**	97	**Melrose**		02176	US	**7816201441**	ADCentral	Promoter Plan	$ (86,046.74)	$ 520,680.56		$ 751,314.60	$ (255,861.65)	$ (948,029.16)	$ 106,939.93	$ -	$ 930,086.77
714683	707775	**dl610**	linda s hackett	**davidbeeba@gmail.com**	**.**	.	**.**	.	02176	US		ADCentral	Promoter Plan		$ 2,300.00	$ (1,375.00)			$ -	$ 2,700.00	$ -	$ 925.00
1809146	707775	dl32ten	Linda S. Hackett	**davidbeeba@gmail.com**	**97 Bellevue Ave.**	.	**Melrose**	Ma	02176	US	**7816201441**	ADCentral	Promoter Plan		$ 2,000.00	$ (1,375.00)	$ 499.00		$ -	$ 2,301.00	$ -	$ 1,124.00
1851312	707775	dl25five	Linda S. Hackett	**davidbeeba@gmail.com**	**97 Bellevue Ave.**	.	**Melrose**	Ma	02176	US	**7816201441**	ADCentral	Promoter Plan		$ 1,900.00	$ (1,375.00)	$ 598.80		$ -	$ 2,401.20	$ -	$ 1,123.80
1851356	707775	dl25six	Linda S. Hackett	**davidbeeba@gmail.com**	**97 Bellevue Ave.**	.	**Melrose**	Ma	02176	US	**7816201441**	ADCentral	Promoter Plan		$ 2,100.00	$ (1,375.00)	$ 598.80		$ -	$ 2,201.20	$ -	$ 1,323.80
1851404	707775	dl25seven	Linda S. Hackett	**davidbeeba@gmail.com**	**97 Bellevue Ave.**	.	**Melrose**	Ma	02176	US	**7816201441**	ADCentral	Promoter Plan		$ 2,100.00	$ (1,375.00)	$ 598.80		$ -	$ 2,301.20	$ -	$ 1,323.80
2462232	707775	dl31one5	Linda S. Hackett	**davidbeeba@gmail.com**	**97 Bellevue Ave.**	.	**Melrose**	Ma	02176	US	**7816201441**	ADCentral	Promoter Plan		$ 2,024.55	$ (1,425.00)	$ 9,281.40		$ (8,829.65)	$ 2,748.25	$ -	$ 9,880.95
2534112	707775	dl25c47	linda hackett	**davidbeeba@gmail.com**	**.**	.	**.**	.	.	US		Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
2534162	707775	dl25c48	linda hackett	**davidbeeba@gmail.com**	**.**	.	**.**	.	.	US		Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
2534214	707775	dl25c49	linda hackett	**davidbeeba@gmail.com**	**.**	.	**.**	.	.	US		Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
7332966	707775	davili0	Linda Suzanne Hackett	**davidbeeba@gmail.com**	**97 Bellevue avenue**	.	**Melrose**	Ma	02176	US	**7816200168**	ADCentral	Promoter Plan		$ 30,475.92	$ (1,425.00)	$ 1,347.30		$ (26,094.00)	$ 11,383.69	$ -	$ 30,398.22
7383722	707775	davilic2	Linda S. Hackett	**davidbeeba@gmail.com**	**97 Bellevue Avenue**	.	**Melrose**	Ma	02176	US	**7816201441**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
7539346	707775	davicli5	Linda S. Hackett	**davidbeeba@gmail.com**	**97 Bellevue Avenue**	.	**Melrose**	Ma	02176	US	**7816201441**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
7548850	707775	davicli29	Linda S. Hackett	**davidbeeba@gmail.com**	**97 Bellevue Avenue**	.	**Melrose**	Ma	02176	US	**7816201441**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
13186253	707775	dlsh3c30	linda hackett	**davidbeeba@gmail.com**	**.**	.	**melrose**	MA	.	US		Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.6: #6 Top Net Loser - Parent ID 12242105 ($481,000 Net Loser) v. Parent ID 4987334 ($948,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
12242105	$ (481,227.70)	517

Example of User Accounts:

													$ -	$ -	$ (726,956.40)	$ 245,728.70	$ -	$ (621,864.91)	$ 1,008,813.48	$ -	**$ (481,227.70)**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
16653383	12242105	**xml004**	Jun Zhang	telexfreexml@gmail.com	134-17 Cherry Ave #2C		Flushing	NY	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 598.80		$ -	$ 1,724.91	$ -	$ (826.20)
16653661	12242105	**xml005**	Jun Zhang	telexfreexml@gmail.com	134-17 Cherry Ave #2C		Flushing	NY	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 598.80		$ -	$ 1,724.91	$ -	$ (826.20)
16653705	12242105	**xml006**	Jun Zhang	telexfreexml@gmail.com	134-17 Cherry Ave #2C		Flushing	NY	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 598.80		$ -	$ 1,724.91	$ -	$ (826.20)
16654007	12242105	**xml007**	Jun Zhang	telexfreexml@gmail.com	134-17 Cherry Ave #2C		Flushing	NY	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 548.90		$ -	$ 904.91	$ -	$ (876.10)
17116793	12242105	fax001	Jun Zhang	telexfreexml@gmail.com	134-17 Cherry Ave #2C		Flushing	NY	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 698.60		$ -	$ 3,144.91	$ -	$ (726.40)
17117045	12242105	fax002	Jun Zhang	telexfreexml@gmail.com	134-17 Cherry Ave #2C		Flushing	NY	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 698.60		$ -	$ 3,144.91	$ -	$ (726.40)
17117655	12242105	fax003	Jun Zhang	telexfreexml@gmail.com	134-17 Cherry Ave #2C		Flushing	NY	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 499.00		$ -	$ 1,724.91	$ -	$ (926.00)
17117717	12242105	fax004	Jun Zhang	telexfreexml@gmail.com	134-17 Cherry Ave #2C		Flushing	NY	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 499.00		$ -	$ 1,724.91	$ -	$ (926.00)
17118481	12242105	fax005	Jun Zhang	telexfreexml@gmail.com	134-17 Cherry Ave #2C		Flushing	NY	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 499.00		$ -	$ 1,724.91	$ -	$ (926.00)
18616627	12242105	shao01	Jun Zhang	telexfree1k@gmail.com	136-45 Cherry Ave #2C		Flushing	NY	11355	US	3478228766	ADFamily	Promoter Plan			$ (339.60)			$ -	$ 1,174.00	$ -	$ (339.60)
18616819	12242105	shao02	Jun Zhang	telexfree1k@gmail.com	136-45 Cherry Ave #2C		Flushing	NY	11355	US	3478228766	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
18617009	12242105	shao03	Jun Zhang	telexfree1k@gmail.com	136-45 Cherry Ave #2C		Flushing	NY	11355	US	3478228766	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
18617273	12242105	shao04	Jun Zhang	telexfree1k@gmail.com	136-45 Cherry Ave #2C		Flushing	NY	11355	US	3478228766	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
18731909	12242105	shao11	Jun Zhang	telexfree1k@gmail.com	136-45 Cherry Ave #2C		Flushing	NY	11355	US	3478228766	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
18732293	12242105	shao12	Jun Zhang	telexfree1k@gmail.com	136-45 Cherry Ave #2C		Flushing	NY	11355	US	3478228766	ADFamily	Promoter Plan			$ (339.60)			$ -	$ 1,344.71	$ -	$ (339.60)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
4987334	$ 947,923.10	211

Example of User Accounts:

													$ -	$ -	$ (13,251.30)	$ -	$ -	$ (1,038,550.37)	$ 132,966.90	$ -	**$ 947,923.10**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
16653553	4987334	**xml002c1**	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
16653899	4987334	**xml003c1**	Peng Zhang	james3s593@gmail.com	4345 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
16654269	4987334	**xml004c1**	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
16654581	4987334	**xml005c1**	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
16654951	4987334	**xml006c1**	Peng Zhang	james359s3@gmail.com	4345 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
17210903	4987334	fax012c1	Penseg Zhang	james3s593@gmail.com	4345 Byrd St		Flushing	HI	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
17211447	4987334	fax013c1	Penseg Zhang	james3s593@gmail.com	4345 Byrd St		Flushing	HI	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
17212199	4987334	fax014c1	Penseg Zhang	james3s593@gmail.com	4345 Byrd St		Flushing	HI	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
17212925	4987334	fax015c1	Penseg Zhang	james3s593@gmail.com	4345 Byrd St		Flushing	HI	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17213595	4987334	fax016c1	Penseg Zhang	james3s593@gmail.com	4345 Byrd St		Flushing	HI	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17214209	4987334	fax017c1	Penseg Zhang	james3s593@gmail.com	4345 Byrd St		Flushing	HI	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18617971	4987334	shao01c1	Pen3g Zha3ng	james359s3@gmail.com	4545 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18618071	4987334	shao01c2	Pen3g Zha3ng	james359s3@gmail.com	4545 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18618127	4987334	shao01c3	Pen3g Zha3ng	james359s3@gmail.com	4545 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18618209	4987334	shao01c4	Pen3g Zha3ng	james359s3@gmail.com	4545 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18618287	4987334	shao01c5	Pen3g Zha3ng	james359s3@gmail.com	4545 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18618499	4987334	shao01c6	Pen3g Zha3ng	james359s3@gmail.com	4545 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18618837	4987334	shao01c10	Pen3g Zha3ng	james359s3@gmail.com	4545 Byrd St		Flushing	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.7: #7 - Parent ID 12072673 ($438,000 Net Loser) v. Parent ID 1922040 ($1.3M Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
12072673	$ (438,337.60)	378

Example of User Accounts:

														$ -	$ -	$ (463,109.50)	$ 24,771.90	$ -	$ (178,773.64)	$ 317,416.43	$ -	**$ (438,337.60)**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
13783125	12072673	telecomm1	Telecommunication group inc .	Intertelecomm@gmail.com	Case dr	62	Revere	MA	02151	US	5089715961	ADCentral	Promoter Plan			$ (1,425.00)	$ 12,138.80		$ (65,912.64)	$ 55,541.06	$ -	$ 10,713.80
13783445	12072673	telecomm2	Telecommunication group inc .	Intertelecomm@gmail.com	Case dr	62	Revere	MA	02151	US	5089715961	ADCentral	Promoter Plan			$ (1,495.00)	$ 11,285.80		$ (5,636.00)	$ 500.00	$ -	$ 9,790.80
15203077	12072673	**eboston52**	telex boston inc .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 800.00	$ -	$ (1,425.00)
16100251	12072673	bostoniano2	telecomm boston .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MA	02451	US	6172998408	Partner	Promoter Plan			$ (50.00)			$ -	$ -	$ -	$ (50.00)
16100407	12072673	bostoniano4	telecomm boston ..	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MA	02451	US	6172998408	Partner	Promoter Plan			$ (50.00)			$ -	$ -	$ -	$ (50.00)
16870169	12072673	vestorck2	telex boston inc .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 704.00	$ -	$ (1,425.00)
17029807	12072673	**eboston0100**	telex boston inc .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 400.00	$ -	$ (1,425.00)
17029955	12072673	**eboston0103**	telex boston inc .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 400.00	$ -	$ (1,425.00)
17629985	12072673	vestorck4	telex boston inc .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
17924427	12072673	**eboston160**	telex boston inc .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
17934285	12072673	**eboston175**	telex boston inc .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MI	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
17934649	12072673	**eboston176**	telex boston inc .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	w	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
18006241	12072673	sboston119	telex boston inc .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
18007845	12072673	sboston127	telex boston inc .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
18018019	12072673	**bboston76**	telex boston inc .	**edstorck@hotmail.com**	**1128 trapelp rd**	.	**waltham**	MA	02451	US	7	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
1922040	$ 1,299,826.40	181

Example of User Accounts:

														$ -	$ 58,500.00	$ (227,469.30)	$ 1,468,795.70	$ -	$ (2,782,977.22)	$ 1,738,694.17	$ -	**$ 1,299,826.40**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
1997892	1922040	storcknavarro8	Edilene Storck Navarro	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,375.00)			$ -	$ 4,800.00	$ -	$ (1,375.00)
1998306	1922040	storcknavarro9	Edilene Storck Navarro	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,375.00)			$ -	$ 5,000.00	$ -	$ (1,375.00)
1998538	1922040	storcknavarro10	Edilene Storck Navarro	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,375.00)			$ -	$ 5,000.00	$ -	$ (1,375.00)
9742155	1922040	edsto51	edilene s. navarro	**edstorck@hotmail.com**	**1128 trapelo rd**		**waltham**	Ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,824.55	$ -	$ (1,425.00)
9742279	1922040	edsto52	edilene s navarro	**edstorck@hotmail.com**	**1128 trapelo rd**		**waltham**	Ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,524.55	$ -	$ (1,425.00)
10995193	1922040	telexboston01	Edilene Storck Navarro	**edstorck@hotmail.com**	**1128 trapelo**		**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)	$ 58,427.90		$ (364,471.71)	$ 196,651.47	$ -	$ 57,002.90
10995797	1922040	telexboston02	Edilene Storck Navarro	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)	$ 92,762.20		$ (123,738.68)	$ 64,559.94	$ -	$ 91,337.20
11483849	1922040	telexboston02c	Edilene s navarro	**edstorck@hotmail.com**	e	e	r	e	s	US		Partner	Phone Plan / Other			$ (349.30)			$ -	$ -	$ -	$ (349.30)
11506343	1922040	tboston13	Edilene s navarro	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 6,164.55	$ -	$ (1,425.00)
11506705	1922040	tboston15	Edilene s navarro	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 6,164.55	$ -	$ (1,425.00)
11507133	1922040	tboston16	Edilene s navarro	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 5,854.73	$ -	$ (1,425.00)
11507193	1922040	tboston17	Edilene s navarro	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 4,854.73	$ -	$ (1,425.00)
11544787	1922040	**eboston25**	edilene s navarro	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,000.00	$ -	$ (1,425.00)
11545105	1922040	eboston28	edilene s navarro	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,000.00	$ -	$ (1,425.00)
11545135	1922040	**eboston29**	edilene s navarro	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,000.00	$ -	$ (1,425.00)
12094587	1922040	**bboston38**	edilene s navarro	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,600.00	$ -	$ (1,425.00)
12520455	1922040	iuniversla3	edilene s navarro	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	MA	02451	US	**7818913536**	Partner	Promoter Plan			$ (50.00)			$ -	$ -	$ -	$ (50.00)
12520545	1922040	iuniversal5	edilene s navarro	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,495.00)	$ 7,597.30		$ (9,206.94)	$ 3,328.00	$ -	$ 6,102.30
12520819	1922040	iuniversal3	edilene s navarro	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**		02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,495.00)	$ 99.80		$ -	$ 1,900.00	$ -	$ (1,395.20)

Source(s): Based upon queries of the Huron Data Tables; and Exhibit 9.9

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.8: #8 Top Net Loser - Parent ID 19298561 ($407,000 Net Loser) v. Parent ID 18296273 ($43,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
19298561	$ (407,383.60)	8,163

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two	4 - Triangular Receipt Two	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (407,383.60)	$ -	$ -	$ -	$ -	$ -	**$ (407,383.60)**
19360631	19298561	patricklh001	yabin zhang	**gdllwz@gmail.com**	dongdajie	1231	suzhou	44	215004	CN	13004590687	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19360663	19298561	patricklh001002	yabin zhang	**gdllwz@gmail.com**	dongdajie	1231	suzhou	44	215004	CN	13004590687	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19360681	19298561	patricklh001003	yabin zhang	**gdllwz@gmail.com**	dongdajie	1231	suzhou	44	215004	CN	13004590687	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19360699	19298561	patricklh001004	yabin zhang	**gdllwz@gmail.com**	dongdajie	1231	suzhou	44	215004	CN	13004590687	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19360723	19298561	patricklh001005	yabin zhang	**gdllwz@gmail.com**	dongdajie	1231	suzhou	44	215004	CN	13004590687	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19360731	19298561	patricklh001006	yabin zhang	**gdllwz@gmail.com**	dongdajie	1231	suzhou	44	215004	CN	13004590687	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19360747	19298561	patricklh001007	yabin zhang	**gdllwz@gmail.com**	dongdajie	1231	suzhou	44	215004	CN	13004590687	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21716813	19298561	myc169401001016	yabin zhang		dongdajie	1231	suzhou	32	215004	CN	13004590687	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
18296273	$ 43,207.50	11

User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two	4 - Triangular Receipt Two	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (858.50)	$ 44,066.00	$ -	$ (350,758.41)	$ 324,121.20	$ -	**$ 43,207.50**
18296273	18296273	liangliang168	lingliang liu	llwznet@gmail.com	dongguan	358	dongguan	44	523620	CN	8613556762602	ADFamily	Promoter Plan			$ (359.50)	$ 44,066.00		$ (350,758.41)	$ 324,121.20	$ -	$ 43,706.50
18297443	18296273	liangliang001	lingliang liu	**gdllwz@gmail.com**	dongguan	358	dongguan	44	523620	CN	8613556762602	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18297561	18296273	liangliang002	lingliang liu	**gdllwz@gmail.com**	dongguan	358	dongguan	44	523620	CN	8613556762602	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18297591	18296273	liangliang003	lingliang liu	**gdllwz@gmail.com**	dongguan	358	dongguan	44	523620	CN	8613556762602	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18297625	18296273	liangliang004	lingliang liu	**gdllwz@gmail.com**	dongguan	358	dongguan	44	523620	CN	8613556762602	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18297761	18296273	liangliang005	lingliang liu	**gdllwz@gmail.com**	dongguan	358	dongguan	44	523620	CN	8613556762602	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18298109	18296273	liangliang006	lingliang liu	**gdllwz@gmail.com**	dongguan	358	dongguan	44	523620	CN	8613556762602	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18298127	18296273	liangliang007	lingliang liu	**gdllwz@gmail.com**	dongguan	358	dongguan	12	523620	DZ	8613556762602	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18298191	18296273	liangliang008	lingliang liu	**gdllwz@gmail.com**	dongguan	358	dongguan	43	523620	CN	8613556762602	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18298291	18296273	liangliang009	lingliang liu	**gdllwz@gmail.com**	dongguan	358	dongguan	44	523620	CN	8613556762602	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18298353	18296273	liangliang0010	lingliang liu	**gdllwz@gmail.com**	dongguan	358	dongguan	11	523620	CN	8613556762602	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.9: #9 Top Net Loser - Parent ID 2876246 ($407,000 Net Loser) v. Parent ID 1922040 ($1.3M Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
2876246	$ (406,590.90)	1,142

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (424,674.50)	$ 18,083.60	$ -	$ (54,841.55)	$ 1,243,612.99	$ -	**$ (406,590.90)**
2876246	2876246	storck11	**edilene**	**edstorck@hotmail.com**	.	.	.	.	.	US		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 10,158.89	$ -	$ (1,425.00)
2877144	2876246	storck12	**edilene**	**edstorck@hotmail.com**	.	.	.	.	.	US		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 18,597.24	$ -	$ (1,425.00)
2877778	2876246	storck13	**edilene**	**edstorck@hotmail.com**	.	.	.	.	.	US		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 5,449.10	$ -	$ (1,425.00)
3575968	2876246	sto1211c1	**edilene**	**edstorck@hotmail.com**	.	.	.	.	.	US		Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
8169028	2876246	bstorck01	**edilene**	edstorck@hotmail.com	..	.	.		.	US		ADCentral	Promoter Plan			$ (1,425.00)	$ 7,174.90		$ (54,841.55)	$ 51,591.20	$ -	$ 5,749.90
8913852	2876246	**edsto24**	**Edilene**	edilenestorck@gmail.com	.	.	.	.	.	US		ADCentral	Promoter Plan			$ (1,425.00)	$ 399.20		$ -	$ 3,924.55	$ -	$ (1,025.80)
9742509	2876246	**edsto54**	**edilene**	edstorck@hotmail.com	**1128 trapelo rd**		**waltham**	Ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,824.55	$ -	$ (1,425.00)
10428335	2876246	**edsto53c5**	**edilene**	**edstorck@hotmail.com**	.	.	.	.	.	US		Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10428347	2876246	**edsto54c2**	**edilene**	**edstorck@hotmail.com**	.	.	.	.	.	US		Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10647437	2876246	edstorck1cl1	**edilene**	**edstorck@hotmail.com**	.	.	...	.	..	US		Partner	Phone Plan / Other			$ (349.30)			$ -	$ -	$ -	$ (349.30)
10850905	2876246	**edsto77**	**edilene**	**edstorck@hotmail.com**	.	.	.	.	.	US		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,700.00	$ -	$ (1,425.00)
10995645	2876246	telex01cl1	**edilene**	**edstorck@hotmail.com**	.	.	.	.	.	US		Partner	Phone Plan / Other			$ (399.20)			$ -	$ -	$ -	$ (399.20)
10996317	2876246	**telexb03cl1**	**edilene**	**edstorck@hotmail.com**	.	.	.	.	.	US		Partner	Phone Plan / Other			$ (399.20)			$ -	$ -	$ -	$ (399.20)
10999707	2876246	**telexb04cl1**	**edilene**	**edstorck@hotmail.com**	edstorck@hotmail.com	.	.	.	.	US		Partner	Phone Plan / Other			$ (399.20)			$ -	$ -	$ -	$ (399.20)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
1922040	$ 1,299,826.40	181

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ 58,500.00	$ (227,469.30)	$ 1,468,795.70	$ -	$ (2,782,977.22)	$ 1,738,694.17	$ -	**$ 1,299,826.40**
1997892	1922040	storcknavarro8	**Edilene Storck Navarro**	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,375.00)			$ -	$ 4,800.00	$ -	$ (1,375.00)
1998306	1922040	storcknavarro9	**Edilene Storck Navarro**	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,375.00)			$ -	$ 5,000.00	$ -	$ (1,375.00)
1998538	1922040	storcknavarro10	**Edilene Storck Navarro**	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,375.00)			$ -	$ 5,000.00	$ -	$ (1,375.00)
9741797	1922040	**edsto48**	edilene s. navarro	**edstorck@hotmail.com**	1128 trapelo rd		waltham	Ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,824.55	$ -	$ (1,425.00)
9741913	1922040	**edsto49**	edilene s. navarro	**edstorck@hotmail.com**	1128 trapelo rd		waltham	Ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,824.55	$ -	$ (1,425.00)
9742027	1922040	**edsto50**	edilene s. navarro	**edstorck@hotmail.com**	1128 trapelo rd		waltham	Ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,824.55	$ -	$ (1,425.00)
9742155	1922040	**edsto51**	**edilene s. navarro**	**edstorck@hotmail.com**	**1128 trapelo rd**		**waltham**	Ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,824.55	$ -	$ (1,425.00)
9742279	1922040	**edsto52**	**edilene s navarro**	**edstorck@hotmail.com**	**1128 trapelo rd**		**waltham**	Ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,524.55	$ -	$ (1,425.00)
10995193	1922040	telexboston01	**Edilene Storck Navarro**	**edstorck@hotmail.com**	**1128 trapelo**		**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)	$ 58,427.90		$ (364,471.71)	$ 196,651.47	$ -	$ 57,002.90
10995797	1922040	telexboston02	**Edilene Storck Navarro**	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)	$ 92,762.20		$ (123,738.68)	$ 64,559.94	$ -	$ 91,337.20
10996151	1922040	**telexb03**	Edilene storck navarro	**edstorck@hotmail.com**	.	.	waltham	ma	02451	US	7815899726	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 8,292.77	$ -	$ (1,425.00)
10996431	1922040	**telexb04**	Edilene storck navarro	**edstorck@hotmail.com**	1128 trapelo rd	.	waltham	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 4,532.37	$ -	$ (1,425.00)
10996505	1922040	**telexb05**	Edilene storck navarro	**edstorck@hotmail.com**	1128 trapelo rd	.	waltham	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 8,205.68	$ -	$ (1,425.00)
10999437	1922040	**telexb06**	Edilene storck navarro	**edstorck@hotmail.com**	1128 trapelo rd	.	waltham	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,500.00	$ -	$ (1,425.00)
10999507	1922040	**telexb07**	Edilene storck navarro	**edstorck@hotmail.com**	1128 trapelo rd	.	waltham	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,500.00	$ -	$ (1,425.00)
11483849	1922040	telexboston02c	**Edilene s navarro**	**edstorck@hotmail.com**	e	e	r	e	s	US		Partner	Phone Plan / Other			$ (349.30)			$ -	$ -	$ -	$ (349.30)
11506343	1922040	tboston13	**Edilene s navarro**	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 6,164.55	$ -	$ (1,425.00)
11506705	1922040	tboston15	**Edilene s navarro**	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 6,164.55	$ -	$ (1,425.00)
11507133	1922040	tboston16	**Edilene s navarro**	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 5,854.73	$ -	$ (1,425.00)
11507193	1922040	tboston17	**Edilene s navarro**	**edstorck@hotmail.com**	**1128 Trapelo Rd**	1128	**Waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 4,854.73	$ -	$ (1,425.00)
11544787	1922040	eboston25	**edilene s navarro**	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,000.00	$ -	$ (1,425.00)
11545105	1922040	eboston28	**edilene s navarro**	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,000.00	$ -	$ (1,425.00)
11545135	1922040	eboston29	**edilene s navarro**	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,000.00	$ -	$ (1,425.00)
12094587	1922040	bboston38	**edilene s navarro**	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	ma	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,600.00	$ -	$ (1,425.00)
12520455	1922040	inuniversla3	**edilene s navarro**	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	MA	02451	US	**7818913536**	Partner	Promoter Plan			$ (50.00)			$ -	$ -	$ -	$ (50.00)
12520545	1922040	iuniversal5	**edilene s navarro**	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**	MA	02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,495.00)	$ 7,597.30		$ (9,206.94)	$ 3,328.00	$ -	$ 6,102.30
12520819	1922040	iuniversal3	**edilene s navarro**	**edstorck@hotmail.com**	**1128 trapelo rd**	.	**waltham**		02451	US	**7818913536**	ADCentral	Promoter Plan			$ (1,495.00)	$ 99.80		$ -	$ 1,900.00	$ -	$ (1,395.20)

Source(s): Based upon queries of the Huron Data Tables; and Exhibit 9.7

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.10: #10 Top Net Loser - Parent ID 15747789 ($390,000 Net Loser) v. Parent ID 16655191 ($171,000 Net Winner) and Parent ID 10844649 ($1.4M Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
15747789	$ (390,125.20)	371

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (393,000.00)	$ 2,874.80	$ -	$ (138,705.09)	$ 285,110.48	$ -	**$ (390,125.20)**
15791161	15747789	usqun003	qun zheng	**py900531@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	**6463215658**	ADCentral	Promoter Plan			$ (1,425.00)	$ 49.90		$ -	$ 3,789.90	$ -	$ (1,375.10)
15810237	15747789	usqun040	qun zheng	**py900531@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	**6463215658**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,016.91	$ -	$ (1,425.00)
15810283	15747789	usqun041	qun zheng	**py900531@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	**6463215658**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,016.91	$ -	$ (1,425.00)
16766669	15747789	usying039	qun zheng	**py900531@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	**6463215658**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 616.91	$ -	$ (1,425.00)
16766815	15747789	usying041	qun zheng	**py900531@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	**6463215658**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 614.90	$ -	$ (1,425.00)
16779683	15747789	usying090	qun zheng	**py900531@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	**6463215658**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
20710069	15747789	usnew257	qun zheng	**py900531@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
20710107	15747789	usnew258	qun zheng	**py900531@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
20720297	15747789	usnew286	qun zheng	**py900531@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	35200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
20720367	15747789	usnew288	qun zheng	**py900531@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	35200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
21529887	15747789	usnewying082	qun zheng	**panbingj88@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
21532239	15747789	usnewying085	qun zheng	**panbingj88@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
21532351	15747789	usnewying117	qun zheng	**panbingj88@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
16655191	$ 171,287.10	22

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (8,556.60)	$ 179,843.70	$ -	$ (348,282.21)	$ 218,945.17	$ -	**$ 171,287.10**
16655191	16655191	usying001	sheng jin che	**py900531@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	**6463215658**	ADCentral	Promoter Plan			$ (1,425.00)	$ 179,843.70		$ (290,507.83)	$ 117,443.00	$ -	$ 178,418.70
19635995	16655191	usche008	sheng jin che	**py900531@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ 1,255.38	$ -	$ (339.60)
19636015	16655191	usche016	sheng jin che	**py900531@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ 779.70	$ -	$ (339.60)
19636043	16655191	usche005	sheng jin che	**py900531@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ 2,348.30	$ -	$ (339.60)
19636071	16655191	usche011	sheng jin che	**py900531@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ 1,259.30	$ -	$ (339.60)
19636097	16655191	usche023	sheng jin che	**py900531@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ 779.70	$ -	$ (339.60)
19638007	16655191	usche006	sheng jin che	**py900531@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ 2,348.30	$ -	$ (339.60)
19638069	16655191	usche012	sheng jin che	**py900531@gmail.com**	**wu Hang Zhen Jie Fang Lu Tuan Jie Xiao Qu 11 Zuo 160**		**chang le shi**	35	350200	CN	**6463215658**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ 1,263.22	$ -	$ (339.60)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
10844649	$ 1,381,439.70	50

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (2,850.00)	$ -	$ (45,998.90)	$ 1,430,288.60	$ -	$ (2,228,455.23)	$ 1,138,084.44	$ -	**$ 1,381,439.70**
10844649	10844649	ng823	chai hock ng	panbingjian@yahoo.com	13708 31 st rd		**flushing**	ny	11354	US	6467013208	ADCentral	Promoter Plan			$ (1,425.00)	$ 249.50		$ -	$ 8,093.37	$ -	$ (1,175.50)
10871647	10844649	ng823a	chai hock ng	panbingjian@yahoo.com	135-25 Roosevelt ave	2d	**flushing**	ny	11354	US	6467013208	ADCentral	Promoter Plan			$ (1,425.00)	$ 199.60		$ -	$ 5,495.67	$ -	$ (1,225.40)
10879591	10844649	ng823g	chai hock ng	ngchaihock@yahoo.com	135-25 Roosevelt ave		**flushing**	NY	11354	US	6467013208	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 3,791.47	$ -	$ (1,325.20)
10879843	10844649	ng823e	chai hock ng	ngchaihock@yahoo.com	135-25 Roosevelt ave		**flushing**	NY	11354	US	6467013208	ADCentral	Promoter Plan			$ (1,425.00)	$ 199.60		$ -	$ 4,346.18	$ -	$ (1,225.40)
10917529	10844649	ng829lc1	chai hock ng	ngchaihock@yahoo.com	144-30roosevelt ave apt515		**flushing**	ny	11354	US	6467013208	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
10976249	10844649	chen829cc1	chai hock ng	baoyinc@yahoo.com	144-16 35 ave		**flushing**	ny	11355	US	6468231133	Partner	Phone Plan / Other			$ (399.20)			$ -	$ -	$ -	$ (399.20)
10990333	10844649	ng829y	chai hock ng	ngchaihock@yahoo.com	4170 main st		**flushing**	ny	11354	US	6467013208	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,427,144.90		$ (2,228,455.23)	$ 1,029,862.91	$ -	$ 1,425,719.90
29301278	10844649	us0407bb	chai hock ng	**panbingj88@gmail.com**	135-25 roosevelt ave		**flshing**	NY	11354	US	13478915475	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
33310406	10844649	newpan014	chai hock ng	**panbingj88@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	3479718531	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
33310808	10844649	newpan015	chai hock ng	**panbingj88@gmail.com**	**5735 lawrence st.**		**flushing**	NY	11355	US	3479718531	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.11: #11 Top Net Loser - Parent ID 3392494 ($378,000 Net Loser) v. Parent ID 4486096 ($28,000 Net Winner), Parent ID 4987334 ($948,000 Net Winner) and Parent ID 3780002 ($387,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
3392494	$ (377,696.00)	351

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (407,358.20)	$ 30,887.60	$ -	$ (468,158.48)	$ 748,238.28	$ -	$ (377,696.00)
4246864	3392494	weiwei0407c1	yuching wang	yuc6688@gmail.com	2929saint helen cir	2929	Silver Spring	md	20906	US	3019065998	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
4426182	3392494	weiwei0407c2	yuching wang	yuc6688@gmail.com	2929saint helen cir		Silver Spring	md	20906	US	3019065998	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
4426228	3392494	weiwei0407c3	yuching wang	yuc6688@gmail.com	2929saint helen cir		Silver Spring	md	20906	US	3018068899	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
4426290	3392494	weiwei0407c4	yuching wang	yuc6688@gmail.com	2929saint helen cir		Silver Spring	md	20906	US	3019065998	Partner	Phone Plan / Other			$ (149.70)			$ -	$ -	$ -	$ (49.90)
4426710	3392494	weiwei0407c5	yuching wang	yuc6688@gmail.com	2929saint helen cir		Silver Spring	md	20906	US	3018068899	Partner	Phone Plan / Other			$ (1,425.00)	$ 49.90		$ -	$ -	$ -	$ (149.70)
4493816	3392494	canada0416c1	yuching wang	yuc6688@gmail.com	2929saint helen cir		Silver Spring	md	20906	US	3019065998	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
6330274	3392494	canada0416c22	yuching wang	yuc6688@gmail.com	2212 bear valley terrace		Silver Spring	md	20906	US	3013647993	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (99.80)
6330328	3392494	canada0416c33	yuching wang	yuc6688@gmail.com	2212 bear valley terrace		Silver Spring	md	20906	US	3013647993	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
6330444	3392494	canada0416c44	yuching wang	yuc6688@gmail.com	2212 bear valley terrace		Silver Spring	md	20906	US	3014609589	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
6330494	3392494	canada0416c55	yuching wang	yuc6688@gmail.com	2212 bear valley terrace		Silver Spring	md	20906	US	3014609589	Partner	Phone Plan / Other			$ (1,425.00)			$ -	$ -	$ -	$ (49.90)
11472035	3392494	63yuqin01	youqin wang	speedyup@gmail.com	2212 bear valley terr		silver spring	MD	20906	US	3019065998	ADCentral	Promoter Plan			$ (349.30)			$ (110,657.11)	$ 118,673.77	$ -	$ (1,225.40)
11472337	3392494	63yuqin01a1	yuching wang	speedyup@gmail.com	2212 bear valley terr		silver spring	MD	20906	US	3019065998	Partner	Phone Plan / Other			$ (349.30)			$ -	$ -	$ -	$ (349.30)
11472497	3392494	63yuqin02	youqin wang	speedyup@gmail.com	2212 bear valley terr		silver spring	md	20906	US	3019065998	ADCentral	Promoter Plan			$ (1,425.00)	$ 149.70		$ -	$ 5,145.36	$ -	$ (1,275.30)
11472627	3392494	63yuqin01a2	yuching wang	speedyup@gmail.com	2212 bear valley terr		silver spring	md	20906	US	3019065998	Partner	Phone Plan / Other			$ (1,425.00)	$ 149.70		$ -	$ -	$ -	$ (349.30)
11473015	3392494	63yuqin003	youqin wang	speedyup@gmail.com	2212 bear valley terr		silver spring	MD	20906	US	3019065998	ADCentral	Promoter Plan			$ (349.30)			$ -	$ 5,075.06	$ -	$ (1,225.40)
11473177	3392494	63yuqin003a1	yuching wang	speedyup@gmail.com	2212 bear valley terr	2212	silver spring	md	20906	US	3019065998	Partner	Phone Plan / Other			$ (1,425.00)	$ 199.60		$ -	$ -	$ -	$ (349.30)
11473453	3392494	63yuqin04	youqin wang	speedyup@gmail.com	2212 bear valley terr	2212	silver spring	md	20906	US	3019065998	ADCentral	Promoter Plan			$ (349.30)			$ -	$ 3,667.36	$ -	$ (1,275.30)
13454643	3392494	yunyun2013c888	yuching wang	yuc6688@mail.com	2212 bear valley terrace		Silver Spring	MD	20905	US	3018068899	Partner	Phone Plan / Other			$ (149.70)			$ -	$ -	$ -	$ (49.90)
14809499	3392494	yunyun2013c9	yuching wang	yuc6688@mail.com	1648 18th st nw	85200	washington	DC	20011	US	2023048632	Partner	Phone Plan / Other						$ -	$ -	$ -	$ (149.70)
17497805	3392494	cyou001	Youqin Wang	telexfreewang@gmail.com	309 Shaoyao Ju Beili 4-601		Chaoyang	11	11355	CN	13019065998	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,144.91	$ -	$ (1,425.00)
17497901	3392494	cyou002	Youqin Wang	telexfreewang@gmail.com	309 Shaoyao Ju Beili 4-601		Chaoyang	11	11355	CN	13019065998	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,144.91	$ -	$ (1,425.00)
17498369	3392494	cyou003	Youqin Wang	telexfreewang@gmail.com	309 Shaoyao Ju Beili 4-601		Chaoyang	11	11355	CN	13019065998	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,724.91	$ -	$ (1,425.00)
17498435	3392494	cyou004	Youqin Wang	telexfreewang@gmail.com	309 Shaoyao Ju Beili 4-601		Chaoyang	11	11355	CN	13019065998	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,724.91	$ -	$ (1,425.00)
17498813	3392494	cyou005	Youqin Wang	telexfreewang@gmail.com	309 Shaoyao Ju Beili 4-601		Chaoyang	11	11355	CN	13019065998	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,724.91	$ -	$ (1,425.00)
17498905	3392494	cyou006	Youqin Wang	telexfreewang@gmail.com	309 Shaoyao Ju Beili 4-601		Chaoyang	11	11355	CN	13019065998	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,724.91	$ -	$ (1,425.00)
17536049	3392494	cyou007	Youqin Wang	telexfreewang@gmail.com	309 Shaoyao Ju Beili 4-601		Chaoyang	11	100102	CN	13019065998	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 904.91	$ -	$ (1,425.00)
17536145	3392494	cyou008	Youqin Wang	telexfreewang@gmail.com	309 Shaoyao Ju Beili 4-601		Chaoyang	11	100102	CN	13019065998	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 904.91	$ -	$ (1,425.00)
17536771	3392494	cyou009	Youqin Wang	telexfreewang@gmail.com	309 Shaoyao Ju Beili 4-601		Chaoyang	11	100102	CN	13019065998	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 904.91	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
4486096	$ 28,222.70	50

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (11,993.10)	$ 40,215.80	$ -	$ (408,612.84)	$ 403,809.05	$ -	$ 28,222.70
6069136	4486096	weiwei0407c11	zhexin wang	yuc6688@gmail.com	2212 bear valley terrace		Silver Spring	md	20906	US	3018068899	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
6069908	4486096	weiwei0407c22	zhexin wang	yuc6688@gmail.com	2212 bear valley terrace		Silver Spring	md	20906	US	3013647992	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
6070194	4486096	weiwei0407c33	zhexin wang	kivuvani@gmail.com	2212 bear valley terrace		Silver Spring	md	20906	US	3014609589	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
6070422	4486096	weiwei0407c44	zhexin wang	kolukying@gmail.com	2929saint helen cir		Silver Spring	md	20906	US	3014609589	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
6093290	4486096	canada0416c2	zhexin wang	yuc6688@gmail.com	2212 bear valley terrace		Silver Spring	md	20906	US	3014609589	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
6093340	4486096	canada0416c3	zhexin wang	yuc6688@gmail.com	2212 bear valley terrace		Silver Spring	md	20906	US	3013647993	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
6093366	4486096	canada0416c4	zhexin wang	yuc6688@gmail.com	2212 bear valley terrace		Silver Spring	md	20906	US	3013647992	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
6093380	4486096	canada0416c5	zhexin wang	yuc6688@gmail.com	2212 bear valley terrace		Silver Spring	md	20906	US	3013647993	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.11: #11 Top Net Loser - Parent ID 3392494 ($378,000 Net Loser) v. Parent ID 4486096 ($28,000 Net Winner), Parent ID 4987334 ($948,000 Net Winner) and Parent ID 3780002 ($387,000 Net Winner)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
4987334	$ 947,923.10	211

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (13,251.30)	$ 961,174.40	$ -	$ (1,038,550.37)	$ 132,966.90	$ -	**$ 947,923.10**
17498181	4987334	cyou001c1	Peng Zhang	jamess359s3@gmail.com	4345 Byrd St		Flushing	NY	11355	US	3478228765	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17498641	4987334	cyou002c1	Peng Zhang	jamess359s3@gmail.com	4345 Byrd St		Flushing	NY	11355	US	3478228765	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17535521	4987334	cyou003c1	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	3478228765	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17536477	4987334	cyou004c1	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	3478228765	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17537261	4987334	cyou005c1	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	3478228765	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17537715	4987334	cyou006c1	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	3478228765	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17541879	4987334	cyou007c1	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	3478228765	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17542427	4987334	cyou008c1	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	3478228765	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17543947	4987334	cyou009c1	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	3478228765	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17544193	4987334	cyou010c1	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	3478228765	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17544463	4987334	cyou011c1	Peng Zhang	james35s93@gmail.com	4345 Byrd St		Flushing	NY	11355	US	3478228765	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
3780002	$ 386,837.70	34

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (2,850.00)	$ 665.00	$ (43,498.50)	$ 432,521.20	$ -	$ (1,052,189.33)	$ 863,538.84	$ -	**$ 386,837.70**
3780002	3780002	liangyy0426	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	13588352073	ADCentral	Promoter Plan			$ (1,474.90)	$ 88,743.50		$ (173,992.50)	$ 139,960.64	$ -	$ 87,268.60
3780262	3780002	yunyun2013	Liang Yun Yun	**yuc6688@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	NULL	200040	CN	13588352073	ADCentral	Promoter Plan			$ (1,474.90)	$ 219,383.10		$ (715,331.46)	$ 514,615.28	$ -	$ 217,908.20
3808198	3780002	liangyy1001	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	62792354	ADCentral	Promoter Plan			$ (1,425.00)	$ 9,021.20		$ (17,341.73)	$ 27,101.71	$ -	$ 7,596.20
3808356	3780002	liangyy1002	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	US	62792354	ADCentral	Promoter Plan			$ (1,425.00)	$ 6,460.90		$ (9,885.16)	$ 29,364.49	$ -	$ 5,035.90
3911496	3780002	liangyy1003	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	62792354	ADCentral	Promoter Plan	$ (1,425.00)			$ 7,553.00		$ (6,199.64)	$ 3,671.33	$ -	$ 6,128.00
3911802	3780002	liangyy1004	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	62792354	ADCentral	Promoter Plan	$ (1,425.00)			$ 798.40		$ -	$ 3,885.43	$ -	$ (626.60)
3912004	3780002	liangyy1005	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	62792354	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,537.20		$ -	$ 4,140.72	$ -	$ 112.20
3940838	3780002	liangyy1009	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	62792354	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,147.70		$ (1,239.00)	$ 5,474.13	$ -	$ (277.30)
3941040	3780002	liangyy1010	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	62792354	ADCentral	Promoter Plan		$ 665.00	$ (1,425.00)	$ 30,296.40		$ (46,716.00)	$ 21,037.23	$ -	$ 29,536.40
4037686	3780002	michelle25	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	62792354	ADCentral	Promoter Plan			$ (1,425.00)	$ 5,694.30		$ (1,408.00)	$ 4,472.17	$ -	$ 4,269.30
4444078	3780002	michelle2501	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	62792354	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,047.90		$ -	$ 3,934.93	$ -	$ (377.10)
4444310	3780002	michelle2502	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	62792354	ADCentral	Promoter Plan			$ (1,425.00)	$ 2,922.00		$ (1,138.00)	$ 3,198.83	$ -	$ 1,497.00
11647087	3780002	atlanda01	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	62792354	ADCentral	Promoter Plan			$ (1,425.00)	$ 299.40		$ (3,136.72)	$ 10,384.27	$ -	$ (1,125.60)
11649697	3780002	jeffpan01	Liang Yun Yun	**speedyup@gmail.com**	Jingan Qu Julu Rd. 588 Long 13 Hao	588	Shanghai	Sh	200040	CN	62792354	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ (420.00)	$ 4,037.95	$ -	$ (1,325.20)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.12: #12 Top Net Loser - Parent ID 15084253 ($361,000 Net Loser) v. Parent ID 632789 ($968,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
15084253	$ (360,774.50)	266

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	Neighborhood	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
															$ -	$ -	$ (361,822.40)	$ 1,047.90	$ -	$ -	$ 95,008.55	$ -	**$ (360,774.50)**
15088101	15084253	**grupoboston21**	bryan matt	**grupoboston@icloud.com**	av ecuador	10		santiago	AI	8980000	CL	617344782	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15088217	15084253	**grupoboston22**	bryan matt	**grupoboston@icloud.com**	av ecuador	10		santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15088333	15084253	**grupoboston23**	bryan matt	**grupoboston@icloud.com**	av ecuador	10		santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15088415	15084253	**grupoboston24**	bryan matt	**grupoboston@icloud.com**	av ecuador	10		santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15088543	15084253	**grupoboston25**	bryan matt	**grupoboston@icloud.com**	av ecuador	10		santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15088897	15084253	**grupoboston26**	bryan matt	**grupoboston@icloud.com**	av ecuador	10		santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
15088991	15084253	**grupoboston27**	bryan matt	**grupoboston@icloud.com**	av ecuador	10		santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 700.00	$ -	$ (1,425.00)
16466553	15084253	**hyannis40**	bryan matt	alvarodesouza11@gmail.com	10 ave ecuador			santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
16466667	15084253	**hyannis41**	bryan matt	alvarodesouza11@gmail.com	10 ave ecuador			santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
16466831	15084253	**hyannis42**	bryan matt	alvarodesouza11@gmail.com	10 ave ecuador			santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
16466949	15084253	**hyannis43**	bryan matt	alvarodesouza11@gmail.com	10 ave ecuador			santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
16467041	15084253	**hyannis44**	bryan matt	alvarodesouza11@gmail.com	10 ave ecuador			santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
16467129	15084253	**hyannis45**	bryan matt	**grupoboston@icloud.com**	10 ave ecuador			santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 500.00	$ -	$ (1,425.00)
16467255	15084253	**hyannis46**	bryan matt	**grupoboston@icloud.com**	10 ave ecuador			santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 500.00	$ -	$ (1,425.00)
16467347	15084253	**hyannis47**	bryan matt	**grupoboston@icloud.com**	10 ave ecuador			santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 500.00	$ -	$ (1,425.00)
16769893	15084253	**grupoboston35**	bryan matt	**grupoboston@icloud.com**	10 ave ecuador			santiago	AI	8980000	CL	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)	$ 399.20		$ -	$ 2,884.91	$ -	$ (1,025.80)
15378255	15084253	saugus94	bryan matt	**grupoboston@icloud.com**	10 ave ecuador			santiago	AI	8980000	CL	6173344780	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
632789	$ 968,101.20	1,331

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	Neighborhood	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
															$ -	$ -	$ (671,068.10)	$ 1,727,034.30	$ (133,540.00)	$ (2,146,258.63)	$ 1,777,185.23	$ -	**$ 968,101.20**
4640888	632789	**grupoboston**	wesley dias	**grupoboston@icloud.com**	30 Franklin St	30	centro	Malden		02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)	$ 773,991.10		$ (1,014,124.36)	$ 433,756.81	$ -	$ 772,566.10
6443448	632789	**grupoboston2**	wesley dias	wesley_comcast@yahoo.com	30 Franklin St	30	centro	Malden	ma	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)	$ 399.20		$ -	$ 3,900.00	$ -	$ (1,025.80)
11947687	632789	**grupoboston3**	Wesley Gome	**grupoboston@icloud.com**	30 Franklin St	30		Malden	ma	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)	$ 399.20		$ -	$ 1,700.00	$ -	$ (1,025.80)
11947719	632789	**grupoboston4**	Wesley Gome	**grupoboston@icloud.com**	30 Franklin St	30		Malden	ma	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)	$ 399.20		$ -	$ 1,700.00	$ -	$ (1,025.80)
11947755	632789	**grupoboston5**	Wesley Gome	**grupoboston@icloud.com**	30 Franklin St	30		Malden	ma	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)	$ 399.20		$ -	$ 1,700.00	$ -	$ (1,025.80)
11948047	632789	**grupoboston10**	Wesley Gome	**grupoboston@icloud.com**	30 Franklin St	30		Malden	ma	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,100.00	$ -	$ (1,425.00)
11948073	632789	**grupoboston11**	Wesley Gome	**grupoboston@icloud.com**	30 Franklin St	30		Malden	ma	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,100.00	$ -	$ (1,425.00)
14354855	632789	**hyannis13**	Wesley Dias	**grupoboston@icloud.com**	30 Franklin st	30		Malden	MA	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)	$ 49.90		$ -	$ 4,460.91	$ -	$ (1,375.10)
14355219	632789	**hyannis**	Wesley Dias	**grupoboston@icloud.com**	30 Franklin st	30		Malden	MA	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)	$ 15,675.00		$ (15,816.01)	$ 1,141.01	$ -	$ 14,250.00
14355577	632789	**hyannis1**	Wesley Dias	**grupoboston@icloud.com**	30 Franklin st	30		Malden	MA	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,000.00	$ -	$ (1,425.00)
14356025	632789	**hyannis10**	Wesley Dias	**grupoboston@icloud.com**	30 Franklin st			Malden	MA	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 400.00	$ -	$ (1,425.00)
14356101	632789	**hyannis11**	Wesley Dias	**grupoboston@icloud.com**	30 Franklin st			Malden	MA	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 400.00	$ -	$ (1,425.00)
14356135	632789	**hyannis12**	Wesley Dias	**grupoboston@icloud.com**	30 Franklin st			Malden	MA	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 400.00	$ -	$ (1,425.00)
14356429	632789	**hyannis16**	Wesley Dias	**grupoboston@icloud.com**	30 Franklin st			Malden	MA	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 4,761.92	$ -	$ (1,325.20)
14356481	632789	**hyannis15**	Wesley Dias	**grupoboston@icloud.com**	30 Franklin st			Malden	MA	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 400.00	$ -	$ (1,425.00)
18654749	632789	**grupoboston100**	Wesley Gome	**grupoboston@icloud.com**	30 Franklin st	30		Malden	MA	02148	US	**6173344782**	ADFamily	Promoter Plan			$ (339.60)			$ (74,880.11)	$ 11,493.00	$ -	$ (339.60)
11773739	632789	revere44	wesley dias	alvarodesouza11@gmail.com	30 Franklin St	30		Malden	ma	02148	US	**6173344782**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,800.00	$ -	$ (1,425.00)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.13: #13 Top Net Loser - Parent ID 10630965 ($350,000 Net Loser) v. Parent ID 312457 ($260,000 Net Winner) and Parent ID 4993498 ($211,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
10630965	$ (349,610.40)	431

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (362,308.80)	$ 12,698.40	$ -	$ (36,781.92)	$ 413,935.19	$ -	**$ (349,610.40)**
10630965	10630965	mainnetwork2	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz Miraflores	1	Lima	LI	00000	PE	**7742395643**	ADCentral	Promoter Plan			$ (1,425.00)	$ 4,574.40		$ (19,658.92)	$ 19,929.35	$ -	$ 3,149.40
10781505	10630965	mainnetwork3	Gerano Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz Miraflores		Lima	LI	00000	PE	7742393096	ADCentral	Promoter Plan			$ (1,425.00)	$ 49.90		$ -	$ 2,350.10	$ -	$ (1,375.10)
11083969	10630965	aintercash22	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
11084041	10630965	intercash24	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,300.00	$ -	$ (1,425.00)
11084093	10630965	intercash25	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 5,896.55	$ -	$ (1,425.00)
11084147	10630965	aintercash25	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
11084189	10630965	intercash26	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	0	PE	5087911869	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 4,000.55	$ -	$ (1,425.00)
11084231	10630965	intercash27	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 4,372.55	$ -	$ (1,425.00)
11084285	10630965	aintercash26	Genaro Olortegui Carrera	**olorteguiomar@live.com**	calle Mariano Melgar 373 Santa Cruz		Miraflores	li	00000	PE	5087911869	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
11084337	10630965	intercash28	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,268.55	$ -	$ (1,425.00)
11084373	10630965	intercash29	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,671.64	$ -	$ (1,425.00)
11084403	10630965	aintercash27	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
11084463	10630965	intercash30	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,274.55	$ -	$ (1,425.00)
11084551	10630965	aintercash28	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
11084659	10630965	aintercash29	Genaro Olortegui Carrera	**olorteguiomar@live.com**	calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
11396161	10630965	aintercash23	genaro Olortegui Carrera	**olorteguiomar@live.com**	calle Mariano Melgar 373 Santa Cruz		miraflores	li	00000	PE	5087911869	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
11454765	10630965	aintercash10	genaro Olortegui Carrera	**olorteguiomar@live.com**	calle Mariano Melgar 373 Santa Cruz		miraflores	li	00000	PE	5087911869	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
12025451	10630965	ggladysusa5	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	Li	00000	PE	5087911869	Partner	Phone Plan / Other			$ (149.70)			$ -	$ -	$ -	$ (149.70)
12025507	10630965	ggladysusa7	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		miraflores	LI	00000	PE	5087911869	Partner	Phone Plan / Other			$ (149.70)			$ -	$ -	$ -	$ (149.70)
12025639	10630965	ggladysusa27	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373 Santa Cruz		Miraflores	li	00000	PE	5087911869	Partner	Phone Plan / Other			$ (199.60)			$ -	$ -	$ -	$ (199.60)
12195263	10630965	mainnetwork10	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373		Miraflores	LIM	00000	PE	5087911869	ADCentral	Promoter Plan			$ (1,425.00)	$ 49.90		$ -	$ 1,580.83	$ -	$ (1,375.10)
12195357	10630965	mainnetwork11	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373		Miraflores	LIM	00000	PE	5087911869	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,604.91	$ -	$ (1,425.00)
12195379	10630965	mainnetwork12	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373		Miraflores	LIM	00000	PE	5087911869	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,200.00	$ -	$ (1,425.00)
18203825	10630965	mainnetwork13	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373		Miraflores	LIM	00000	PE	1234567890	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
18203899	10630965	mainnetwork14	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373		Miraflores	LIM	00000	PE	1234567890	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
27479466	10630965	tmt02	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373	373	Miraflores	LIM	00000	PE	1234567890	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
27480602	10630965	tmt03	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373		Miraflores	LIM	00000	PE	1234567890	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
27481068	10630965	tmt04	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373		Miraflores	LIM	00000	PE	1234567890	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
27482716	10630965	tmt05	Genaro Olortegui Carrera	**olorteguiomar@live.com**	Calle Mariano Melgar 373		Miraflores	LIM	00000	PE	1234567890	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
312457	$ 259,813.60	357

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (18,025.00)	$ 410.00	$ (115,714.83)	$ 407,393.43	$ (14,250.00)	$ (496,484.64)	$ 437,021.41	$ -	**$ 259,813.60**
312457	312457	gladysusa	GLADYS ALVARADO	gladys_alvarado65@hotmail.com	177 LINCOLN ST # 2	2	WORCEST	MA	1605	US	7742395643	ADCentral	Promoter Plan		$ 410.00	$ (10,146.83)	$ 339,489.23	$ (14,250.00)	$ (427,970.80)	$ 119,690.93	$ -	$ 315,502.40
1082775	312457	agladysusa5	Gladys Alvarado	**olorteguiomar@live.com**	177 Lincoln st	2	Worcester	MA	1605	US	7742399351	Partner	Phone Plan / Other			$ (249.50)			$ -	$ -	$ -	$ (249.50)
1083047	312457	gladysusa7	Gladys Alvarado	**olorteguiomar@live.com**	177 Lincoln st	2	Worcester	MA	1605	US	7742399351	ADCentral	Promoter Plan			$ (1,375.00)	$ 27,336.30		$ (34,385.43)	$ 29,659.55	$ -	$ 25,961.30
5521752	312457	gladysusa43	Gladys Alvarado	olortegui_omar@aol.com	177 Lincoln St.	1	Worcester	MA	1605	US	7742399351	ADCentral	Promoter Plan			$ (1,425.00)	$ 199.60		$ -	$ 4,224.95	$ -	$ (1,225.40)
5597340	312457	dgladysusa43	Gladys Alvarado	olortegui_omar@aol.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742399351	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
7397886	312457	aintercash6	Gladys Alvarado	olortegui_omar@aol.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742399351	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
9665483	312457	aintercash7	Gladys Alvarado	olortegui_omar@aol.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742399351	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
9670191	312457	aintercash8	Gladys Alvarado	olortegui_omar@aol.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742399351	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10346919	312457	aintercash18	Gladys Alvarado	**olorteguiomar@live.com**	177 Lincoln St.	2	Worcester	MA	1605	US	7742399351	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
10371453	312457	ggladysusa26	Gladys Alvarado	**olorteguiomar@live.com**	177 Lincoln St.	2	Worcester	MA	1605	US	7742399351	Partner	Phone Plan / Other			$ (199.60)			$ -	$ -	$ -	$ (199.60)
10710677	312457	ggladysusa23	Gladys Alvarado	**olorteguiomar@live.com**	177 Lincoln St		Worcester	MA	1605	US	7742399351	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.13: #13 Top Net Loser - Parent ID 10630965 ($350,000 Net Loser) v. Parent ID 312457 ($260,000 Net Winner) and Parent ID 4993498 ($211,000 Net Winner)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
4993498	$ 210,873.00	27

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (1,425.00)	$ 7,000.00	$ (34,879.20)	$ 240,177.20	$ -	$ (645,606.63)	$ 533,750.49	$ -	$ 210,873.00
4993498	4993498	intercash	Omar Olortegui	olortegui_omar@aol.com	177 Lincoln St.	1	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan		$ 7,000.00	$ (1,425.00)	$ 299.40		$ (7,097.03)	$ 3,900.60	$ -	$ 5,874.40
5020388	4993498	intercash3	Omar Olortegui	olortegui_omar@aol.com	177 Lincoln St.	1	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)			$ (11,047.00)	$ 16,311.37	$ -	$ (1,425.00)
5627696	4993498	intercash4	Omar Olortegui	olortegui_omar@aol.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)	$ 249.50		$ -	$ 6,429.09	$ -	$ (1,175.50)
7329870	4993498	intercash5	Omar Olortegui	olortegui_omar@aol.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan	$ (1,425.00)					$ -	$ 3,900.00	$ -	$ (1,425.00)
7397394	4993498	intercash6	Omar Olortegui	olortegui_omar@aol.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)	$ 199.60		$ -	$ 4,657.68	$ -	$ (1,225.40)
7398220	4993498	intercash7	Omar Olortegui	olortegui_omar@aol.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)	$ 638.40		$ -	$ 4,046.88	$ -	$ (786.60)
9669493	4993498	intercash8	Omar Olortegui	olortegui_omar@aol.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)	$ 149.70		$ -	$ 4,063.58	$ -	$ (1,275.30)
9670601	4993498	intercash9	Omar Olortegui	olortegui_omar@aol.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,600.00	$ -	$ (1,425.00)
9730639	4993498	mainnetwork	Omar Olortegui	olortegui_omar@aol.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)	$ 49.90		$ (1,985.00)	$ 5,135.10	$ -	$ (1,375.10)
10344189	4993498	intercash10	Omar Olortegui	olorteguiomar@live.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 4,398.55	$ -	$ (1,425.00)
10344245	4993498	intercash11	Omar Olortegui	olorteguiomar@live.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)	$ 49.90		$ -	$ 6,747.38	$ -	$ (1,375.10)
10344297	4993498	intercash12	Omar Olortegui	olorteguiomar@live.com	177 Lincoln	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)	$ 49.90		$ -	$ 8,901.38	$ -	$ (1,375.10)
10344339	4993498	intercash13	Omar Olortegui	olorteguiomar@live.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,300.00	$ -	$ (1,425.00)
10344419	4993498	intercash14	Omar Olortegui	olorteguiomar@live.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)	$ 2,427.10		$ -	$ 10,184.98	$ -	$ 1,002.10
10344441	4993498	intercash15	Omar Olortegui	olorteguiomar@live.com	177 Lincoln St.	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,400.00	$ -	$ (1,425.00)
10344525	4993498	intercash16	Omar Olortegui	olorteguiomar@live.com	177 Lincoln St	2	Worcester	MA	1605	US	7742395643	ADCentral	Promoter Plan			$ (1,425.00)	$ 3,106.30		$ -	$ 11,097.38	$ -	$ 1,681.30
26845856	4993498	tmt00	Omar Olortegui	dmservices2014@gmail.com	28 New St		Worcester	MA	1605	US	7742395643	Associate	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
26849708	4993498	tmt01	Omar Olortegui	dmservices2014@gmail.com	28		Worcester	MA	1605	US	7742395643	ADFamily	Promoter Plan			$ (339.60)			$ (14,420.00)	$ -	$ -	$ (339.60)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.14: #14 Top Net Loser - Parent ID 15772887 ($307,000 Net Loser)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
15772887	$ (307,008.50)	6,759

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (307,009.10)	$ 300.00	$ (262,993.10)	$ 262,693.70	$ -	$ (297.00)	$ 300.00	$ -	**$ (307,008.50)**
15773745	15772887	fook2	LAI YEE FOOK	rizatelex@gmail.com	Unit 504. 5/F., Oriental Center,		KLN		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 38,523.00		$ -	$ -	$ -	$ 37,098.00
15838691	15772887	fook7	Lai Yee FOOK	rizatelex@gmail.com	Unit 504, 5/F., Oriental Center,67-71 Chatham Road South, TST		KLN		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 888.50		$ -	$ -	$ -	$ (536.50)
15838847	15772887	fook8	Lai Yee FOOK	rizatelex@gmail.com	Unit 504, 5/F., Oriental Center,67-71 Chatham Road South, TST		KLN		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 898.20		$ -	$ -	$ -	$ (526.80)
15839633	15772887	fook9	Lai Yee FOOK	rizatelex@gmail.com	Unit 504, 5/F., Oriental Center,67-71 Chatham Road South, TST		KLN		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 898.20		$ -	$ -	$ -	$ (526.80)
15840157	15772887	fook010	Lai Yee FOOK	rizatelex@gmail.com	Unit 504, 5/F., Oriental Center,67-71 Chatham Road South, TST		KLN		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 888.50		$ -	$ -	$ -	$ (536.50)
16230911	15772887	fook11	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Pok Fu Lam		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 399.20		$ -	$ -	$ -	$ (1,025.80)
16723293	15772887	fook12	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace, Pok Fu Lam, Hong Kong	60133	Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 399.20		$ -	$ -	$ -	$ (1,025.80)
16723623	15772887	fook13	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace, Pok Fu Lam		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 399.20		$ -	$ -	$ -	$ (1,025.80)
16723833	15772887	fook14	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace, Pok Fu Lam		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 399.20		$ -	$ -	$ -	$ (1,025.80)
16723963	15772887	fook15	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace, Pok Fu Lam		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 399.20		$ -	$ -	$ -	$ (1,025.80)
16724139	15772887	fook16	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace, Pok Fu Lam		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 399.20		$ -	$ -	$ -	$ (1,025.80)
17405481	15772887	fook101	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 17,022.90		$ -	$ -	$ -	$ 15,597.90
17405723	15772887	fook102	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 1,018.80		$ -	$ -	$ -	$ (406.20)
17405873	15772887	fook103	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 679.20		$ -	$ -	$ -	$ (745.80)
17406017	15772887	fook104	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 339.60		$ -	$ -	$ -	$ (1,085.40)
17507581	15772887	fook108	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 16,572.40		$ -	$ -	$ -	$ 15,147.40
17507765	15772887	fook112	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 1,018.80		$ -	$ -	$ -	$ (406.20)
17508057	15772887	fook105	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 10,185.20		$ -	$ -	$ -	$ 8,760.20
17508271	15772887	fook106	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 7,546.00		$ -	$ -	$ -	$ 6,121.00
17508465	15772887	fook107	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 7,256.30		$ -	$ -	$ -	$ 5,831.30
17508733	15772887	fook109	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 339.60		$ -	$ -	$ -	$ (1,085.40)
17508967	15772887	fook110	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 339.60		$ -	$ -	$ -	$ (1,085.40)
17509155	15772887	fook111	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 339.60		$ -	$ -	$ -	$ (1,085.40)
17509445	15772887	fook113	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)		$ (49.90)	$ 6,412.20		$ -	$ -	$ -	$ 4,937.30
17509919	15772887	fook114	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)		$ (49.90)	$ 6,801.70		$ -	$ -	$ -	$ 5,326.80
17510949	15772887	fook118	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 6,891.80		$ -	$ -	$ -	$ 5,466.80
17511137	15772887	fook119	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 339.60		$ -	$ -	$ -	$ (1,085.40)
17511411	15772887	fook120	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADCentral	Promoter Plan	$ (1,425.00)			$ 339.60		$ -	$ -	$ -	$ (1,085.40)
18227489	15772887	fook129	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (359.50)	$ 12,456.10		$ -	$ -	$ -	$ 12,096.60
18404805	15772887	fook131	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (359.50)	$ 3,665.40		$ -	$ -	$ -	$ 3,305.90
18405509	15772887	fook132	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (359.50)	$ 3,615.50		$ -	$ -	$ -	$ 3,256.00
18410813	15772887	fook133	LAI YEE FOOK	rizatelex@gmail.com	Flat A,3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan	$ (339.60)		$ (19.90)	$ 3,515.70		$ -	$ -	$ -	$ 3,156.20
18411045	15772887	fook134	LAI YEE FOOK	rizatelex@gmail.com	Flat A,3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan	$ (339.60)		$ (19.90)	$ 3,555.50		$ -	$ -	$ -	$ 3,196.00
18411989	15772887	fook136	LAI YEE FOOK	rizatelex@gmail.com	Flat A,3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan	$ (339.60)		$ (19.90)	$ 3,005.60		$ -	$ -	$ -	$ 2,646.10
18776147	15772887	fook151	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan	$ (339.60)		$ (19.90)	$ 5,648.90		$ -	$ -	$ -	$ 5,289.40
18776431	15772887	fook152	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden,15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan	$ (339.60)		$ (19.90)	$ 5,671.10		$ -	$ -	$ -	$ 5,311.60
18777059	15772887	fook153	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan	$ (339.60)		$ (19.90)	$ 4,441.60		$ -	$ -	$ -	$ 4,082.10
18778077	15772887	fook155	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan	$ (339.60)		$ (19.90)	$ 3,295.30		$ -	$ -	$ -	$ 2,935.80
19982691	15772887	fook255	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden, 15-19 To Li Terrace		Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan	$ (339.60)					$ -	$ -	$ -	$ (339.60)
21541399	15772887	fook386	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan	$ (339.60)					$ -	$ -	$ -	$ (339.60)
23304935	15772887	fook455	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
23305047	15772887	fook456	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
23678666	15772887	fook513	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
23687822	15772887	fook524	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
23752968	15772887	fook551	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
24277630	15772887	fook563	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
24287122	15772887	fook582	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
24287290	15772887	fook583	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25527970	15772887	fook609	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25528202	15772887	fook610	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25575870	15772887	fook621	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25585400	15772887	fook639	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25585542	15772887	fook640	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
26891114	15772887	fook5831	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
26891338	15772887	fook5832	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
30702310	15772887	fook696	LAI YEE FOOK	rizatelex@gmail.com	Flat A, 3/F., To Li Garden	00852	Hong Kong		00852	HK	60133662	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.15: #15 Top Net Loser - Parent ID 7127624 ($291,000 Net Loser) v. Parent ID 1183234 ($41,000 Net Winner) and Parent ID 1309022 ($814,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
7127624	$ (291,301.25)	319

Example of User Accounts:

														$ (14,250.00)	$ 56,282.55	$ (366,069.60)	$ 35,235.80	$ (2,500.00)	$ (131,332.41)	$ 491,315.90	$ - **$ (291,301.25)**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
7127624	7127624	imadthweib	Imad Thweib	zthweib@gmail.com	44 princess rahma street	.	amman	.	.	US	011962797696868	ADCentral	Promoter Plan		$ 1,826.65	$ (1,425.00)	$ 16,473.40	$ (250.00)	$ (83,588.41)	$ 23,169.56	$ -	$ 16,625.05
11161119	7127624	sthweib	Imad Thweib	zthweib@gmail.com	44 princess rahma st	.	amman	.	.	JO	011962797696868	ADCentral	Promoter Plan		$ 47,856.00	$ (1,425.00)	$ 798.40		$ (47,744.00)	$ 5,932.52	$ -	$ 47,229.40
11780117	7127624	**rlinetsb**	Imad Thweib	**zthweib@gmail.com**	44 princes st	.	amman	.	.	JO	011962797696868	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 4,063.10	$ -	$ (1,425.00)
11780317	7127624	**rlinetsb1**	Imad Thweib	**zthweib@gmail.com**	44 princes st	.	amman	.	.	JO	011962797696868	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,124.55	$ -	$ (1,425.00)
11780359	7127624	**rlinetsb2**	Imad Thweib	**zthweib@gmail.com**	44 princes st	.	amman	.	.	JO	011962797696868	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,824.55	$ -	$ (1,425.00)
11780397	7127624	**rlinetsb3**	Imad Thweib	**zthweib@gmail.com**	.	.	amman	.	.	JO	011962797696868	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,124.55	$ -	$ (1,425.00)
11780431	7127624	**rlinetsb4**	Imad Thweib	**zthweib@gmail.com**	44 princes st	.	amman	.	.	JO	011962797696868	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,124.55	$ -	$ (1,425.00)
11780471	7127624	**rlinetsb5**	Imad Thweib	**zthweib@gmail.com**	44 princes st	.	amman	.	.	JO	011962797696868	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,124.55	$ -	$ (1,425.00)
11780519	7127624	**rlinetsb6**	Imad Thweib	**zthweib@gmail.com**	44 princes st	.	amman	.	.	JO	011962797696868	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,824.55	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
1183234	$ 40,898.93	17

Example of User Accounts:

														$ (1,375.00)	$ 22,170.63	$ (14,249.00)	$ 49,827.30	$ (15,475.00)	$ (31,334.80)	$ 107,068.68	$ - **$ 40,898.93**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
1273648	1183234	zthweib8	zaid thweib	**zthweib@gmail.com**	1307 commonwealth ave	.	allston	Ma	2134	US	6178882435	ADCentral	Promoter Plan		$ 900.00	$ (1,375.00)	$ 299.40		$ -	$ 3,800.60	$ -	$ (175.60)
1273748	1183234	zthweib9	zaid thweib	**zthweib@gmail.com**	1307 commonwealth ave	.	allston	Ma	2134	US	6178882435	ADCentral	Promoter Plan		$ 900.00	$ (1,375.00)	$ 299.40		$ -	$ 3,900.60	$ -	$ (175.60)
1273870	1183234	zthweib10	zaid thweib	**zthweib@gmail.com**	1307 commonwealth ave	.	allston	Ma	2134	US	6178882435	ADCentral	Promoter Plan		$ 900.00	$ (1,375.00)	$ 299.40		$ -	$ 3,800.60	$ -	$ (175.60)
1274044	1183234	zthweibcustomer	zaid thweib	**zthweib@gmail.com**	1307 commonwealth ave	.	allston	Ma	2134	US	6178882435	Partner	Phone Plan / Other			$ (249.50)			$ -	$ -	$ -	$ (249.50)
3441116	1183234	zthweibc1	zaid thweib	**zthweib@gmail.com**	22 9th st, apt # 702	.	medford	ma	2155	US	6178882435	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
7956472	1183234	**rlinetsc6**	zaid thweib	**zthweib@gmail.com**	.	.	.	.	.	US		Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
7956756	1183234	**rlinetsc7**	zaid thweib	**zthweib@gmail.com**	.	.	.	.	.	US		Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
7957254	1183234	**rlinetsc10**	zaid thweib	**zthweib@gmail.com**	.	.	.	.	.	US		Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
7957664	1183234	**rlinetsc13**	zaid thweib	**zthweib@gmail.com**	.	.	.	.	.	US		Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
1309022	$ 814,074.10	224

Example of User Accounts:

														$ (14,250.00)	$ 35,196.60	$ (57,891.70)	$ 853,519.20	$ (2,500.00)	$ (1,472,156.61)	$ 1,046,781.40	$ - **$ 814,074.10**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
1309022	1309022	athweib	Adel Thweib	**zthweib@gmail.com**	44 Princess rahma st	.	khalda	NULL	.	JO	6178882435	ADCentral	Promoter Plan		$ 300.00	$ (1,455.00)	$ 812,290.00		$ (1,435,162.41)	$ 754,425.14	$ -	$ 811,135.00
2938828	1309022	athweibc4	Adel Thweib	**zthweib@gmail.com**	.	.	.	.	.	US		Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
2939062	1309022	athweibc5	Adel Thweib	**zthweib@gmail.com**	.	.	.	.	.	US		Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
2940768	1309022	athweib1	Adel Thweib	**zthweib@gmail.com**	44 Princess rahma st	.	khalda	.	.	JO		ADCentral	Promoter Plan		$ 600.00	$ (1,425.00)	$ 3,648.40		$ (2,579.00)	$ 3,654.55	$ -	$ 2,823.40
2941812	1309022	athweib2	Adel Thweib	**zthweib@gmail.com**	44 Princess rahma st	.	khalda	.	.	JO		ADCentral	Promoter Plan		$ 300.00	$ (1,495.00)	$ 798.40		$ -	$ 4,225.55	$ -	$ (396.60)
2941972	1309022	athweib3	Adel Thweib	**zthweib@gmail.com**	.	.	.	.	.	US		ADCentral	Promoter Plan		$ 300.00	$ (1,495.00)	$ 798.40		$ -	$ 4,225.55	$ -	$ (396.60)
2944500	1309022	athweib5	Adel Thweib	**zthweib@gmail.com**	44 Princess rahma st	.	khalda	.	.	JO		ADCentral	Promoter Plan		$ 300.00	$ (1,425.00)	$ 798.40		$ -	$ 4,225.55	$ -	$ (326.60)

Source(s): Based upon queries of the Huron Data Tables

Exhibit 9.16: #16 Top Net Loser - Parent ID 117321 ($278,000 Net Loser) v. Parent ID 11034097 ($445,000 Net Winner) and Parent ID 12294889 ($17,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
117321	$ (278,072.80)	738

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (320,076.90)	$ 42,004.10	$ -	$ (112,428.79)	$ 215,697.90	$ -	**$ (278,072.80)**
12295111	117321	**rmmmn20**	renato siamo honorio dos santos	**renato_simao@hotmail.com**	av	**836**	**jauru**	mat	78255000	US	**6599745032**	ADCentral	Promoter Plan			$ (1,574.70)	$ 3,721.80		$ (1,200.00)	$ 8,792.34	$ -	$ 2,147.10
14310967	117321	**rmmmn108**	renato simao honorio dos santos	**renato_simao@hotmail.com**	av brasil	**836**	**jauru**	1	78255000	PY	**6599745032**	ADCentral	Promoter Plan			$ (1,524.80)	$ 11,499.80		$ (14,069.60)	$ 3,678.71	$ -	$ 9,975.00
29083366	117321	**rmmmn270**	renato simao honorio dos santos	**renato_simao@hotmail.com**	av brasil	**836**	**jauru**	AR	78255000	UY	**6599745032**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
29084414	117321	**rmmmn271**	renato simao honorio dos santos	**renato_simao@hotmail.com**	av brasil	**836**	**jauru**	AR	78255000	UY	**6599745032**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
29085430	117321	**rmmmn272**	renato simao honorio dos santos	**renato_simao@hotmail.com**	av brasil	**836**	**jauru**	AR	78255000	UY	**6599745032**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
29086590	117321	**rmmmn273**	renato simao honorio dos santos	**renato_simao@hotmail.com**	av brasil	**836**	**jauru**	AR	78255000	UY	**6599745032**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
29087452	117321	**rmmmn274**	renato simao honorio dos santos	**renato_simao@hotmail.com**	av brasil	**836**	**jauru**	AR	78255000	UY	**6599745032**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
29088920	117321	**rmmmn275**	renato simao honorio dos santos	**renato_simao@hotmail.com**	av brasil	**836**	**jauru**	AR	78255000	UY	**6599745032**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
11034097	$ 444,519.50	1

User Account:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (339.00)	$ 444,858.50	$ -	$ (446,154.47)	$ 59,820.62	$ -	**$ 444,519.50**
11034097	11034097	costatfgbi	ANDERSON NEVES	**renato_simao@hotmail.com**	RUA FLORIANO PEIXOTO	45	GUANAMBI	AL	46430000	US	7799942259	ADCentral	Promoter Plan			$ (339.00)	$ 444,858.50		$ (446,154.47)	$ 59,820.62	$ -	$ 444,519.50

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
12294889	$ 17,299.60	1

User Account:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (1,574.70)	$ 18,874.30	$ -	$ (20,537.06)	$ 16,899.00	$ -	**$ 17,299.60**
12294889	12294889	**rmmmn19**	renato siamo honoris dos santos	**renato_simao@hotmail.com**	av	**836**	**jauru**	mat	78255000	US	**6599745032**	ADCentral	Promoter Plan			$ (1,574.70)	$ 18,874.30		$ (20,537.06)	$ 16,899.00	$ -	$ 17,299.60

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.17: #17 Top Net Loser - Parent ID 227597 ($278,000 Net Loser) v. Parent ID 10525055 ($34,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
227597	$ (277,789.20)	12

User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (7,966.30)	$ 154,502.10	$ (424,325.00)	$ (755,700.53)	$ 1,387,345.90	$ -	**$ (277,789.20)**
10525055	227597	**reclap**	**Cleonice Ribeiro**	cleoniceclr@gmail.com	Silver Leaf Way # 1534	15	Peabody		01960	US	**6178035838**	ADCentral	Promoter Plan			$ (1,495.00)	$ 54,913.80		$ (505,204.21)	$ 725,817.09	$ -	$ 53,418.80
10525467	227597	**reclap2**	**Cleonice Ribeiro**	cleoniceclr@gmail.com	silver Leaf Way # 1534	15	peabody	MA	01960	US	**6178035838**	ADCentral	Promoter Plan			$ (1,495.00)	$ 98,989.50	$ (424,325.00)	$ (215,939.33)	$ 597,503.99	$ -	$ (326,830.50)
10527337	227597	**reclapfone**	**Cleonice Ribeiro**	cleoniceclr@gmail.com	silver Leaf Way # 1534	15	peabody	MA	01960	US	**6178035838**	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
10666317	227597	cofeecity	**Cleonice Ribeiro**	niceecr@hotmail.com	Martin #2	41	Medford	MA	02155	US	6172335901	ADCentral	Promoter Plan			$ (1,495.00)	$ 149.70		$ (34,556.99)	$ 59,785.26	$ -	$ (1,345.30)
10767931	227597	**reclap3**	**Cleonice Ribeiro**	cleoniceclr@gmail.com	**silver leaf way 1534**	**15**	**Peabody**	MA	01960	US	**6178035838**	ADCentral	Promoter Plan			$ (1,495.00)	$ 449.10		$ -	$ 4,239.56	$ -	$ (1,045.90)
10805315	227597	**reclap2phone**	**Cleonice Ribeiro**	cleoniceclr@gmail.com	Martin St apt 2	41	Medford	MA	02155	US	6172335901	Partner	Phone Plan / Other			$ (399.20)			$ -	$ -	$ -	$ (399.20)
10805475	227597	cofeecityphone	**Cleonice Ribeiro**	cleoniceclr@gmail.com	Martin St Apt 2	41	Medford	MA	02155	US	6172335901	Partner	Phone Plan / Other			$ (249.50)			$ -	$ -	$ -	$ (249.50)
10830435	227597	**reclap1fone**	**Cleonice Ribeiro**	cleoniceclr@gmail.com	Martin st	41	Medford	MA	02155	US	6172335901	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
11183563	227597	**reclapphone2**	**Cleonice Ribeiro**	cleoniceclr@gmail.com	Harvard St	99	Everett	MA	02149	US	6172335901	Partner	Phone Plan / Other			$ (349.30)			$ -	$ -	$ -	$ (349.30)
11556867	227597	**reclap3phone**	**Cleonice Ribeiro**	cleoniceclr@gmail.com	Martin st	41	Medford	MA	02155	US	6172335901	Partner	Phone Plan / Other			$ (249.50)			$ -	$ -	$ -	$ (249.50)
12256941	227597	cofeecity1phone	**Cleonice Ribeiro**	cleoniceclr@gmail.com	Martin st	41	Medford	MA	02155	US	6172335901	Partner	Phone Plan / Other			$ (249.50)			$ -	$ -	$ -	$ (249.50)
18656781	227597	niceecr1	**Cleonice Ribeiro**	cleoniceclr@gmail.com	Rua Presidente Leite Figueiredo 542		Cuiaba	MA	78050000	UY	6178035848	Associate	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
10525055	$ 34,124.20	1

User Account:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (339.00)	$ 34,463.20	$ -	$ (46,994.79)	$ 221,300.88	$ -	**$ 34,124.20**
10525055	10525055	**reclap1**	**Cleonice Ribeiro**	**fredfkoch@gmail.com**	**silver Leaf Way # 1534**	15	**Peabody**		1960	US	**6178035838**	ADCentral	Promoter Plan			$ (339.00)	$ 34,463.20		$ (46,994.79)	$ 221,300.88	$ -	$ 34,124.20

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.18: #18 Top Net Loser - Parent ID 13531273 ($262,000 Net Loser) v. Parent ID 966001 ($819,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
13531273	$ (262,113.50)	213

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	Neighborhood	City	State/ Province	Postal Code	Country	Phone	Account State	Entity	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
																$ (5,700.00)	$ -	$ (191,929.50)	$ 2,076.40	$ -	$ (133,924.69)	$ 246,002.54	$ -	**$ (262,113.50)**
13531273	13531273	sirleneduarte14	Gilson Da Silva	sirleneduartetelex@gmail.com	Jackson St	33		Long Branch	NJ	07740	US	7329043516	ADCentral	TelexFree	Promoter Plan			$ (339.00)	$ 99.80		$ -	$ 4,381.41	$ -	$ (239.20)
13532285	13531273	gilsonduarte1v	Gilson Da Silva	**bigfoxnj@hotmail.com**	Jackson St	33		Long Branch	NJ	07740	US	7329043516	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
14895827	13531273	gilsonduarte1vv	Gilson Duarte	**bigfoxnj@hotmail.com**	Jackson St	33		Long Branch	NJ	07740	US	7329043516	Partner	TelexFree	Phone Plan / Other						$ -	$ -	$ -	$ (49.90)
17865375	13531273	gilsonduarte116	Gilson da Silva	gilsonduarte67@msn.com	33 Jackson St			Long Branch	NJ	07740	US	7329043516	ADCentral	TelexFree	Promoter Plan			$ (1,474.90)			$ -	$ 304.91	$ -	$ (1,474.90)
17865503	13531273	gilsonduarte117	Gilson da Silva	gilsonduarte67@msn.com	33 Jackson St			Long Branch	NJ	07740	US	7329043516	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17875367	13531273	gilsonduarte144	Gilson da Silva	gilsonduarte67@msn.com	33 Jackson St			Long branch	NJ	07740	US	7329043516	ADCentral	TelexFree	Promoter Plan			$ (1,474.90)			$ -	$ -	$ -	$ (1,474.90)
17894981	13531273	sirleneduarte44	Gilson da Silva	gilsonduarte67@msn.com	33 Jackson St			Long Branch	NJ	07740	US	7329043516	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
26114710	13531273	voipgilson04	gilson Dasilva	gilsonduarte67@msn.com	33 jackson st			long branch	NJ	07740	US	7329043516	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
966001	$ 819,345.28	250

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	Neighborhood	City	State/ Province	Postal Code	Country	Phone	Account State	Entity	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
																$ (5,600.00)	$ 28,310.68	$ (256,026.00)	$ 1,123,031.60	$ (70,371.00)	$ (1,998,657.34)	$ 1,368,282.09	$ -	**$ 819,345.28**
966001	966001	juniorcaldas	Joelito S Caldas	**bigfoxnj@hotmail.com**	120 Atlantic Ave # 9B	120	120 Atlantic Ave # 9	Long Branch	NJ	7740	US	17327492665	ADCentral	TelexFree	Promoter Plan	$ (1,375.00)	$ 22,280.94		$ 145,753.10	$ (30,370.00)	$ (193,383.13)	$ 73,573.88	$ -	$ 136,289.04
1484814	966001	juniorcaldas2	Joelito S. Caldas	**bigfoxnj@hotmail.com**	120 atlantic ave apt 9B	120		Long Branch	NJ	7740	US	17327492665	ADCentral	TelexFree	Promoter Plan			$ (1,375.00)	$ 199.60		$ -	$ 4,600.40	$ -	$ (1,175.40)
1484992	966001	juniorcaldas3	Joelito S. Caldas	**bigfoxnj@hotmail.com**	120 Atlantic ave apt 9E	120		Long Branch	NJ	7740	US	7327492665	ADCentral	TelexFree	Promoter Plan			$ (1,375.00)	$ 49.90		$ -	$ 4,850.10	$ -	$ (1,325.10)
1486372	966001	juniorcaldas10	Joelito S. Caldas	**bigfoxnj@hotmail.com**	120 Atlantic ave apt 9E	120		Long Branch	NJ	7740	US	7327492665	ADCentral	TelexFree	Promoter Plan			$ (1,375.00)	$ 28,549.90		$ (32,464.00)	$ 8,800.00	$ -	$ 27,174.90
5085890	966001	juniorcaldas13	joelito souza calda	**bigfoxnj@hotmail.com**	atlantic ave apt 9b	120		long branch		7740	US	17326186381	ADCentral	TelexFree	Promoter Plan		$ 5,649.74	$ (339.00)	$ 831,241.70	$ (40,001.00)	$ (1,302,102.68)	$ 605,485.31	$ -	$ 796,551.44
10864299	966001	voip2fabricio	Joelito S Caldas	**bigfoxnj@hotmail.com**	120			long branch	Ne	7740	US	7327492665	Partner	TelexFree	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
12256985	966001	juniorcaldas64	Joelito Souza Cald	gladiadorusa@outlook.com	Atlantic Ave #9 B	120		Long Branch	NJ	7740	US	7327492665	ADCentral	TelexFree	Promoter Plan			$ (339.00)	$ 40,818.30		$ (150,121.50)	$ 140,019.04	$ -	$ 40,479.30
27085560	966001	foxgroup03	Joelito caldas	**bigfoxnj@hotmail.com**	120 atlantic ave apt 9b			long branch	NJ	7740	US	7327492665	ADFamily	TelexFree	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
27085570	966001	foxgroup02	Joelito Caldas Jr	**bigfoxnj@hotmail.com**	120 Atlantic Ave #9 B			Long Branch	NJ	7740	US	7327492665	ADFamily	TelexFree	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
27523548	966001	jscj3	joelito caldas	**bigfoxnj@hotmail.com**	120 atlantic ave 9 B			long branch	NJ	7740	US	7327492665	ADFamily	TelexFree	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
30366236	966001	queijomineiro1	Joelito Caldas	**bigfoxnj@hotmail.com**	120 Atlantic Ave #9 B			Long Branch	NJ	7740	US	7327492665	ADFamily	TelexFree	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
31216178	966001	blacksuzuki5b	Joelito Caldas Jr	**bigfoxnj@hotmail.com**	120 Atlantic Ave #9 B			Long Branch	NJ	7740	US	7327492665	ADFamily	TelexFree	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
31733016	966001	gbs10	Joelito Caldas Jr	**bigfoxnj@hotmail.com**	120 atlantic ave apt 9B			Long Branch	NJ	7740	US	7327492665	ADFamily	TelexFree	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
31742018	966001	gbs14	Joelito Caldas Jr	**bigfoxnj@hotmail.com**	120 atlantic ave apt 9B			Long Branch	NJ	7740	US	7327492665	Associate	TelexFree	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.19: #19 Top Net Loser - Parent ID 10811451 ($242,000 Net Loser) v. Parent ID 10838191 ($296,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
10811451	**$ (241,722.60)**	**1,027**

Example of User Accounts:

														$ (14,736.00)	$ 600.00	$(306,175.10)	$ 78,588.50	$ -	$ (66,228.29)	$ 235,447.15	$ - **$ (241,722.60)**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
10922947	10811451	**sosa00010**	**oscar sosa**	oscarsosa00010@yahoo.es	**1315 nw, 98 CT suite 7**		**miami**	**fl**	**33172**	US	**0019047050817**	ADCentral	Promoter Plan			$ (1,425.00)	$ 299.40		$ -	$ 3,024.55	$ -	$ (1,125.60)
11065869	10811451	**sosa00011**	**oscar sosa**	oscarsosa00011@hotmail.es	**1315 nw 98 ct suite 7**		**miami**	**fl**	**33172**	US	**0019047050817**	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,047.90		$ -	$ 2,124.75	$ -	$ (377.10)
11291073	10811451	**sosa00012**	**oscar sosa**	**ososa_esculpi1@hotmail.com**	**1315 nw 98 ct suite 7**		**miami**	**fl**	**33172**	US	**0019047050817**	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,347.30		$ -	$ 4,770.73	$ -	$ (77.70)
11291313	10811451	**sosa00014**	**oscar sosa**	ososa_esculpi1@hotmail.com	**1315 nw 98 ct suite 7**		**miami**	**fl**	**33172**	US	**0019047050817**	ADCentral	Promoter Plan			$ (1,425.00)	$ 399.20		$ -	$ 2,424.95	$ -	$ (1,025.80)
11291357	10811451	**sosa00015**	**oscar sosa**	ososa_esculpi1@hotmail.com	**1315 nw 98 ct suite 7**		**miami**	**fl**	**33172**	US	**0019047050817**	ADCentral	Promoter Plan			$ (1,425.00)	$ 399.20		$ -	$ 2,424.95	$ -	$ (1,025.80)
13949375	10811451	**sosa107**	**oscar sosa**	**ososa_esculpi1@hotmail.com**	**1315 nw 98 ct suite 7**		**miami**	**FL**	**33172**	US	**0019047050817**	ADCentral	Promoter Plan			$ (1,425.00)	$ 6,441.50		$ (4,167.89)	$ 859.62	$ -	$ 5,016.50
18240987	10811451	dais0001	**oscar sosa**	**ososa_esculpi1@hotmail.com**	**1315 nw 98 ct suite 7**		**miami**	**FL**	1060	US	**0019047050817**	ADFamily	Promoter Plan			$ (359.50)	$ 35,158.10		$ (62,060.40)	$ -	$ -	$ 34,798.60

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
10838191	**$ 295,498.20**	**3**

User Accounts:

														$ (2,850.00)	$ 17,155.00	$ (1,425.00)	$ 282,618.20	$ -	$(273,353.88)	$ 11,102.41	$ - **$ 295,498.20**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
10838191	10838191	telexfreeccs	**Oscar Enrique Sosa Esculpi**	**ososa_esculpi1@hotmail.com**	**1315 nw 98 ct suite 7**		**miami**	**fl**	**33172**	US	**19047050817**	ADCentral	Promoter Plan	$ (1,425.00)	$ 17,155.00		$ 279,424.60		$(273,353.88)	$ 1,050.00	$ -	$ 295,154.60
10850563	10838191	**sosa0001**	**Oscar Enrique Sosa Esculpi**	**oscarsosa0001@live.com**	**1315 nw 98 ct suite 7**		**miami**	**fl**	**33172**	US	**19047050817**	ADCentral	Promoter Plan	$ (1,425.00)			$ 1,646.70		$ -	$ 6,459.20	$ -	$ 221.70
10850933	10838191	**sosa0002**	**Oscar Enrique Sosa Esculpi**	**oscarsosa0002@live.com**	**1315 nw 98 ct suite 7**		**miami**	**fl**	**33172**	US	**19047050817**	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,546.90		$ -	$ 3,593.21	$ -	$ 121.90

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.20: #20 Top Net Loser - Parent ID 29758080 ($240,000 Net Loser) v. Parent ID 10539115 ($1.7M Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
29758080	$ (240,140.60)	2,705

Example of User Accounts:

														$ -	$ -	$ (240,140.60)	$ -	$ -	$ -	$ -	$ (240,140.60)
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
29988526	29758080	04ztb0x6g41e0ad	Mgs marketing inc inc	mgsmarketinginc@gmail.com	**334 chestnut farm way**		**Rayham**	**MA**	02767	US	**7818838099**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
30018714	29758080	077nrfyj09miqhe	mgsmarketinginc.inc inc	mgsmarketinginc@gmail.com	**334 Chesnut farm way**		**rayham**	**MA**	02767	US	**7818838099**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
30147966	29758080	05zfwue4whgajl4	MGS Marketing Inc	mgsmarketinginc@gmail.com	**334 Chestnut Farm Way**	81	**Raynham**	**MA**	02767	US	**7818838099**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
30148060	29758080	043dvtzc76i9czj	MGS Marketing Inc	mgsmarketinginc@gmail.com	**334 Chestnut Farm Way**	81	**Raynham**	**MA**	02767	US	**7818838099**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
31201240	29758080	04z3j9vbzj25f3p	Mgsmarketinginc.inc sudre	mgsmarketinginc@gmail.com	**334 chestnut farm way**	111	**rayham**	**MA**	02767	US	**7818838099**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
31201922	29758080	03m542pd8mjiosc	Mgsmarketinginc.inc sudre	mgsmarketinginc@gmail.com	**334 chestnut farm way**	111	**rayham**	**MA**	02767	US	**7818838099**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
31283194	29758080	07gg10v4rsodr7n	MGS Marketing Inc	mgsmarketinginc@gmail.com	**334 Chestnut Farm Way**	99	**Raynham**	**MA**	02767	US	**7818838099**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
31283650	29758080	05e1wtkmz4moduv	MGS Marketing Inc	mgsmarketinginc@gmail.com	**334 Chestnut Farm Way**	99	**Raynham**	**MA**	02767	US	**7818838099**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
31284388	29758080	049q9yg11ra5hb3	MGS Marketing Inc	mgsmarketinginc@gmail.com	**334 Chestnut Farm Way**	99	**Raynham**	**MA**	02767	US	**7818838099**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
31667616	29758080	02ii9cjiz1t2rzw	MGS Marketing Inc	mgsmarketinginc@gmail.com	**334 Chestnut Farm Way**	142	**Raynham**	**MA**	02767	US	**7818838099**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
31676090	29758080	04bdrusvmwl0c3e	MGS Marketing Inc	mgsmarketinginc@gmail.com	**334 Chestnut Farm Way**	177	**Raynham**	**MA**	02767	US	**7818838099**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
10539115	$ 1,677,542.30	144

Example of User Accounts:

														$ -	$ 2,000.00	$ (192,698.90)	$ 1,987,141.20	$(118,900.00)	$(4,627,521.36)	$ 3,422,430.33	$ -	$ 1,677,542.30
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
10885105	10539115	dady2013	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 Chestnut Farm Way**	334	**Raynham**	**Ma**	02767	US	**7818838099**	Partner	Promoter Plan			$ (50.00)			$ -	$ -	$ -	$ (50.00)
10539115	10539115	mariasudre5	euzebio sudre neto	mariasudre@yahoo.com	**334 Chestnut Farm Way**	334	**raynham**		02767	US	5087251309	ADCentral	Promoter Plan		$ 2,000.00	$ (1,425.00)	$ 1,489,353.30	$(118,900.00)	$(3,952,967.22)	$ 2,813,021.19	$ -	$ 1,371,028.30
12140735	10539115	gruporns13	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 chestnut farm way**	334	**Raynham**		02767	US	**7818838099**	ADCentral	Promoter Plan			$ (1,425.00)	$ 168,537.80		$ (196,191.58)	$ 40,931.00	$ -	$ 167,112.80
12140867	10539115	gruporns14	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 chestnut farm way**	334	**Raynham**	**Ma**	02767	US	**7818838099**	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,047.90		$ -	$ 6,411.57	$ -	$ (377.10)
12140879	10539115	gruporns15	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 chestnut farm way**	334	**Raynham**	**Ma**	02767	US	**7818838099**	ADCentral	Promoter Plan			$ (1,425.00)	$ 998.00		$ -	$ 5,458.57	$ -	$ (427.00)
13664669	10539115	gruporns105	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 chestnut farm way**	334	**Raynham**		02767	US	**7818838099**	ADCentral	Promoter Plan			$ (1,425.00)	$ 499.00		$ -	$ 800.00	$ -	$ (926.00)
10552959	10539115	mariasudre7	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 chestnut farm way**	334	**Raynham**	**Ma**	02767	US	5087251309	ADCentral	Promoter Plan			$ (1,425.00)	$ 83,171.10		$ (161,739.94)	$ 91,242.20	$ -	$ 81,746.10
13664691	10539115	gruporns106	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 chestnut farm way**	334	**Raynham**		02767	US	**7818838099**	ADCentral	Promoter Plan			$ (1,425.00)	$ 499.00		$ -	$ 800.00	$ -	$ (926.00)
13664759	10539115	gruporns107	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 chestnut farm way**	334	**Raynham**	**MA**	02767	US	**7818838099**	ADCentral	Promoter Plan			$ (1,425.00)	$ 499.00		$ -	$ 800.00	$ -	$ (926.00)
12794563	10539115	mgts	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 chestnut farm way**	334	**Raynham**		02767	US	**7818838099**	ADCentral	Promoter Plan			$ (1,425.00)	$ 149,092.50		$ (195,565.39)	$ 57,152.00	$ -	$ 147,667.50
13842987	10539115	drinds2014	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 chestnut farm way**	334	**Raynham**	**MA**	02767	US	**7818838099**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,100.00	$ -	$ (1,425.00)
14585613	10539115	gruporns129	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 chestnut farm way**	334	**Raynham**	**MA**	02767	US	**7818838099**	ADCentral	Promoter Plan			$ (1,425.00)	$ 499.00		$ -	$ 500.00	$ -	$ (926.00)
14585649	10539115	gruporns130	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 chestnut farm way**	334	**Raynham**	**MA**	02767	US	**7818838099**	ADCentral	Promoter Plan			$ (1,425.00)	$ 499.00		$ -	$ 500.00	$ -	$ (926.00)
17682955	10539115	gruporns160	Euzebio Sudre Neto	mariasudre@yahoo.com	**334 Chestnut Farm Way**	.	**Raynham**	**MA**	02767	US	**17818838099**	ADCentral	Promoter Plan			$ (1,425.00)	$ 249.50		$ (8,545.00)	$ 10,929.00	$ -	$ (1,175.50)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.21: #21 Top Net Loser - Parent ID 17439659 ($235,000 Net Loser) v. Parent ID 17248041 ($128,000 Net Winner), Parent ID 5108450 ($121,000 Net Winner) and Parent ID 17957689 ($76,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
17439659	$ (235,125.00)	165

Example of User Accounts:

														$ -	$ -	$ (235,125.00)	$ -	$ -	$ -	$ 27,642.97	$ -	**$ (235,125.00)**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
17447393	17439659	linjie10	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 524.91	$ -	$ (1,425.00)
17448493	17439659	linjie11	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 524.91	$ -	$ (1,425.00)
17448567	17439659	linjie12	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 524.91	$ -	$ (1,425.00)
17449393	17439659	linjie13	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 524.91	$ -	$ (1,425.00)
17449459	17439659	linjie14	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17543383	17439659	linjie15	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 304.91	$ -	$ (1,425.00)
17543453	17439659	linjie16	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 304.91	$ -	$ (1,425.00)
17544481	17439659	linjie17	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 304.91	$ -	$ (1,425.00)
17554333	17439659	npnp31	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17554387	17439659	npnp32	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17615633	17439659	npnp33	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17615665	17439659	npnp34	cincin zhao	telexchina668@gmail.com	24607 64th ave.		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17620333	17439659	npnp35	cincin zhao	telexchina668@gmail.com	24607 64ave		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17620451	17439659	npnp36	cincin zhao	telexchina668@gmail.com	24607 64ave		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17621099	17439659	npnp37	cincin zhao	telexchina668@gmail.com	24607 64ave		little neck	NY	11362	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
17248041	$ 128,244.30	24

Example of User Accounts:

														$ -	$ -	$ (34,200.00)	$ 162,444.30	$ -	$ (167,700.24)	$ 27,318.11	$ -	**$ 128,244.30**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
17248041	17248041	okko1	Xiao Ling Xu	telexasian168@gmail.com	11 feng road		guangzhou	44	133298	CN	6617480241	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17248229	17248041	okkoo	Xiao Ling Xu	telexasian168@gmail.com	11 feng road		guangzhou	44	133298	CN	6617480241	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17248505	17248041	okkoo1	Xiao Ling Xu	telexasian168@gmail.com	11 feng road		guangzhou	44	133298	CN	6617480241	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17401937	17248041	chairs	Xiao Ling Xu	telexasian168@gmail.com	7 Rivington St		Washington	37	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17402229	17248041	chairs1	Xiao Ling Xu	telexasian168@gmail.com	7 Rivington St		Washington	44	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17403385	17248041	notes	Xiao Ling Xu	telexasian168@gmail.com	7 Rivington St		Washington	GA	11366	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17404217	17248041	127127a	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	GA	11366	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)	$ 162,444.30		$ (167,700.24)	$ 11,220.00	$ -	$ 161,019.30
17404529	17248041	npnp	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	GA	11366	US	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,304.91	$ -	$ (1,425.00)
17404579	17248041	linjie	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	CYI	11366	TW	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,724.91	$ -	$ (1,425.00)
17404981	17248041	npnp01	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	44	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,504.91	$ -	$ (1,425.00)
17405009	17248041	npnp02	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	44	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,724.10	$ -	$ (1,425.00)
17405157	17248041	linjie01	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	31	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,724.91	$ -	$ (1,425.00)
17405211	17248041	linjie02	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	CAL	11366	CO	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17405443	17248041	npnp03	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	44	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 904.91	$ -	$ (1,425.00)
17405587	17248041	npnp04	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	44	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 684.91	$ -	$ (1,425.00)
17405883	17248041	npnp05	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	AT	11366	CL	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 904.91	$ -	$ (1,425.00)
17405965	17248041	linjie03	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	37	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 904.91	$ -	$ (1,425.00)
17406013	17248041	linjie04	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	44	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 904.91	$ -	$ (1,425.00)
17406453	17248041	npnp06	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	44	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 904.91	$ -	$ (1,425.00)
17406769	17248041	linjie05	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	44	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 904.91	$ -	$ (1,425.00)
17406819	17248041	linjie06	Xiao Ling Xu	telexchina668@gmail.com	7 Rivington St		Washington	44	11366	CN	9172885485	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.21: #21 Top Net Loser - Parent ID 17439659 ($235,000 Net Loser) v. Parent ID 17248041 ($128,000 Net Winner), Parent ID 5108450 ($121,000 Net Winner) and Parent ID 17957689 ($76,000 Net Winner)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
5108450	**$ 120,729.50**	**40**

Example of User Accounts:

													$ -	$ -	$ (35,192.20)	$ 155,921.70	$ -	$ (422,743.00)	$ 328,361.51	$ -	**$ 120,729.50**
Rep. Id	**Parent Id**	**Login**	**Name**	**Email**	**Street**	**Address Number**	**City**	**State/ Province**	**Postal Code**	**Country**	**Phone**	**Account State**	**Participant Type**	**1 - Direct Payment Two Year**	**2 - Direct Receipt Two Year**	**3 - Triangular Payment Two Year**	**4 - Triangular Receipt Two Year**	**5 - Purchase of Credits Two Year**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**8 - Chargebacks**	**Net Equity**
17030829	5108450	hkg061	Jian Chen	**telexasian168@gmail.com**	145 Birch Drive		Manhasset Hills	NY	10014	US	**19172885485**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
17030907	5108450	hkg062	Jian Chen	**telexasian168@gmail.com**	145 Birch Drive		Manhasset Hills	NY	10014	US	**19172885485**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
20636345	5108450	newyorkc	Jenny Chen	bidifychen@gmail.com	7 Rivington St		New York	NY	10002	US	16466218305	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
20824631	5108450	newdollar	Jenny Chen	bidifychen@gmail.com	7606 169th street	NULL	Flushing	NY	11366	US	9172885485	ADFamily	Promoter Plan			$ (339.60)	$ 3,445.90		$ (320,920.00)	$ 225,025.00	$ -	$ 3,106.30
20827997	5108450	dollarb	Jenny Chen	telexlin@gmail.com	**11 dong feng nan**	NULL	beijing	11	11233	CN	16466218305	ADFamily	Promoter Plan			$ (339.60)	$ 150,898.40		$ (100,848.00)	$ 27,954.90	$ -	$ 150,558.80

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
17957689	**$ 75,737.00**	**8**

Example of User Accounts:

													$ -	**$ -**	**$ (11,400.00)**	**$ 87,137.00**	**$ -**	**$ (98,339.57)**	**$ 7,004.55**	**$ -**	**$ 75,737.00**
Rep. Id	**Parent Id**	**Login**	**Name**	**Email**	**Street**	**Address Number**	**City**	**State/ Province**	**Postal Code**	**Country**	**Phone**	**Account State**	**Participant Type**	**1 - Direct Payment Two Year**	**2 - Direct Receipt Two Year**	**3 - Triangular Payment Two Year**	**4 - Triangular Receipt Two Year**	**5 - Purchase of Credits Two Year**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**8 - Chargebacks**	**Net Equity**
17957689	17957689	killermoney	Jean Lau	**telexasian168@gmail.com**	**11 feng road**		guangzhou	44	11354	CN	**19172885485**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17959317	17957689	moneyspot	Jean Lau	**telexasian168@gmail.com**	**11 feng road**		guangzhou	44	11354	CN	**19172885485**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17960877	17957689	moneyspot2	Jean Lau	**telexasian168@gmail.com**	**11 feng road**		guangzhou	44	11354	CN	**19172885485**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17961679	17957689	dollarmoney	Jean Lau	**telexasian168@gmail.com**	**11 feng road**		guangzhou	KHQ	11354	TW	**19172885485**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17962547	17957689	dollarmoney2	Jean Lau	**telexasian168@gmail.com**	**11 feng road**		guangzhou	HSQ	11354	TW	**19172885485**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 7,004.55	$ -	$ (1,425.00)
17965893	17957689	127dm	Jean Lau	**telexasian168@gmail.com**	**11 feng road**		guangzhou	ILA	11354	TW	**19172885485**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17967135	17957689	moneyfood	Jean Lau	**telexasian168@gmail.com**	**11 feng road**		guangzhou	TPQ	11354	TW	**19172885485**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17969837	17957689	moneywater	Jean Lau	**telexasian168@gmail.com**	**11 feng road**		guangzhou	TPQ	11354	TW	**19172885485**	ADCentral	Promoter Plan			$ (1,425.00)	$ 87,137.00		$ (98,339.57)	$ -	$ -	$ 85,712.00

Source(s): Based upon queries of the Huron Data Tables

Exhibit 9.22: #22 Top Net Loser - Parent ID 11487123 ($231,000 Net Loser) v. Parent ID 11482181 ($53,000 Net Winner) and Parent ID 10594737 ($225,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
11487123	$ (230,922.00)	383

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Entity	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
															$ -	$ -	$ (300,323.70)	$ 69,401.70	$ -	$ (54,229.20)	$ 74,593.45	$ -	**$ (230,922.00)**
12290113	11487123	barry1300	BARRY MILLER	**expositoresorla@gmail.com**	3190 n atlantic ave- unit 115	3190	Cocoa beach	FLO	32931	US	3522052709	Partner	TelexFree	Phone Plan / Other			$ (249.50)			$ -	$ -	$ -	$ (249.50)
13490427	11487123	barry2400	BARRY MILER	**expositoresorla@gmail.com**	**1310 PARTRIDGE CLOSE**	1310	POMPANO BEACH	FL	33064	US	3522952709	Partner	TelexFree	Phone Plan / Other			$ (199.60)			$ -	$ -	$ -	$ (199.60)
23079387	11487123	bmiller164	Barry Miller	**thiaguimts@hotmail.com**	N Atlantic Ave	3190	Cocoa Beach	FL	32931	US	3522092709	ADFamily	TelexFree	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
23230289	11487123	bmiller166cl	Barry Miller	**thiaguimts@hotmail.com**	N Atlantic Ave.	3190	Cocoa Beach	FL	32931	US	3522092709	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
23230419	11487123	bmiller166cl1	Barry Miller	**thiaguimts@hotmail.com**	N Atlantic Ave.	3190	Cocoa Beach	FL	32931	US	3522092709	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
23230589	11487123	bmiller166cl2	Barry Miller	**thiaguimts@hotmail.com**	N Atlantic Ave.	3190	Cocoa Beach	FL	32931	US	3522092709	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
23230735	11487123	bmiller166cl3	Barry Miller	**thiaguimts@hotmail.com**	N Atlantic Ave.	3190	Cocoa Beach	FL	32931	US	3522092709	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
23230917	11487123	bmiller166cl4	Barry Miller	**thiaguimts@hotmail.com**	N Atlantic Ave.	3190	Cocoa Beach	FL	32931	US	3522092709	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
23231127	11487123	bmiller166cl5	Barry Miller	**thiaguimts@hotmail.com**	N Atlantic Ave.	3190	Cocoa Beach	FL	32931	US	3522092709	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
23231279	11487123	bmiller166cl6	Barry Miller	**thiaguimts@hotmail.com**	N Atlantic Ave.	3190	Cocoa Beach	FL	32931	US	3522092709	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
23231441	11487123	bmiller166cl7	Barry Miller	**thiaguimts@hotmail.com**	N Atlantic Ave.	3190	Cocoa Beach	FL	32931	US	3522092709	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
23231611	11487123	bmiller166cl8	Barry Miller	**thiaguimts@hotmail.com**	N Atlantic Ave.	3190	Cocoa Beach	FL	32931	US	3522092709	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
23231801	11487123	bmiller166cl9	Barry Miller	**thiaguimts@hotmail.com**	N Atlantic Ave.	3190	Cocoa Beach	FL	32931	US	3522092709	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
11482181	$ 53,340.10	47

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Entity	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
															$ 3,470.80	$ (24,322.60)	$ 74,191.90	$ -	$ -	$ (83,667.64)	$ 46,885.55	$ -	**$ 53,340.10**
11482181	11482181	udson00	udson nascimento vieira	udsontelex83@gmail.com	165 jefferson drive	165	norwood		2062	US	7812671718	ADCentral	TelexFree	Promoter Plan		$ (1,425.00)	$ 10,953.60			$ (11,721.15)	$ 3,675.27	$ -	$ 9,528.60
11482397	11482181	udson001	udson nascimento vieira	udsontelex83@gmail.com	165 jefferson drive	165	norwood		2062	US	7812671718	ADCentral	TelexFree	Promoter Plan		$ (1,425.00)	$ 14,948.60			$ (16,207.00)	$ 3,595.20	$ -	$ 13,523.60
13749577	11482181	fialhom	Udson Vieira	udsontelex83@gmail.com	jefferson dr	165	norwood		2062	US	7812671718	ADCentral	TelexFree	Promoter Plan	$ 3,470.80	$ (1,425.00)	$ 44,152.00			$ (47,827.18)	$ 16,868.91	$ -	$ 46,197.80
13749747	11482181	fialhomcl	Udson Vieira	**thiaguimts@hotmail.com**	jefferson dr.	165	norwood	MA	2062	US	7812671718	Partner	TelexFree	Phone Plan / Other		$ (199.60)				$ -	$ -	$ -	$ (199.60)
13750167	11482181	fialhom1cl	Udson Vieira	**thiaguimts@hotmail.com**	jefferson dr.	165	norwood	MA	2062	US	7812671718	Partner	TelexFree	Phone Plan / Other		$ (199.60)				$ -	$ -	$ -	$ (199.60)
13750433	11482181	fialhom2cl	Udson Vieira	**thiaguimts@hotmail.com**	jefferson dr.	165	norwood	MA	2062	US	7812671718	Partner	TelexFree	Phone Plan / Other		$ (199.60)				$ -	$ -	$ -	$ (199.60)
31035454	11482181	5viec10	udson vieira	udsontelex83@gmail.com	165 jefferson drive	1	normood	MA	2062	US	7812671718	Partner	TelexFree	Phone Plan / Other		$ (49.90)				$ -	$ -	$ -	$ (49.90)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
10594737	$ 224,692.60	95

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Entity	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
															$ -	$ -	$ (68,655.40)	$ 293,348.00	$ -	$ (373,418.83)	$ 308,411.66	$ -	**$ 224,692.60**
10594737	10594737	rrpoba204	NILTON S VIEIRA	gessosudoeste@gmail.com	326 BRODWAY	326	SAMERVILLE	MA	2145	US	9784739622	ADCentral	TelexFree	Promoter Plan			$ (339.00)	$ 50,148.50		$ (48,829.16)	$ 27,101.07	$ -	$ 49,809.50
10594939	10594737	edil200	NILTON S VIEIRA	tontelex@yahoo.com.br	326 BROADWAY	326	SAMERVILLE	MA	2145	US	9784739622	ADCentral	TelexFree	Promoter Plan			$ (339.00)	$ 21,774.20		$ (34,705.31)	$ 19,061.48	$ -	$ 21,435.20
10669279	10594737	ton202	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	326 BROADWAY	326	SAMERVILLE	MA	2145	US	9784739622	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)			$ -	$ 400.00	$ -	$ (1,425.00)
10669295	10594737	ton203	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	326 BROADWAY	326	SAMERVILLE	MA	2145	US	9784739622	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
10669703	10594737	wendyd1	NILTON SILVA VIEIRA	edilsonld1317@hotmail.com	326 BROADWAY	326	SAMERVILLE	MA	2145	US	9547093979	ADCentral	TelexFree	Promoter Plan			$ (339.00)	$ 33,555.60		$ (54,829.18)	$ 33,627.29	$ -	$ 33,216.60
10670675	10594737	andba100	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	a	a	a	ma	2145	US	9547093979	Partner	TelexFree	Phone Plan / Other			$ (199.60)			$ -	$ -	$ -	$ (199.60)
10670959	10594737	wendyd2	NILTON SILVA VIEIRA	edilsonld1317@hotmail.com	326 BROADWAY	326	SAMERVILLE	MA	2145	US	9547093979	ADCentral	TelexFree	Promoter Plan			$ (339.00)	$ 67,585.50		$ (64,077.69)	$ 3,891.07	$ -	$ 67,246.50
10670997	10594737	wendyd100	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	expositoresorla@gmail.com	a	a	a	2145	US	9547093979	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10671297	10594737	wendyd200	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	a	a	a	ma	2145	US	9547093979	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10699043	10594737	wendyd400	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	326 BROADWAY	326	SAMERVILLE	MA	2145	US	9547093979	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10699537	10594737	wendyd600	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	326 BROADWAY	326	SAMERVILLE	MA	2145	US	9547093979	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10719673	10594737	wendyd800	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	**1310 PARTRIDGE CLOSE**	a	a	a	33064	US	9547093979	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10719743	10594737	wendyd900	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	326 BROADWAY	326	SAMERVILLE	MA	2145	US	9547093979	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10896185	10594737	rrpoba2050	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	**1310 PARTRIDGE CLOSE**	1310	Weston	FL	33064	US	9547093979	Partner	TelexFree	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
10896419	10594737	ferraz206	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	**1310 PARTRIDGE CLOSE**	1310	POMPANO BEACH	FL	33064	US	9547093979	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 399.20		$ -	$ 1,500.60	$ -	$ (1,025.80)
10896435	10594737	ferraz207	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	**1310 PARTRIDGE CLOSE**	1310	POMPANO BEACH		33064	US	9547093979	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 99.80		$ (610.00)	$ 2,719.11	$ -	$ (1,325.20)
10896553	10594737	ferraz2040	NILTON SILVA VIEIRA	**expositoresorla@gmail.com**	**1310 PARTRIDGE CLOSE**	1310	POMPANO BEACH	FL	33064	US	9547093979	Partner	TelexFree	Phone Plan / Other			$ (299.40)			$ -	$ -	$ -	$ (299.40)
10954781	10594737	rrpoba206	NILTON SILVA VIEIRA	gessosudoeste@hotmail.com	**1310 PARTRIDGE CLOSE**	1310	POMPANO BEACH	FL	33064	US	9546874023	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 26,997.50		$ (30,662.83)	$ 8,148.42	$ -	$ 25,572.50
11750339	10594737	theo1113	NILTON SILVA VIEIRA	niltonld@hotmail.com	326 BROADWAY	326	ma	ma	2145	US	9547093979	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Source(s): Based upon queries of the Huron Data Tables

Exhibit 9.23: #23 Top Net Loser - Parent ID 14473091 ($226,000 Net Loser) v. Parent ID 13557817 ($368,000 Net Winner) and Parent ID 13559825 ($53,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
14473091	$ (226,402.40)	213

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Entity	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two	4 - Triangular Receipt Two	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
															$ -	$ -	$ (289,625.70)	$ 63,223.30	$ -	$ (16,976.33)	$ 112,397.27	$ -	**$ (226,402.40)**
14483747	14473091	telexinove021	Jonatas Vasconcelos	arthurbrazsoares@gmail.com	**augusto correia de melo**	**331**	catende	MA	55400000	US	08199386573	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 598.80		$ -	$ 624.95	$ -	$ (826.20)
14483773	14473091	telexinove022	Jonatas Vasconcelos	arthurbrazsoares@gmail.com	**augusto correia de melo**	**331**	catende	MA	55400000	US	08199386573	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 1,218.40		$ -	$ 2,204.95	$ -	$ (206.60)
16553369	14473091	101telexinove	Jonatas Vasconcelos	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**08199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 300.00	$ -	$ (1,325.20)
16553391	14473091	102telexinove	Jonatas Vasconcelos	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**08199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 300.00	$ -	$ (1,325.20)
16553489	14473091	103telexinove	Jonatas Vasconcelos	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**08199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 300.00	$ -	$ (1,325.20)
16553523	14473091	104telexinove	Jonatas Vasconcelos	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**08199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 300.00	$ -	$ (1,325.20)
16553575	14473091	105telexinove	Jonatas Vasconcelos	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**08199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
16553615	14473091	106telexinove	Jonatas Vasconcelos	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**08199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
16553651	14473091	107telexinove	Jonatas Vasconcelos	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**08199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
16553685	14473091	108telexinove	Jonatas Vasconcelos	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**08199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 300.00	$ -	$ (1,325.20)
17884897	14473091	telexinove169	Jonatas Vasconcelos	arthurbrazsoares@gmail.com	**Av augusto correia de melo**		Catende	06	55400000	PT	08199386573	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
13557817	$ 368,173.10	6

User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Entity	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two	4 - Triangular Receipt Two	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
															$ -	$ -	$ (7,174.90)	$ 375,348.00	$ -	$ (359,204.70)	$ 12,563.97	$ -	**$ 368,173.10**
13557817	13557817	rildobrazi9pe	Rildo Braz	**arthurbrazsoares@gmail.com**	Av. Juscelino kubitschek	60	Catende	MA	55400000	US	558132287097	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 372,952.80		$ (359,204.70)	$ 8,863.37	$ -	$ 371,527.80
13558279	13557817	rildope	Rildo Braz	**arthurbrazsoares@gmail.com**	Av. Juscelino kubitschek	60	Catende	MA	55400000	US	558132287097	Partner	TelexFree	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
13823823	13557817	**6pernambucoi9**	Rildo Braz	**arthurbrazsoares@gmail.com**	**Av. Augusto Correia de Melo**	**331**	Catende	NULL	55400000	US	8199386574	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 598.80		$ -	$ 925.15	$ -	$ (826.20)
13849453	13557817	**7pernambucoi9**	Rildo Braz	**arthurbrazsoares@gmail.com**	Av. Juscelino kutischek	60	Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 598.80		$ -	$ 925.15	$ -	$ (826.20)
13849587	13557817	**8pernambucoi9**	Rildo Braz	**arthurbrazsoares@gmail.com**	Av. Juscelino kutischek	60	Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 598.80		$ -	$ 925.15	$ -	$ (826.20)
13849623	13557817	**9pernambucoi9**	Rildo Braz	**arthurbrazsoares@gmail.com**	Av. Juscelino kutischek	60	Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 598.80		$ -	$ 925.15	$ -	$ (826.20)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
13559825	$ 52,556.30	65

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Entity	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
															$ -	$ -	$ (92,625.00)	$ 145,181.30	$ -	$ (121,267.89)	$ 40,235.19	$ -	**$ 52,556.30**
13560247	13559825	3pernambucoi9	Solange Braz	**arthurbrazsoares@gmail.com**	**Av. Augusto correias e melo**	**331**	Catende	MA	55400000	US	558136732345	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 114,122.60		$ (121,267.89)	$ 8,749.04	$ -	$ 112,697.60
14756395	13559825	**100pernambucoi9**	Solange Braz	**arthurbrazsoares@gmail.com**	**Augustos Correia de Melo**	**331**	Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 648.70		$ -	$ 2,179.76	$ -	$ (776.30)
16098945	13559825	**151pernambucoi9**	Solange Braz	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 399.20		$ -	$ 300.00	$ -	$ (1,025.80)
16098973	13559825	**152pernambucoi9**	Solange Braz	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 299.40		$ -	$ 300.00	$ -	$ (1,125.60)
16099055	13559825	**153pernambucoi9**	Solange Braz	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 399.20		$ -	$ 300.00	$ -	$ (1,025.80)
16099079	13559825	**154pernambucoi9**	Solange Braz	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 439.40		$ -	$ 300.00	$ -	$ (985.60)
16099113	13559825	**155pernambucoi9**	Solange Braz	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 439.40		$ -	$ 380.00	$ -	$ (985.60)
16099133	13559825	**156pernambucoi9**	Solange Braz	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 300.00	$ -	$ (1,325.20)
16099181	13559825	**157pernambucoi9**	Solange Braz	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 300.00	$ -	$ (1,325.20)
16099199	13559825	**158pernambucoi9**	Solange Braz	**arthurbrazsoares@gmail.com**	**Av augusto correia de melo**		Catende	MA	55400000	US	**8199386573**	ADCentral	TelexFree	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 300.00	$ -	$ (1,325.20)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.24: #24 Top Net Loser - Parent ID 15486069 ($225,000 Net Loser) v. Parent ID11583717 ($4,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
15486069	$ (224,723.00)	170

Example of User Accounts:

													$ -	$ -	$(227,672.80)	$ -	$ -	$ (10,144.00)	$ 31,031.19	$ -	**$ (224,723.00)**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
16260111	15486069	aide013	EDIVALDO REIS	**heavenlyperfumes@live.com**	9 blacksmith way		Saugus	MA	01906	US	6175907401	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 400.00	$ -	$ (1,425.00)
16260229	15486069	aide014	EDIVALDO REIS	**heavenlyperfumes@live.com**	9 blacksmith way		Saugus	MA	01906	US	6175907401	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 400.00	$ -	$ (1,425.00)
16261717	15486069	cookiea002	EDIVALDO REIS	**heavenlyperfumes@live.com**	9 blacksmith way		Saugus	MA	01906	US	6175907401	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
16289845	15486069	aide021	EDIVALDO REIS	**heavenlyperfumes@live.com**	9 blacksmith way		Saugus	MA	01906	US	6175907401	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
17773431	15486069	aide022	EDIVALDO REIS	**heavenlyperfumes@live.com**	9 blacksmith way		Saugus	FL	01906	US	6175907401	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17773723	15486069	aide023	EDIVALDO REIS	**heavenlyperfumes@live.com**	9 blacksmith way		Saugus	FL	01906	US	6175907401	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17849181	15486069	aide025	EDIVALDO REIS	**heavenlyperfumes@live.com**	9 blacksmith way		Saugus	FL	01906	US	6175907401	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
11583717	$ 3,628.30	2

User Accounts:

														$ -	$ -	$ (1,544.90)	$ 5,173.20	$ -	$ (9,155.92)	$ 11,963.00	$ -	**$ 3,628.30**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
11583717	11583717	neyvera	WSGV Services Inc	**heavenlyperfumes@live.com**	West Lake Drive	4948	Deerfield Beach		33442	US	7542350408	ADCentral	Promoter Plan			$ (1,495.00)	$ 5,173.20		$ (9,155.92)	$ 11,963.00	$ -	$ 3,678.20
11583907	11583717	neyvera1	WSGV Services Inc	neyflorida2@hotmail.com	West Lake Drive	4948	Deerfield Beach	FL	33442	US	7542350408	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.25: #25 Top Net Loser - Parent ID 12112907 ($222,000 Net Loser) v. Parent ID 4987334 ($948,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
12112907	$ (221,879.50)	264

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (357,569.60)	$ 135,690.10	$ -	$ (556,489.71)	$ 936,884.10	$ -	**$ (221,879.50)**
12241123	12112907	art059c1	Guixiang Sun	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	New	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (199.60)			$ -	$ -	$ -	$ (199.60)
12756495	12112907	gee007	Guixiang Sun	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	New	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 2,554.53	$ -	$ (1,325.20)
12756527	12112907	gee008	Guixiang Sun	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	New	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 2,556.93	$ -	$ (1,325.20)
12756643	12112907	gee004c1	Guixiang Sun	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	New	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (149.70)			$ -	$ -	$ -	$ (149.70)
12756709	12112907	gee009	Guixiang Sun	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	New	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 2,556.93	$ -	$ (1,325.20)
12756771	12112907	gee010	Guixiang Sun	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	New	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 2,556.13	$ -	$ (1,325.20)
12756929	12112907	gee011	Guixiang Sun	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	New	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 2,556.13	$ -	$ (1,325.20)
12756963	12112907	gee012	Guixiang Sun	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	New	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 2,554.53	$ -	$ (1,325.20)
12757171	12112907	gee013	Guixiang Sun	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	New	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 2,536.93	$ -	$ (1,325.20)
12757209	12112907	gee014	Guixiang Sun	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	New	11355	US	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 99.80		$ -	$ 2,558.13	$ -	$ (1,325.20)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
4987334	$ 947,923.10	211

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (13,251.30)	$ 961,174.40	$ -	$ (1,038,550.37)	$ 132,966.90	$ -	**$ 947,923.10**
4987334	4987334	dandong	Peng Zhang	**james3593@gmail.com**	41 Si Wei Rd Zhen Xing Qu		Dandong	LN	118000	CN	**3478228765**	ADCentral	Promoter Plan			$ (1,425.00)	$ 961,174.40		$ (1,038,550.37)	$ 132,966.90	$ -	$ 959,749.40
16537977	4987334	xml138c1	Peng Zhang	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
17751261	4987334	cyou030c1	Peng Zhang	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17777449	4987334	cyou050c1	Peng Zhang	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17777977	4987334	cyou051c1	Peng Zhang	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17778781	4987334	cyou053c1	Peng Zhang	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17779293	4987334	cyou054c1	Peng Zhang	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17779711	4987334	cyou055c1	Peng Zhang	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17780109	4987334	cyou056c1	Peng Zhang	**james3593@gmail.com**	**4345 Byrd St**		**Flushing**	NY	11355	US	**3478228765**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Source(s): Based upon queries of the Huron Data Tables

Exhibit 9.26: #26 Top Net Loser - Parent ID 10572419 ($215,000 Net Loser) v. Parent ID 10572045 ($107,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
10572419	$ (214,739.30)	169

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$(115,425.00)	$ 16,600.00	$(116,263.60)	$ 349.30	$ -	$ (22,261.73)	$ 74,500.20	$ -	**$ (214,739.30)**
10572419	10572419	**bb333999002**	Brahim Boulbahaiem	bb33399@hotmail.com	**Beskenstraat 101**		**Zonhoven**	Li	**3520**	BE	**0032475954337**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,100.00	$ -	$ (1,425.00)
18155561	10572419	**bb333999ao16**	Brahim Boulbahaiem	telexfreemousab@hotmail.com	**Beskenstraat 101 zonhoven**		Hasselt	VLI	**3520**	BE	**0032475954337**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
18156007	10572419	**bb333999ao15**	brahim boulbahaiem	telexfreemousab@hotmail.com	**beskensstraat**	**101**	hasselt	VLI	**3520**	BE	**0032475954337**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
18157423	10572419	**bb333999ao18**	Brahim Boulbahaiem	telexfreemousab@hotmail.com	**Beskenstraat 101 zonhoven**		Hasselt	VLI	**3520**	BE	**0032475954337**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
10572045	$ 106,856.90	28

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (39,900.00)	$ -	$ (99.80)	$ 146,856.70	$ -	$(183,608.34)	$ 110,380.94	$ -	**$ 106,856.90**
10572045	10572045	**bb333999**	Innovatech bvba	bb@ds4e.net	**Beskensstraat 101**	**3520**	**Zonhoven**	Zo	3520	BE	**32475954337**	ADCentral	Promoter Plan	$ (1,425.00)		$ (49.90)	$ 146,657.10		$(183,608.34)	$ 45,623.00	$ -	$ 145,182.20
10572207	10572045	**bb333999001**	Innovatech bvba	bb@ds4e.net	**Beskensstraat 101**	**3520**	**Zonhoven**	Zo	3520	BE	**32475954337**	ADCentral	Promoter Plan	$ (1,425.00)		$ (49.90)	$ 199.60		$ -	$ 5,457.94	$ -	$ (1,275.30)
10644541	10572045	**bb333999aoo4**	Innovatech bvba	bb@ds4e.net	**Beskensstraat 101**	**101**	**Zonhoven**	Zo	3520	BE	**32475954337**	ADCentral	Promoter Plan	$ (1,425.00)					$ -	$ 2,200.00	$ -	$ (1,425.00)
10644545	10572045	**bb333999aoo5**	Innovatech bvba	bb@ds4e.net	**Beskensstraat 101**	**101**	**Zonhoven**	Zo	3520	BE	**32475954337**	ADCentral	Promoter Plan	$ (1,425.00)					$ -	$ 1,900.00	$ -	$ (1,425.00)
10644549	10572045	**bb333999aoo6**	Innovatech bvba	bb@ds4e.net	**Beskensstraat 101**	**101**	**Zonhoven**	Zo	3520	BE	**32475954337**	ADCentral	Promoter Plan	$ (1,425.00)					$ -	$ 1,900.00	$ -	$ (1,425.00)
10644555	10572045	**bb333999aoo7**	Innovatech bvba	bb@ds4e.net	**Beskensstraat 101**	**101**	**Zonhoven**	Zo	3520	BE	**32475954337**	ADCentral	Promoter Plan	$ (1,425.00)					$ -	$ 2,200.00	$ -	$ (1,425.00)

Source(s): Based upon queries of the Huron Data Tables

Exhibit 9.27: #27 Top Net Loser - Parent ID 14745823 ($210,000 Net Loser) v. Parent ID 10396601 ($289,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
14745823	$ (210,408.00)	193

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$(217,564.50)	$ 7,156.50	$ -	$ -	$ 111,700.82	$ -	**$ (210,408.00)**
17161037	14745823	annew100	anne wang	**telexfree988@yahoo.com**	21205 haidian		beijing	11	010010	CN	13952158874	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
17161871	14745823	annew101	anne wang	**telexfree988@yahoo.com**	21205 haidian		beijing	34	010010	CN	13952158874	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17162027	14745823	annew102	anne wang	**telexfree988@yahoo.com**	21205 haidian		beijing	11	010010	CN	13952158874	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
17162597	14745823	annew103	anne wang	**telexfree988@yahoo.com**	21205 haidian		beijing	11	010010	CN	13952158874	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
17162671	14745823	annew104	anne wang	**telexfree988@yahoo.com**	21205 haidian		beijing	11	010010	CN	13952158874	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
17165733	14745823	annew105	anne wang	**telexfree988@yahoo.com**	21205 haidian		beijing	11	010010	CN	13952158874	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
17165797	14745823	annew106	anne wang	**telexfree988@yahoo.com**	21205 haidian		beijing	11	010010	CN	13952158874	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17166769	14745823	annew107	anne wang	**telexfree988@yahoo.com**	21205 haidian		beijing	11	010010	CN	13952158874	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
10396601	$ 288,064.20	50

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (47,496.20)	$ 335,560.40	$ -	$(547,328.84)	$ 343,401.70	$ -	**$ 288,064.20**
10396601	10396601	susan888	Susan	xiaofangwu888@gmail.com	?		?		?	US	1111111111	ADCentral	Promoter Plan			$ (1,425.00)	$ 332,966.90		$(542,358.84)	$ 244,447.04	$ -	$ 331,541.90
10770583	10396601	susan88824	Susan	**telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)	$ 1,574.70		$ -	$ 4,404.66	$ -	$ 149.70
10770589	10396601	susan88825	Susan	**telexfree988@yahoo.com**	?		?		?	ZM		ADCentral	Promoter Plan			$ (1,425.00)	$ 1,018.80		$ -	$ 2,700.00	$ -	$ (406.20)
10770595	10396601	susan88826	Susan	**telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,000.00	$ -	$ (1,425.00)
10770921	10396601	susan88829	Susan	**telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,000.00	$ -	$ (1,425.00)
10770943	10396601	susan88832	Susan	**Telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,000.00	$ -	$ (1,425.00)
10770961	10396601	susan88835	Susan	**telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,000.00	$ -	$ (1,425.00)
10770969	10396601	susan88836	Susan	**telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,000.00	$ -	$ (1,425.00)
10770973	10396601	susan88837	Susan	**telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,000.00	$ -	$ (1,425.00)
10770981	10396601	susan88838	Susan	**telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,000.00	$ -	$ (1,425.00)
10770989	10396601	susan88839	Susan	**telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,000.00	$ -	$ (1,425.00)
10770997	10396601	susan88841	Susan	**telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,000.00	$ -	$ (1,425.00)
10787491	10396601	susan88853	Susan	**telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,000.00	$ -	$ (1,425.00)
10787495	10396601	susan88854	Susan	**telexfree988@yahoo.com**	?		?	?	?	ZM		ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 3,000.00	$ -	$ (1,425.00)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.28: #28 Top Net Loser - Parent ID 12242931 ($210,000 Net Loser) v. Parent ID 2004122 ($199,000 Net Winner) and Parent ID 5510906 ($400,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
12242931	$ (209,780.80)	204

Example of User Accounts:

													$ -	$ -	$ (226,562.30)	$ 16,781.50	$ -	$ (41,701.76)	$ 186,162.22	$ -	**$ (209,780.80)**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
12553117	12242931	lulily3	vacerly petty	**lucio_mascena@hotmail.com**	499 erza st		bridgeport	ct	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)	$ 199.60		$ -	$ 1,996.91	$ -	$ (1,225.40)
12553169	12242931	lulily4	vacerly petty	**lucio_mascena@hotmail.com**	499 erza st		bridgeport	ct	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,300.00	$ -	$ (1,425.00)
12717429	12242931	vpetty30	Vacerly S Petty	fernellusa@gmail.com	Ezra st	499	Bridgeport	ct	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,100.00	$ -	$ (1,425.00)
12717547	12242931	vpetty31	Vacerly S Petty	fernellusa@gmail.com	Ezra st	499	Bridgeport	ct	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,200.00	$ -	$ (1,425.00)
12717645	12242931	vpetty32	Vacerly S Petty	fernellusa@gmail.com	Ezra st	499	Bridgeport	ct	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,000.00	$ -	$ (1,425.00)
12717817	12242931	vpetty33	Vacerly S Petty	fernellusa@gmail.com	Ezra st	499	Bridgeport	Ct	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,000.00	$ -	$ (1,425.00)
12726177	12242931	vpetty34	Vacerly S Petty	fernellusa@gmail.com	Ezra st	499	Bridgeport	ct	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 900.00	$ -	$ (1,425.00)
13666783	12242931	lulily30	vacerly petty	**rhadassausa@hotmail.com**	499 erza st		bridgeport	CT	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,200.00	$ -	$ (1,425.00)
14337347	12242931	lulily35	vacerly petty	**rhadassausa@hotmail.com**	499 erza st		bridgeport	CT	06606	US	2033830216	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,000.00	$ -	$ (1,425.00)
16278415	12242931	**lubb54**	vacerly petty	**rhadassausa@hotmail.com**	499 erza st		bridgeport	CT	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
16279925	12242931	**lubb57**	vacerly petty	**rhadassausa@hotmail.com**	499 erza st		bridgeport	CT	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17491165	12242931	**lubb70**	vacerly petty	**rhadassausa@hotmail.com**	499 erza st		bridgeport	CT	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)
17492005	12242931	**lubb71**	vacerly petty	**rhadassausa@hotmail.com**	499 erza st		bridgeport	CT	06606	US	2032097044	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 300.00	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
2004122	$ 199,241.40	573

Example of User Accounts:

													$ (26,175.00)	$ -	$ (485,649.80)	$ 817,591.20	$ (106,525.00)	$ (1,252,104.80)	$1,171,814.69	$ -	**$ 199,241.40**
Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
11320765	2004122	usalucio3	lucio jose macena	**rhadassausa@hotmail.com**	21 main st	....	framingham	MA	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)	$ 49.90		$ -	$ 10,997.05	$ -	$ (1,375.10)
11320831	2004122	usalucio4	lucio jose macena	**lucio_mascena@hotmail.com**	21 main st	...	framingham	MA	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 2,300.00	$ -	$ (1,425.00)
12549445	2004122	**lubb3**	lucio jose mascena	**lucio_mascena@hotmail.com**	21 main st		framingham	ma	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,300.00	$ -	$ (1,425.00)
12549525	2004122	**lubb4**	lucio jose mascena	**lucio_mascena@hotmail.com**	21 main st		framingham	ma	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,300.00	$ -	$ (1,425.00)
12549597	2004122	**lubb5**	lucio jose mascena	**lucio_mascena@hotmail.com**	21 main st		framingham	ma	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,300.00	$ -	$ (1,425.00)
12549655	2004122	**lubb6**	lucio jose mascena	**lucio_mascena@hotmail.com**	21 main st		framingham	ma	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,200.00	$ -	$ (1,425.00)
2004122	2004122	mascena	Lucio Mascena	lucio_mascena@hotmail.com	21 Main St	21	framingham	MA	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,375.00)	$ 132,796.60	$ (90,800.00)	$ (168,161.16)	$ 138,218.86	$ -	$ 40,621.60
10539645	2004122	usalucio	lucio Macena	lucio_mascena@hotmail.com	21 Main St	.	framingham	MA	1701	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)	$ 281,087.30	$ (15,725.00)	$ (448,555.08)	$ 232,127.32	$ -	$ 263,937.30
10539785	2004122	usalucio1	lucio Jose Macena	lucio_mascena@hotmail.com	21 Main St	.	framingham	ma	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)	$ 91,998.40		$ (158,492.48)	$ 70,865.32	$ -	$ 90,573.40
12549671	2004122	**lubb7**	lucio jose mascena	**lucio_mascena@hotmail.com**	21 main st		framingham	ma	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,300.00	$ -	$ (1,425.00)
12549729	2004122	**lubb8**	lucio jose mascena	**lucio_mascena@hotmail.com**	21 main st		framingham	ma	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,200.00	$ -	$ (1,425.00)
12547133	2004122	bulgary	lucio jose mascena	lucio_mascena@hotmail.com	21 main st		framingham	ma	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)	$ 145,072.70		$ (133,952.32)	$ 300.00	$ -	$ 143,647.70
13407523	2004122	starr	lucio jose mascena	rhadassausa@hotmail.com	21 main st		framingham	MA	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)	$ 84,673.80		$ (77,088.04)	$ 1,883.00	$ -	$ 83,248.80
13640959	2004122	**lubb27**	lucio jose mascena	**rhadassausa@hotmail.com**	21 main st		framingham	MA	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 900.00	$ -	$ (1,425.00)
13641037	2004122	**lubb28**	lucio jose mascena	**rhadassausa@hotmail.com**	21 main st		framingham	MA	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,300.00	$ -	$ (1,425.00)
13641103	2004122	**lubb29**	lucio jose mascena	**rhadassausa@hotmail.com**	21 main st		framingham	MA	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,300.00	$ -	$ (1,425.00)
13641155	2004122	**lubb30**	lucio jose mascena	**rhadassausa@hotmail.com**	21 main st		framingham	MA	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,300.00	$ -	$ (1,425.00)
13641209	2004122	**lubb31**	lucio jose mascena	**rhadassausa@hotmail.com**	21 main st		framingham	MA	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,300.00	$ -	$ (1,425.00)
14920737	2004122	lottus	lucio jose mascena	rhadassausa@hotmail.com	21 main st		framingham	MA	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)	$ 41,724.20		$ (53,164.49)	$ 1,673.00	$ -	$ 40,299.20
16756465	2004122	flly6	lucio jose mascena	**rhadassausa@hotmail.com**	21 main st		framingham	MA	1702	US	5088089962	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,045.91	$ -	$ (1,425.00)
17606777	2004122	lottus3cli	lucio jose mascena	**rhadassausa@hotmail.com**	21 main st		framingham	MA	1702	US	5088089962	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.28: #28 Top Net Loser - Parent ID 12242931 ($210,000 Net Loser) v. Parent ID 2004122 ($199,000 Net Winner) and Parent ID 5510906 ($400,000 Net Winner)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
5510906	$ 399,904.40	71

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (22,232.20)	$ 451,336.60	$ (29,200.00)	$ (479,253.49)	$ 129,841.36	$ -	**$ 399,904.40**
5617866	5510906	fernell1	Fernando Figueiredo	fernellusa88@hotmail.com	468 Anson St .		Bridgeport	ct	6606	US	2032970358	ADCentral	Promoter Plan			$ (339.00)	$ 898.50		$ -	$ -	$ -	$ 559.50
10730599	5510906	fernell2	Fernando Figueiredo	fernellusa@gmail.com	Anson st	468	Bridgeport		6606	US	2032970358	ADCentral	Promoter Plan			$ (1,495.00)	$ 411,222.50	$ (29,200.00)	$ (415,712.58)	$ 79,996.00	$ -	$ 380,527.50
12670033	5510906	fernell10	Fernando Figueiredo	fernellusa@gmail.com	Anson st	468	Bridgeport	Ct	6606	US	2032970358	ADCentral	Promoter Plan			$ (1,425.00)	$ 26,198.90		$ (40,275.94)	$ 16,780.95	$ -	$ 24,773.90
14522149	5510906	fernell11	Fernando Figueiredo	fernellusa@gmail.com	Anson st	468	Bridgeport	CT	6606	US	2032970358	ADCentral	Promoter Plan			$ (1,495.00)	$ 4,804.40		$ (10,909.42)	$ 9,916.14	$ -	$ 3,309.40
14525545	5510906	fernell12	Fernando Figueiredo	fernellusa@gmail.com	Anson st	468	Bridgeport	CT	6606	US	2032970358	ADCentral	Promoter Plan			$ (339.00)	$ 149.70		$ -	$ -	$ -	$ (189.30)
15817797	5510906	fernell13	Fernando Figueiredo	fernellusa@gmail.com	Anson st	468	Bridgeport	CT	6606	US	12032970358	ADCentral	Promoter Plan			$ (339.00)	$ 199.60		$ -	$ -	$ -	$ (139.40)
15819005	5510906	fernell14	Fernando Figueiredo	fernellusa@gmail.com	Anson st	468	Bridgeport	CT	6606	US	12032970358	ADCentral	Promoter Plan			$ (339.00)	$ 99.80		$ -	$ -	$ -	$ (239.20)
17033397	5510906	fernell15	Fernando Figueiredo	fernellusa@gmail.com	468 Anson st		Bridgeport	CT	6606	US	2032970358	ADCentral	Promoter Plan			$ (1,425.00)	$ 299.40		$ -	$ 524.91	$ -	$ (1,125.60)
18230493	5510906	fernell16	Fernando Figueiredo	fernellusa@gmail.com	468 Anson st	468	Bridgeport	CT	6606	US	2032970358	Associate	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
20113867	5510906	fernell17	Fernando Figueiredo	fernellusa@gmail.com	Anson St.	468	Bridgeport	CT	6606	US	2032970358	ADFamily	Promoter Plan			$ (339.60)	$ 489.30		$ -	$ -	$ -	$ 149.70

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.29: #29 Top Net Loser - Parent ID 17443697 ($205,000 Net Loser) v. Parent ID 147979 ($2.4M Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
17443697	$ (205,031.70)	938

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (205,031.70)	$ -	$ -	$ -	$ 9,401.17	$ -	**$ (205,031.70)**
17485679	17443697	grupaguia2	grupo aguia usa corp	**eduardomutum@gmail.com**	**271 pulaski blvd**		bellingham	MA	02019	US	**7742876860**	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
19778255	17443697	t01cli1	grupo aguia corp	eduadomutum@gmail.com	**42 east st**	42	framingham	MA	01702	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19778437	17443697	t01cli2	grupo aguia corp	eduadomutum@gmail.com	**42 east st**	42	framingham	MA	01702	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19778631	17443697	t01cli3	grupo aguia corp	eduadomutum@gmail.com	**42 east st**	42	framingham	MA	01702	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19778649	17443697	t01cli7	grupo aguia usa corp	**eduardomutum@gmail.com**	union ave	29	framingham	MA	01702	US	8	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19778859	17443697	t01cli4	grupo aguia corp	eduadomutum@gmail.com	**42 east st**	42	framingham	MA	01702	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19778903	17443697	t01cli8	grupo aguia usa corp	**eduardomutum@gmail.com**	union ave	29	framingham	MA	01702	US	8	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19778993	17443697	t01cli5	grupo aguia corp	eduadomutum@gmail.com	**42 east st**	42	framingham	MA	01702	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19780477	17443697	**topeagle02**	grupo aguia usa corp	**eduardomutum@gmail.com**	24 union av #29		framingham	MA	01702	US	**7742876860**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
19781283	17443697	t02cli6	grupo aguia usa corp	**eduardomutum@gmail.com**	24 union av #29	29	framingham	MA	01702	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19781567	17443697	t02cli2	grupo aguia usa corp	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19782025	17443697	t02cli4	grupo aguia usa corp	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19782513	17443697	t02cli10	grupo aguia usa corp	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19785017	17443697	**topeagle03**	grupo aguia usa corp	**eduardomutum@gmail.com**	24 union av #29	24	framingham	MA	01702	US	**7742876860**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ 4,876.26	$ -	$ (339.60)
19786339	17443697	t03cli1	grupo aguia usa corp	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19786771	17443697	t03cli2	grupo aguia usa corp	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19787519	17443697	t03cli5	grupo aguia usa corp	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
20096033	17443697	**topeagle04**	grupo aguia usa corp	eduardomutum@gmail.com	24 union av #29		framingham	MA	01702	US	**7742876860**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
20097819	17443697	**topeagle05**	grupo aguia usa corp	**eduardomutum@gmail.com**	24 union av #29		framingham	MA	01702	US	**7742876860**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
147979	$ 2,405,821.37	542

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (15,174.90)	$ 40,790.00	$ (253,190.67)	$ 2,815,123.67	$ (209,106.63)	$(3,432,871.94)	$ 3,102,336.14	$ -	**$ 2,405,821.37**
147979	147979	mutum	Eduardo da Silva	**eduardomutum@gmail.com**	**271 Pulaski Blvd**	271	bellingham		02019	US	**7742876860**	ADCentral	Promoter Plan	$ (1,375.00)	$ 3,680.00	$ (51,950.29)	$ 296,891.00	$ (141,225.72)	$ (165,678.50)	$ 123,106.08	$ -	$ 133,399.89
148789	147979	mutum1	eduardo nunes dasilva	**eduardomutum@gmail.com**	**271 Pulaski Blvd**	271	Bellingham	ma	02019	US	**7742876860**	ADCentral	Promoter Plan	$ (1,375.00)	$ 14,310.00	$ (26,826.08)	$ 403,150.10		$ (498,730.77)	$ 215,017.10	$ -	$ 389,259.02
148793	147979	mutum2	eduardo nunes dasilva	**eduardomutum@gmail.com**	**271 Pulaski Blvd**	271	Bellingham	ma	02019	US	**7742876860**	ADCentral	Promoter Plan	$ (1,375.00)	$ 300.00	$ (2,385.00)			$ -	$ 6,400.00	$ -	$ (3,460.00)
9347146	147979	mutummg1	eduardo Nunes Da silva	**eduardomutum@gmail.com**	**271 pulaski blvd**	271	bellingham	MA	02019	US	**7742876860**	ADCentral	Promoter Plan		$ 20,100.00	$ (1,495.00)	$ 255,615.37	$ (44,550.00)	$ (333,944.06)	$ 496,169.71	$ -	$ 229,670.37
19481219	147979	**topeagle1**	Eduardo Silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
19483193	147979	**topeagle1cli**	Eduardo silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19483953	147979	**topeagle1cli2**	Eduardo silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19484061	147979	**topeagle1cli3**	Eduardo Silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19484265	147979	**topeagle1cli4**	Eduardo Silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19484391	147979	**topeagle1cli5**	Eduardo Silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19484493	147979	**topeagle1cli6**	Eduardo Silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19484679	147979	**topeagle1cli7**	Eduardo Silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19484803	147979	**topeagle1cli8**	Eduardo Silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19484911	147979	**topeagle1cli9**	Eduardo Silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19485163	147979	**topeagle1cli10**	Eduardo Silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
19487551	147979	**topeagle1cli1**	Eduardo Silva	**eduardomutum@gmail.com**	**42 East St**		Framingham	MA	01702	US	**17742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
27581520	147979	top09cli	Eduardo Silva	**eduardomutum@gmail.com**	**271 pulaski blvd**		bellingham	MA	02019	US	**7742876860**	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
17595335	147979	foxbrasil23	Eduardo Silva	**eduardomutum@gmail.com**	**271 pulaski blvd**		bellingham	MA	02019	US	7742876860	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.30: #30 Top Net Loser - Parent ID 17495171 ($205,000 Net Loser) v. Parent ID 149697 ($1.2M Net Winner) and Parent ID 132785 ($4.1M Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
17495171	$ (204,723.10)	162

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (204,723.10)	$ -	$ -	$ -	$ 11,928.74	$ -	**$ (204,723.10)**
17495171	17495171	nlm01	Calvary Assembly of God	**zieffj@yahoo.com**	7272 Cerritos Av.		Stanton	CA	92562	US	7146588212	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17499551	17495171	nlm10	Calvary Assembly of God	**zieffj@yahoo.com**	7272 Cerritos Av.		Stanton	CA	90680	US	7146588212	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,524.91	$ -	$ (1,425.00)
17584613	17495171	nlm100	Calvary Assembly of God	**zieffj@yahoo.com**	7272 Cerritos Av.		Stanton	CA	90680	US	7146588212	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17584781	17495171	nlm101	Calvary Assembly of God	**zieffj@yahoo.com**	7272 Cerritos Av.		Stanton	CA	90680	US	7146588212	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17584899	17495171	nlm102	Calvary Assembly of God	**zieffj@yahoo.com**	7272 Cerritos Av.		Stanton	CA	90680	US	7146588212	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17584963	17495171	nlm103	Calvary Assembly of God	**zieffj@yahoo.com**	7272 Cerritos Av.		Stanton	CA	90680	US	7146588212	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17591179	17495171	nlm104	Calvary Assembly of God	**zieffj@yahoo.com**	7272 Cerritos Av.		Stanton	CA	90680	US	7146588212	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17591361	17495171	nlm105	Calvary Assembly of God	**zieffj@yahoo.com**	7272 Cerritos Av.		Stanton	CA	90680	US	7146588212	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17591427	17495171	nlm106	Calvary Assembly of God	**zieffj@yahoo.com**	7272 Cerritos Av.		Stanton	CA	90680	US	7146588212	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17591479	17495171	nlm107	Calvary Assembly of God	**zieffj@yahoo.com**	7272 Cerritos Av.		Stanton	CA	90680	US	7146588212	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
17591623	17495171	nlm108	Calvary Assembly of God	**zieffj@yahoo.com**	7272 Cerritos Av.		Stanton	CA	90680	US	7146588212	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 924.91	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
149697	$ 1,228,447.40	831

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (299.00)	$ -	$ (327,261.80)	$ 588,367.10	$ -	$ (18,590,214.98)	$ 19,553,754.00	$ -	**$ 1,228,447.40**
12996381	149697	flywings11	helio barbosa	**zieffj@yahoo.com**	**2b brackett rd**		framingham	MA	01702	US	5088085532	ADCentral	Promoter Plan						$ -	$ 1,200.00	$ -	$ (1,425.00)
12996505	149697	flywings12	helio barbosa	**zieffj@yahoo.com**	**2b brackett rd**		framingham	MA	01702	US	5088085532	ADCentral	Promoter Plan						$ -	$ 1,200.00	$ -	$ (1,425.00)
12996647	149697	flywings13	helio barbosa	**zieffj@yahoo.com**	**2b brackett rd**		framingham	MA	01702	US	5088085532	ADCentral	Promoter Plan						$ -	$ 1,200.00	$ -	$ (1,425.00)
13714941	149697	floridanews3	helio barbosa	**zieffj@yahoo.com**	**2b bracket road**		framingham	MA	01702	US	5088085532	ADCentral	Promoter Plan						$ -	$ 6,690.00	$ -	$ (876.10)
13715053	149697	floridanews3cl	helio barbosa	**zieffj@yahoo.com**	**2b bracket road**		framingham	MA	01702	US	5088085532	Partner	Phone Plan / Other						$ -	$ -	$ -	$ (199.60)
13715251	149697	floridanews4	helio barbosa	**zieffj@yahoo.com**	**2b bracket road**		framingham	MA	01702	US	5088085532	ADCentral	Promoter Plan						$ -	$ 5,119.64	$ -	$ (1,175.50)
13716025	149697	floridanews4cl	helio barbosa	**zieffj@yahoo.com**	**2b bracket road**		framingham	MA	01702	US	5088085532	Partner	Phone Plan / Other						$ -	$ -	$ -	$ (199.60)
13716229	149697	floridanews5	helio barbosa	**zieffj@yahoo.com**	**2b bracket road**		framingham	MA	01702	US	5088085532	ADCentral	Promoter Plan						$ -	$ 833.00	$ -	$ (926.00)
13716323	149697	floridanews5cl	helio barbosa	**zieffj@yahoo.com**	**2b bracket road**		framingham	MA	01702	US	5088085532	Partner	Phone Plan / Other						$ -	$ -	$ -	$ (199.60)
13829801	149697	floridaoffice	Helio Barbosa	**zieffj@yahoo.com**	**2b Brackett Rd**	2	Framingham	MA	01702	US	5088085532	ADCentral	Promoter Plan						$ -	$ 344.91	$ -	$ (1,425.00)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
132785	$ 4,096,688.59	431

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (9,705.00)	$ -	$ (660,396.53)	$ 4,956,440.12	$ (189,650.00)	$ (7,287,435.94)	$ 6,617,041.09	$ -	**$ 4,096,688.59**
132785	132785	ttmilagres	Maria Teresa Milagres Nev	ttmilagres@gmail.com	1630 Worcester rd	1630	Framingham		01702	US	5089043810	ADCentral	Promoter Plan	$ (299.00)		$ (193,956.43)	$ 2,774,595.60		$ (2,836,972.78)	$ 2,864,686.80	$ -	$ 2,580,340.17
162734	132785	ttmilagres3	Maria T Neves	ttmilagres@gmail.com	1630 Worcester rd	626	Framingham	MA	01702	US	5089043810	ADCentral	Promoter Plan	$ (1,076.00)		$ (299.00)	$ 25,906.00		$ (26,549.89)	$ 5,383.89	$ -	$ 24,531.00
206326	132785	ttmilagres9	Maria Teresa Milagres Nev	ttmilagres@gmail.com	**2B Brackett rd**	2B	Framingham	MA	01702	US	5089043810	ADCentral	Promoter Plan	$ (1,375.00)			$ 12,964.00		$ (12,414.28)	$ 4,350.28	$ -	$ 11,589.00
2240074	132785	ttneves1	Maria Teresa Milagres Nev	teresamilagres@hotmail.com	**2B Brackett rd**	2B	Framingham	MA	01702	US	5089043810	ADCentral	Promoter Plan			$ (1,495.00)			$ -	$ 2,060.00	$ -	$ (1,495.00)
4471758	132785	ttstars	Maria Teresa Milagres Nev	ttmilagres@gmail.com	**2b Brackett rd**	2b	framingham		01702	US	5089043810	ADCentral	Promoter Plan			$ (1,425.00)	$ 1,772,912.02	$ (189,650.00)	$ (3,826,234.82)	$ 3,095,588.97	$ -	$ 1,581,837.02
8653610	132785	callandsave6	Maria Teresa Neves	ttmilagres@gmail.com	**2B Brackett rd**	2B	Framingham	MA	01702	US	5089043810	ADCentral	Promoter Plan			$ (1,425.00)	$ 49.90		$ -	$ 4,364.82	$ -	$ (1,375.10)
10526539	132785	ttmilagres01	Maria Neves	**zieffj@yahoo.com**	**2B Brackett Rd**	.	.	MA	.	US	5089043810	ADCentral	Promoter Plan			$ (1,425.00)	$ 87,267.00		$ (103,667.23)	$ 87,338.20	$ -	$ 85,842.00
11244377	132785	lilimax	Maria Teresa Milagres Nev	ttmilagres@gmail.com	**2 B Brackett rd**	2B	Framingham	MA	01702	US	5089043810	ADCentral	Promoter Plan			$ (1,425.00)	$ 239,137.10		$ (253,349.12)	$ 5,753.00	$ -	$ 237,712.10
11244877	132785	lilimax2	Maria Teresa Milagres Nev	ttmilagres@gmail.com	**2 B brackett rd**	.	Framingham	MA	01702	US	5089043810	ADCentral	Promoter Plan			$ (1,495.00)			$ (8,450.46)	$ 12,402.10	$ -	$ (1,495.00)
11246147	132785	lilimax9	Maria Teresa Milagres Nev	ttmilagres@gmail.com	**2B Brackett rd**	2B	framingham	ma	01702	US	5089043810	ADCentral	Promoter Plan			$ (1,425.00)	$ 49.90		$ -	$ 4,234.65	$ -	$ (1,375.10)
11277987	132785	ttnews1	Maria Teresa m Neves	hbmaltaro@gmail.com	2b Brackett Rd	.	Framingham		01702	US	5089043810	ADCentral	Promoter Plan			$ (1,425.00)	$ 28,599.80		$ (105,542.58)	$ 78,999.00	$ -	$ 27,174.80
13997415	132785	thth	Maria Neves	ttmilagres@gmail.com	**2 B Brakett Rd**	2	Framingham		01702	US	5089043810	ADCentral	Promoter Plan			$ (1,425.00)	$ 49.90		$ (10,320.82)	$ -	$ -	$ (1,375.10)
11458747	132785	ttnews17	Maria Teresa Milagres Nev	ttmilagres@gmail.com	**2B Brackett rd**	.	framingham	ma	01702	US	5089043810	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ 1,700.00	$ -	$ (1,425.00)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 9.31: #31 Top Net Loser - Parent ID 17369623 ($203,000 Net Loser) v. Parent ID 2615682 ($328,000 Net Winner)

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
17369623	$ (202,658.80)	286

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ -	$ -	$ (206,172.60)	$ 3,513.80	$ -	$ (1,287.64)	$ 85,161.50	$ -	**$ (202,658.80)**
26039894	17369623	marira1	francisca Alarcon	**mauricioalarconmartinez@gmail.com**	Col.Guadalupe Calle A San Jose	17	Metapan	SA	00000	SV	75886785	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
26098154	17369623	marira3	francisca alarcon	**mauricioalarconmartinez@gmail.com**	col.guadalupe calle a sanjose	17	metapan	SA	00000	SV	75886785	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
26127064	17369623	marira5	francisca alarcon	**mauricioalarconmartinez@gmail.com**	col.guadalupe calle a sanjose	17	metapan	SA	00000	SV	75886785	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
26128018	17369623	marira6	francisca alarcon	**mauricioalarconmartinez@gmail.com**	col.guadalupe calle a sanjose	17	metapan	SA	00000	SV	75886785	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
26110850	17369623	marira4	francisca alarcon	**mauricioalarconmartinez@gmail.com**	col.guadalupe calle a sanjose	17	metapan	SA	00000	SV	75886785	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
26045780	17369623	marira2	Francisca Alarcon	**mauricioalarconmartinez@gmail.com**	col.guadalupe calle a sanjose	17	metapan	SA	00000	SV	75886785	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
2615682	$ 327,836.80	306

Example of User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
														$ (7,125.00)	$ -	$ (115,752.50)	$ 533,094.30	$ (82,380.00)	$ (618,990.57)	$ 455,543.11	$ -	**$ 327,836.80**
2615682	2615682	mauricio78	Mauricio Alarcon Martinez	alarcontelex@yahoo.com	16 Illinois Ave	2	Somerville	MA	2145	US	6177554646	ADCentral	Promoter Plan	$ (1,425.00)			$ 126,216.00	$ (19,950.00)	$ (217,001.75)	$ 141,818.74	$ -	$ 104,841.00
4310718	2615682	mauricio78i	Mauricio Alarcon Martinez	**mauricioalarconmartinez@gmail.com**	16 Illinois Ave	2	Somerville	MA	2145	US	6177554646	ADCentral	Promoter Plan	$ (1,425.00)			$ 7,826.30		$ (5,078.00)	$ 5,735.97	$ -	$ 6,401.30
5841052	2615682	mauricio78rr	mauricioalarconmartinez	**mauricioalarconmartinez@gmail.com**	16 Illinois Ave	2	Somerville	MA	2145	US	6177554646	ADCentral	Promoter Plan			$ (1,425.00)	$ 199.60		$ -	$ 4,624.00	$ -	$ (1,225.40)
5438522	2615682	mauricio78mm	mauricioalarconmartinez	**mauricioalarconmartinez@gmail.com**	16 Illinois Ave	2	Somerville	MA	2145	US	6177554646	ADCentral	Promoter Plan			$ (1,425.00)	$ 16,024.30		$ (15,447.00)	$ 4,347.00	$ -	$ 14,599.30
5438864	2615682	mauricio78nn	mauricioalarconmartinez	**mauricioalarconmartinez@gmail.com**	16 Illinois Ave	2	Somerville	MA	2145	US	6177554646	ADCentral	Promoter Plan			$ (1,425.00)	$ 199.60		$ -	$ 4,624.00	$ -	$ (1,225.40)
5917950	2615682	mauricio7815	Mauricio Alarcon Martinez	**mauricioalarconmartinez@gmail.com**	16 Illinois Ave	2	Somerville	MA	2145	US	6177554646	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
5918006	2615682	mauricio7828	mauricioalarconmartinez	**mauricioalarconmartinez@gmail.com**	16 illinois ave	2	somerville	ma	2145	US	6177554646	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
5918230	2615682	mauricio7816	Mauricio Alarcon Martinez	**mauricioalarconmartinez@gmail.com**	16 Illinois Ave	2	Somerville	MA	2145	US	6177554646	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10815617	2615682	yeseniakk	mauricio alarcon martinez	**mauricioalarconmartinez@gmail.com**	16 Illinois Ave	2	Somerville	MA	2145	US	6177554646	ADCentral	Promoter Plan			$ (1,425.00)	$ 38,197.70		$ (36,610.00)	$ 4,543.69	$ -	$ 36,772.70
10815795	2615682	yeseniamm	mauricio alarcon martinez	**mauricioalarconmartinez@gmail.com**	16 Illinois Ave	2	Somerville	MA	2145	US	6177554646	ADCentral	Promoter Plan			$ (1,425.00)	$ 163,110.50	$ (53,880.00)	$ (155,587.44)	$ 51,870.34	$ -	$ 107,805.50
11549381	2615682	yeseniaal11	Mauricio Alarcon Martinez	**yruizalarcon@gmail.com**	16 illinois av	2	somerville	MA	2145	US	6177554646	ADCentral	Promoter Plan			$ (1,425.00)	$ 145,334.60		$ (144,649.09)	$ 1,754.00	$ -	$ 143,909.60
17216693	2615682	mama5r	mauricio Martinez	mauricioalarconmartinez@gmail.com	16 illinois av	2	somerville	MA	2145	US	6177554646	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Source(s): Based upon queries of the Huron Data Tables

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 10: Frantz Balan - Parent ID 7692912 ($519,000 Net Winner) and Lusette Balan - Parent ID 10391181 ($721,000 Net Winner) User Account Detail

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
7692912	$ 518,540.40	211

Frantz and Lusette Balan													**$ -**	**$ 500.00**	**$ (25,534.20)**	**$ 1,490,381.50**	**$ (225,520.00)**	**$ (2,133,288.81)**	**$ 1,285,643.43**	**$ -**	**$ 1,239,827.30**
													$ -	$ 500.00	$ (22,273.70)	$ 762,984.10	$ (222,670.00)	$ (625,513.96)	$ 323,569.03	$ -	**$ 518,540.40**

User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
7692912	7692912	frantzbalan	Frantz Balan	balan.f@gmail.com	51 Grover St	2	Everett		2149	US	6172832735	ADCentral	Promoter Plan		$ 500.00	$ (488.70)	$ 639,547.80	$ (222,670.00)	$ (520,606.44)	$ 298,944.54	$ -	$ 416,889.10
10482725	7692912	frantzbalanf	frantz balan	balan.f@yahoo.com	balan.f@gmail.com		Malden	MA	2148	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10552045	7692912	majojo5	frantz balan	frantzbalan@yahoo.com	51 grover st		everett	MA	2149	US	6138942435	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10552151	7692912	majojo6	frantz balan	frantzbalan@yahoo.com	55 grover		everett	MA	2149	US	13057647449	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10623309	7692912	mavolcyc1	frantz balan	frantzbalan@yahoo.com	55 grover		everett	MA	2149	US	7813912804	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10623603	7692912	duvivi3	frantz balan	frantzbalan@yahoo.com	55 grover		everett	MA	2149	US	7813912804	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10623721	7692912	duvivi9	frantz balan	frantzbalan@yahoo.com	55 grover		everett	MA	2149	US	7813912804	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
10623757	7692912	duvivi11	frantz balan	frantzbalan@yahoo.com	55 grover		everett	MA	2149	US	7813912804	Partner	Phone Plan / Other			$ (99.80)			$ -	$ -	$ -	$ (99.80)
10747699	7692912	frantzp	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	ADCentral	Promoter Plan			$ (388.90)	$ 5,013.20		$ -	$ 2,310.00	$ -	$ 4,624.30
10820493	7692912	vancol1c2	frantz balan	frantzbalan@yahoo.com	55 grover		everett	MA	2149	US	7813912804	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10844859	7692912	duvivi4c2	frantz balan	balan.f@gmail.com	51 grover st		everett	Ma	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10855765	7692912	duvivi12c1	frantz balan	frantzbalan@yahoo.com	55 grover		everett	MA	2149	US	7813912804	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10855835	7692912	duvivi14c2	frantz balan	frantzbalan@yahoo.com	55 grover		everett	MA	2149	US	7813912804	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10904819	7692912	duvivi17c1	frantz balan	balan.f@gmail.com	51 Grover st		Everett	ma	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
10938241	7692912	beauvois10	frantz balan	balan.f@gmail.om	51 grovwer st		everett	ma	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
11098891	7692912	yourway55	frantz Balan	balan.f@gmail.com	51grover st		Everett		2149	US	6172832735	ADCentral	Promoter Plan			$ (339.00)	$ 438.80		$ -	$ 5,091.94	$ -	$ 99.80
11125853	7692912	yourway44	frantz balan	balan.f@gmail.com	51 grover st		everett	ma	2149	US	6172832735	Partner	Promoter Plan			$ (50.00)			$ -	$ -	$ -	$ (50.00)
11367831	7692912	frantz2013	frantz balan	frantzbalan@yahoo.com	51 grover st		everett	ma	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
11751885	7692912	frantzbalan1	frantz balan	balan.f@gmail.com	51 grover st		everett	ma	2149	US	6172832735	Partner	Promoter Plan			$ (50.00)			$ -	$ -	$ -	$ (50.00)
11888033	7692912	frantzbalan2	frantz balan	frantzbalan@yahoo.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Promoter Plan			$ (50.00)			$ -	$ -	$ -	$ (50.00)
11912787	7692912	frantzbalan3	frantz balan	frantzbalan@yahoo.com	51 grover st		everett		2149	US	6172832735	ADCentral	Promoter Plan			$ (339.00)	$ 114,679.80		$ (104,907.52)	$ 16,066.33	$ -	$ 114,340.80
12186985	7692912	frantzbalan4	frantz balan	balan.f@gmail.com	51 Grover st		everett	ma	2149	US	6172832735	ADCentral	Promoter Plan			$ (438.80)			$ -	$ -	$ -	$ (438.80)
13255335	7692912	frantzbalan5	frantz balan	balan.f@gmail.com	51 gropver st		everett		2149	US	6172832735	ADCentral	Promoter Plan			$ (388.90)			$ -	$ 656.22	$ -	$ (388.90)
13320749	7692912	frantzbiz	frantz balan	balan.f@gmail.com	51 grover st		Everett		2149	US	6172832735	ADCentral	Promoter Plan			$ (339.00)			$ -	$ -	$ -	$ (339.00)
13322093	7692912	frantzbalan7	frantz balan	balan.f@gmail.com	51 gropver st		everett	MA	2149	US	6172832735	ADCentral	Promoter Plan			$ (339.00)			$ -	$ -	$ -	$ (339.00)
13586697	7692912	frantzbalanbiz1	frantz balan	balan.f@gmail.com	55 grover		everett	MA	2149	US	6172832735	ADCentral	Promoter Plan			$ (1,425.00)			$ -	$ -	$ -	$ (1,425.00)
14210457	7692912	frantzbalan8	frantz balan	balan.f@gmail.com	55 grover		everett	ND	2149	HT	7813912804	ADCentral	Promoter Plan			$ (1,524.80)	$ 3,254.60		$ -	$ 500.00	$ -	$ 1,729.80
15038021	7692912	frantzbalanbiz3	frantz balan	balan.f@gmail.com	51 gropver st		everett	MA	2149	US	6172832735	ADCentral	Promoter Plan			$ (388.90)			$ -	$ -	$ -	$ (388.90)
15241895	7692912	frantzbalanbiz4	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	ADCentral	Promoter Plan			$ (339.00)			$ -	$ -	$ -	$ (339.00)
15513041	7692912	yourway01	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Promoter Plan			$ (50.00)			$ -	$ -	$ -	$ (50.00)
15515169	7692912	yourway02	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Promoter Plan			$ (50.00)			$ -	$ -	$ -	$ (50.00)
17419093	7692912	mytelexfree005c	frantz balan	balan.f@gmail.com	27 rector rd		Everett	MA	2126	US	6172822821	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18034733	7692912	yourway55c	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
18462583	7692912	balanglobal	FRANTZ BALAN	JCBOSTON01@GMAIL.COM	564 LOWELL ST		PEABODY	MA	1960	US	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
18632623	7692912	mytelexfree09	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Associate	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
20980287	7692912	frantzbalan11	frantz balan	vjmanigat@gmail.com	21burt street		Dorchester	MA	2124	US	7819225151	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
21025937	7692912	frantzbalan13	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Associate	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
21026123	7692912	frantzbalan12	frantz balan	vjmanigat@gmail.com	22 fargo rd		Holbrook	MA	2343	US	7819225151	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
21028399	7692912	frantzbalan14	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
21163287	7692912	frantz98	frantz balan	balan.f@gmail.com	68 edwin st		bridgewater	MA	2368	US	5086487816	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21170705	7692912	frantz58	frantz balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832537	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21171241	7692912	frantz59	frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832357	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21171949	7692912	frantz61	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	617283456	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21172271	7692912	frantz62	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	617283123	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21172581	7692912	frantz63	Frantz balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172382735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21172897	7692912	frantz64	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832736	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21173293	7692912	frantz65	Frantz balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832734	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21173621	7692912	frantz66	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832731	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21174179	7692912	frantz67	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832732	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21177479	7692912	frantz43	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832740	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21177977	7692912	frantz44	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832741	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21178343	7692912	frantz45	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832742	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21178673	7692912	frantz46	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832743	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21179117	7692912	frantz47	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832744	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21179469	7692912	frantz48	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832745	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21179929	7692912	frantz49	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832746	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21249993	7692912	frantz20	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832720	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21250489	7692912	frantz21	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832721	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21251105	7692912	frantz22	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832722	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21251549	7692912	frantz23	frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832723	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21251913	7692912	frantz24	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832724	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21252385	7692912	frantz25	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832725	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21252761	7692912	frantz26	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832726	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21253161	7692912	frantz27	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832727	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21253841	7692912	frantz28	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832728	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21254141	7692912	frantz29	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832729	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 10: Frantz Balan - Parent ID 7692912 ($519,000 Net Winner) and Lusette Balan - Parent ID 10391181 ($721,000 Net Winner) User Account Detail

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
7692912	$ 518,540.40	211

Frantz and Lusette Balan													**$ -**	**$ 500.00**	**$ (25,534.20)**	**$ 1,490,381.50**	**$ (225,520.00)**	**$ (2,133,288.81)**	**$ 1,285,643.43**	**$ -**	**$ 1,239,827.30**
													$ -	$ 500.00	$ (22,273.70)	$ 762,984.10	$ (222,670.00)	$ (625,513.96)	$ 323,569.03	$ -	**$ 518,540.40**

User Accounts:

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
21255175	7692912	frantz30	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832730	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21275919	7692912	frantzbalan15	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6712632735	ADFamily	Promoter Plan			$ (339.60)	$ 49.90		$ -	$ -	$ -	$ (289.70)
21336625	7692912	frantzbal1	frantz balan	balan.f@gmail.com	564 b rear lowell st		peabody	MA	9160	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21337463	7692912	frantzbal2	frantz balan	balan.f@gmail.com	564 b rear lowell st		peabody	MA	9160	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21339541	7692912	frantzbala7	frantz balan	balan.f@gmail.com	564 grover st		peabody	MA	1960	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21340635	7692912	frantzbal8	frantz balan	balan.f@gmail.com	564 grover st		peabody	MA	1960	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21341065	7692912	frantzbal3	frantz balan	balan.f@gmail.com	564 b rear lowell st		peabody	MA	9160	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21341519	7692912	frantzbal9	frantz balan	balan.f@gmail.com	564 grover st		peabody	MA	1960	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21342077	7692912	frantzbal4	frantz balan	balan.f@gmail.com	564 b rear lowell st		peabody	MA	9160	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21343209	7692912	frantzbal10	frantz balan	balan.f@gmail.com	564 grover st		peabody	MA	1960	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21343601	7692912	frantzbal5	frantz balan	balan.f@gmail.com	564 b rear lowell st		peabody	MA	9160	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21352779	7692912	frantzbal11	frantz balan	balan.f@gmail.com	564 b rear lowell st		peabody	MA	1960	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21366081	7692912	frantz31	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832710	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21369655	7692912	frantz32	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2148	US	6172832601	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21370093	7692912	frantz33	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832602	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21370611	7692912	frantz34	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832603	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21371113	7692912	frantz35	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832604	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21371449	7692912	frantz36	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832605	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21371933	7692912	frantz37	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832606	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21372259	7692912	frantz38	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832607	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21372629	7692912	frantz39	Frantz Balan	balan.f@gmail.com	Grover Street	55	Everett	MA	2149	US	6172832608	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21624913	7692912	balanglobalc1	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21625155	7692912	balanglobalc2	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21625855	7692912	balanglobalc3	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21626039	7692912	balanglobalc4	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21626201	7692912	balanglobalc5	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21626421	7692912	balanglobalc6	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21626623	7692912	balanglobalc7	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21626887	7692912	balanglobalc8	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21627143	7692912	balanglobalc9	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
21627651	7692912	balanglobalc10	frantz balan	balan.f@gmail.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25014538	7692912	yourway20	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25014862	7692912	yourway21	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25015772	7692912	yourway22	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25016518	7692912	yourway23	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25021416	7692912	yourway24	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25023542	7692912	yourway25	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25063424	7692912	yourway26	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25072992	7692912	yourway27	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25074404	7692912	yourway28	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25075430	7692912	yourway29	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25076712	7692912	yourway30	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25077286	7692912	yourway31	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	ADFamily	Promoter Plan			$ (339.60)			$ -	$ -	$ -	$ (339.60)
25096060	7692912	yourway20c1	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25096318	7692912	yourway20c2	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25096534	7692912	yourway20c3	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25097188	7692912	yourway20c4	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25097496	7692912	yourway20c5	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25098038	7692912	yourway20c6	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25098430	7692912	yourway20c7	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25100598	7692912	yourway20c8	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25101330	7692912	yourway20c9	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25101824	7692912	yourway20c10	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25103046	7692912	yourway21c1	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25103286	7692912	yourway21c2	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25103466	7692912	yourway21c3	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25103880	7692912	yourway21c4	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25104106	7692912	yourway21c5	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25104926	7692912	yourway21c6	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25105154	7692912	yourway21c7	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25105540	7692912	yourway21c8	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25105712	7692912	yourway21c9	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25105912	7692912	yourway21c10	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25109392	7692912	yourway22c1	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25109672	7692912	yourway22c2	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25112000	7692912	yourway22c3	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59	HT	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Exhibit 10: Frantz Balan - Parent ID 7692912 ($519,000 Net Winner) and Lusette Balan - Parent ID 10391181 ($721,000 Net Winner) User Account Detail

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
7692912	$ 518,540.40	211

Frantz and Lusette Balan													**$ -**	**$ 500.00**	**$ (25,534.20)**	**$ 1,490,381.50**	**$ (225,520.00)**	**$ (2,133,288.81)**	**$ 1,285,643.43**	**$ -**	**$ 1,239,827.30**

User Accounts:

													$ -	$ 500.00	$ (22,273.70)	$ 762,984.10	$ (222,670.00)	$ (625,513.96)	$ 323,569.03	$ -	**$ 518,540.40**
Rep. Id	**Parent Id**	**Login**	**Name**	**Email**	**Street**	**Address Number**	**City**	**State/ Province**	**Postal Code**	**Country**	**Phone**	**Account State**	**Participant Type**	**1 - Direct Payment Two Year**	**2 - Direct Receipt Two Year**	**3 - Triangular Payment Two Year**	**4 - Triangular Receipt Two Year**	**5 - Purchase of Credits Two Year**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**8 - Chargebacks**	**Net Equity**
25112156	7692912	yourway22c4	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25112304	7692912	yourway22c5	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25112486	7692912	yourway22c6	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25112684	7692912	yourway22c7	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25112864	7692912	yourway22c8	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25113046	7692912	yourway22c9	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25113198	7692912	yourway22c10	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25114712	7692912	yourway23c1	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25115374	7692912	yourway23c2	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25115546	7692912	yourway23c3	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25115870	7692912	yourway23c4	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25116074	7692912	yourway23c5	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25116200	7692912	yourway23c6	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25116326	7692912	yourway23c7	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25116506	7692912	yourway23c8	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25116642	7692912	yourway23c9	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25116830	7692912	yourway23c10	frantz balan	balan.f@gmail.com	41 rue millien		decmas	NE	59 HT		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25123122	7692912	yourway28c1	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25123268	7692912	yourway28c2	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25123388	7692912	yourway28c3	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25123476	7692912	yourway28c4	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25123592	7692912	yourway28c5	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25123834	7692912	yourway28c6	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25123934	7692912	yourway28c7	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25124138	7692912	yourway28c8	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25124224	7692912	yourway28c9	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25124310	7692912	yourway28c10	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25126318	7692912	yourway29c1	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25126464	7692912	yourway29c2	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25126532	7692912	yourway29c3	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25126592	7692912	yourway29c4	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25126704	7692912	yourway29c5	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25126798	7692912	yourway29c6	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25126906	7692912	yourway29c7	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25127014	7692912	yourway29c8	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25127086	7692912	yourway29c9	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25127170	7692912	yourway29c10	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25127568	7692912	yourway30c4	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25127736	7692912	yourway30c5	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25127848	7692912	yourway30c6	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25127958	7692912	yourway30c7	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25128058	7692912	yourway30c8	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25128154	7692912	yourway30c9	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25128246	7692912	yourway30c10	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25128566	7692912	yourway31c1	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25128640	7692912	yourway31c2	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25128748	7692912	yourway31c3	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25128942	7692912	yourway31c4	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25129014	7692912	yourway31c5	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25129086	7692912	yourway31c6	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25129154	7692912	yourway31c7	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25129248	7692912	yourway31c8	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25129638	7692912	yourway31c9	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25129746	7692912	yourway31c10	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25130320	7692912	yourway27c1	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25130402	7692912	yourway27c2	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25130502	7692912	yourway27c3	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25130576	7692912	yourway27c4	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25130666	7692912	yourway27c5	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25130774	7692912	yourway27c6	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25130896	7692912	yourway27c7	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25130972	7692912	yourway27c8	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25131066	7692912	yourway27c9	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25131170	7692912	yourway27c10	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25132260	7692912	yourway26c6	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25132400	7692912	yourway26c7	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149 US		6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

TelexFree LLC, et. Al --Adversary Proceeding Nos. 16-4006 & 16-4007

Exhibit 10: Frantz Balan - Parent ID 7692912 ($519,000 Net Winner) and Lusette Balan - Parent ID 10391181 ($721,000 Net Winner) User Account Detail

Net Loser Summary:

Parent ID	Total Net Equity	Number of User Accounts
7692912	$ 518,540.40	211

Frantz and Lusette Balan	**$ -**	**$ 500.00**	**$ (25,534.20)**	**$ 1,490,381.50**	**$ (225,520.00)**	**$ (2,133,288.81)**	**$ 1,285,643.43**	**$ -**	**$ 1,239,827.30**										
User Accounts:	$ -	$ 500.00	$ (22,273.70)	$ 762,984.10	$ (222,670.00)	$ (625,513.96)	$ 323,569.03	$ -	**$ 518,540.40**										

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
25132484	7692912	yourway26c8	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25132574	7692912	yourway26c9	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25132688	7692912	yourway26c10	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25132802	7692912	yourway26c5	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25305828	7692912	lsemexantc10	frantz balan	balan.f@gmail.com		12	12	AR	12	US	6172832635	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25306518	7692912	lsemexantc13	frantz balan	balan.f@gmail.com		12	12	AR	12	US	6172832635	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25307246	7692912	lsemexantc14	frantz balan	balan.f@gmail.com		12	12	AR	12	US	6172832635	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
25308148	7692912	lsemexantc15	frantz balan	balan.f@gmail.com		12	12	AR	12	US	6172832635	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
30782550	7692912	fastrosie13	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
30782886	7692912	fastrosie14	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
30783256	7692912	fastrosie15	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
30783676	7692912	fastrosie16	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
30784240	7692912	fastrosie17	frantz balan	balan.f@gmail.com	51 grover st		Everett	MA	2149	US	6172832735	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)

Net Winner Summary:

Parent ID	Total Net Equity	Number of User Accounts
10391181	$ 721,286.90	5

User Accounts:	$ -	$ -	$ (3,260.50)	$ 727,397.40	$ (2,850.00)	$ (1,507,774.85)	$ 962,074.40	$ -	**$ 721,286.90**										

Rep. Id	Parent Id	Login	Name	Email	Street	Address Number	City	State/ Province	Postal Code	Country	Phone	Account State	Participant Type	1 - Direct Payment Two Year	2 - Direct Receipt Two Year	3 - Triangular Payment Two Year	4 - Triangular Receipt Two Year	5 - Purchase of Credits Two Year	6 - Credit Transfers In	7 - Credit Transfers Out	8 - Chargebacks	Net Equity
10391181	10391181	lsemexant	lusette balan	balan.f@gmail.com	51 grover st		everett		2149	US	9787175760	ADCentral	Promoter Plan			$ (1,844.30)	$ 726,858.80	$ (2,850.00)	$ (1,507,774.85)	$ 958,438.75	$ -	$ 722,164.50
10869169	10391181	lsemexant3	lusette balan	frantzbalan@yahoo.com	51 grover st		everett	MA	2149	US	6172832735	Partner	Promoter Plan			$ (50.00)			$ -	$ -	$ -	$ (50.00)
10869917	10391181	lsemexant4	lusette balan	frantzbalan@yahoo.com	51 grover st		Everett		2149	US	6172832735	ADCentral	Promoter Plan			$ (877.50)	$ 339.00		$ -	$ 2,736.54	$ -	$ (538.50)
10870003	10391181	duvivi13c6	lusette balan	balan.f@gmail.com	29 glendale street		everett	MA	2149	US	6173090155	Partner	Phone Plan / Other			$ (49.90)			$ -	$ -	$ -	$ (49.90)
11181769	10391181	lsemexant5	lusette Balan	frantzbalan@yahoo.com	51grover st		Everett	Ma	2149	US	6172832735	ADCentral	Promoter Plan			$ (438.80)	$ 199.60		$ -	$ 899.11	$ -	$ (239.20)

Source(s): Based upon queries of the Huron Data Tables

Exhibit 2

Example Names (Less than Four Characters)

full_name	# of User Accounts	Aggregate Net Loss
l l	40,687	$ (2,063,193.00)
. .	36,498	$ (1,986,986.00)
a a	30,363	$ (1,551,269.00)
.	8,329	$ (535,314.00)
q q	5,339	$ (278,741.00)
w w	3,191	$ (161,427.00)
c c	3,009	$ (151,197.00)
m m	2,077	$ (108,882.00)
x x	2,166	$ (108,732.00)
d d	2,021	$ (102,095.00)
l l	1,966	$ (98,403.00)
Jo	1,849	$ (92,354.00)
a b	1,770	$ (89,421.00)
0 0	1,667	$ (84,082.00)
k k	1,512	$ (76,497.00)
s s	1,455	$ (72,954.00)
j	1,328	$ (68,650.00)
j j	1,298	$ (65,619.00)
p p	1,242	$ (62,076.00)
f f	1,171	$ (59,481.00)

Exhibit 3

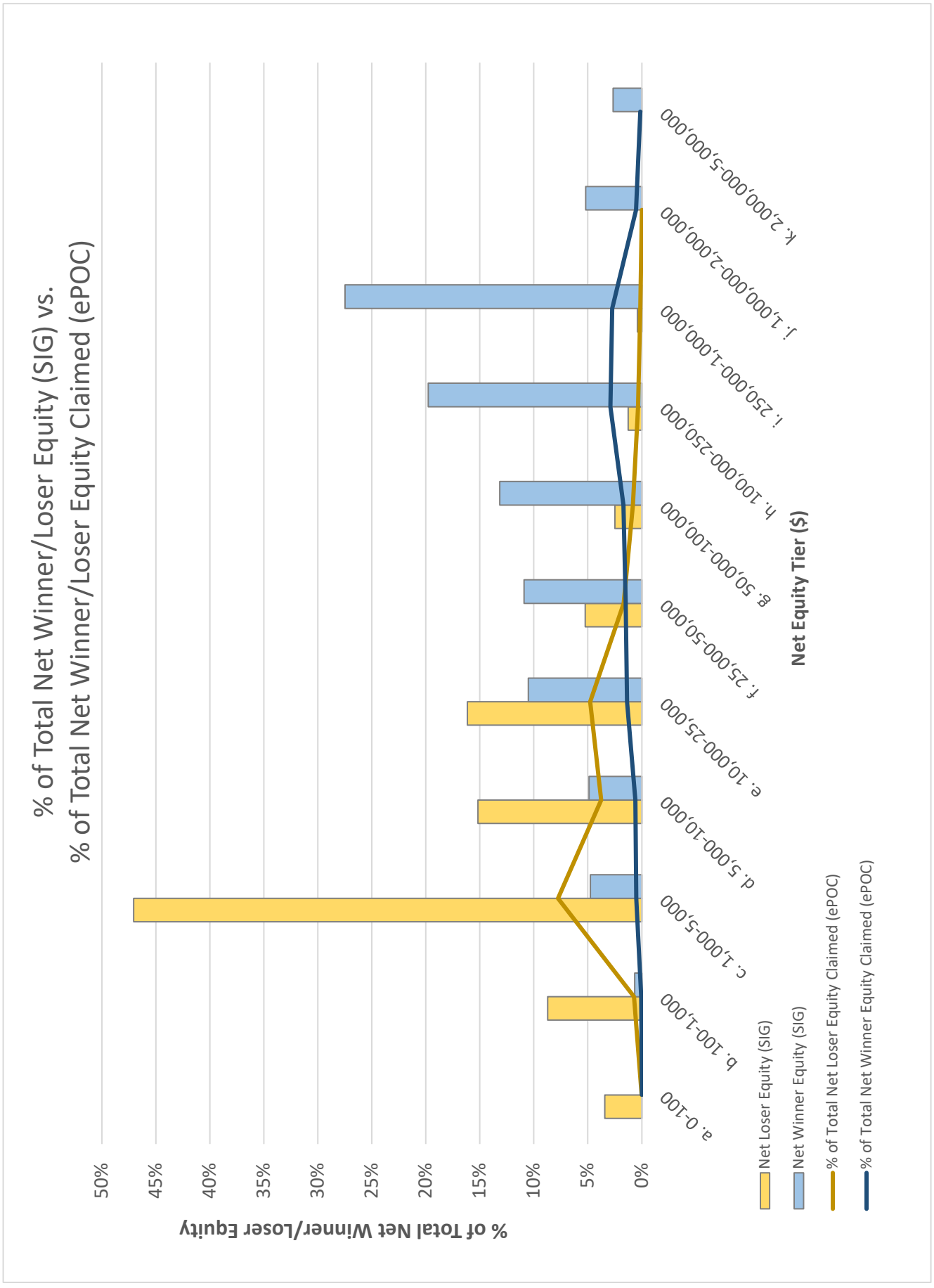

% of Total Net Winner/Loser Equity (SIG) vs.
% of Total Net Winner/Loser Equity Claimed (ePOC)
% of Total Net Winner/Loser Equity
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
a. 0-100
b. 100-1,000
c. 1,000-5,000
d. 5,000-10,000
e. 10,000-25,000
f. 25,000-50,000
g. 50,000-100,000
h. 100,000-250,000
i. 250,000-1,000,000
j. 1,000,000-2,000,000
k. 2,000,000-5,000,000
Net Equity Tier ($)
Net Loser Equity (SIG)
Net Winner Equity (SIG)
% of Total Net Loser Equity Claimed (ePOC)
% of Total Net Winner Equity Claimed (ePOC)

Exhibit 4

Example Impact of Credit Transfers In/Out on Net Equity

Participant Type	Parent ID	Net Equity Excluding Transfers In/Out	*6 - Credit Transfers In*	*7 - Credit Transfers Out*	Net Equity Including Transfers In/Out
Net Loser	13043527	**$ (1,010,727.00)**	*$ (576,124.00)*	*$ 831,217.00*	**$ (755,634.00)**
Net Loser	11849449	**$ (622,342.00)**	*$ (1,904,355.00)*	*$ 2,254,352.00*	**$ (272,344.00)**
Net Loser	12242105	**$ (481,228.00)**	*$ (621,865.00)*	*$ 1,008,813.00*	**$ (94,279.00)**
Net Loser	2876246	**$ (406,591.00)**	*$ (54,842.00)*	*$ 1,243,613.00*	**$ 782,181.00**
Net Loser	1020407	**$ (198,009.00)**	*$ (5,143,557.00)*	*$6,146,96*	**$ 805,396.00**
Net Winner	132785	**$ 4,096,689.00**	*$ (7,287,436.00)*	*$ 6,617,041.00*	**$ 3,426,294.00**
Net Winner	2522900	**$ 4,016,153.00**	*$ (7,128,607.00)*	*$ 3,921,949.00*	**$ 809,495.00**
Net Winner	323967	**$ 3,159,865.00**	$ (5,155,263.00)	$ 3,260,978.00	**$ 1,265,580.00**
Net Winner	10853999	**$ 1,308,377.00**	$ (1,581,410.00)	$ 176,389.00	**$ (96,644.00)**
Net Winner	1826886	**$ 1,163,543.00**	$ (7,882,760.00)	$ 6,592,369.00	**$ (126,847.00)**

Case 16-04007 Doc 608-5 Filed 09/11/23 Entered 09/11/23 16:36:38 Desc Exhibit 5 Page 101 of 109

Exhibit 5

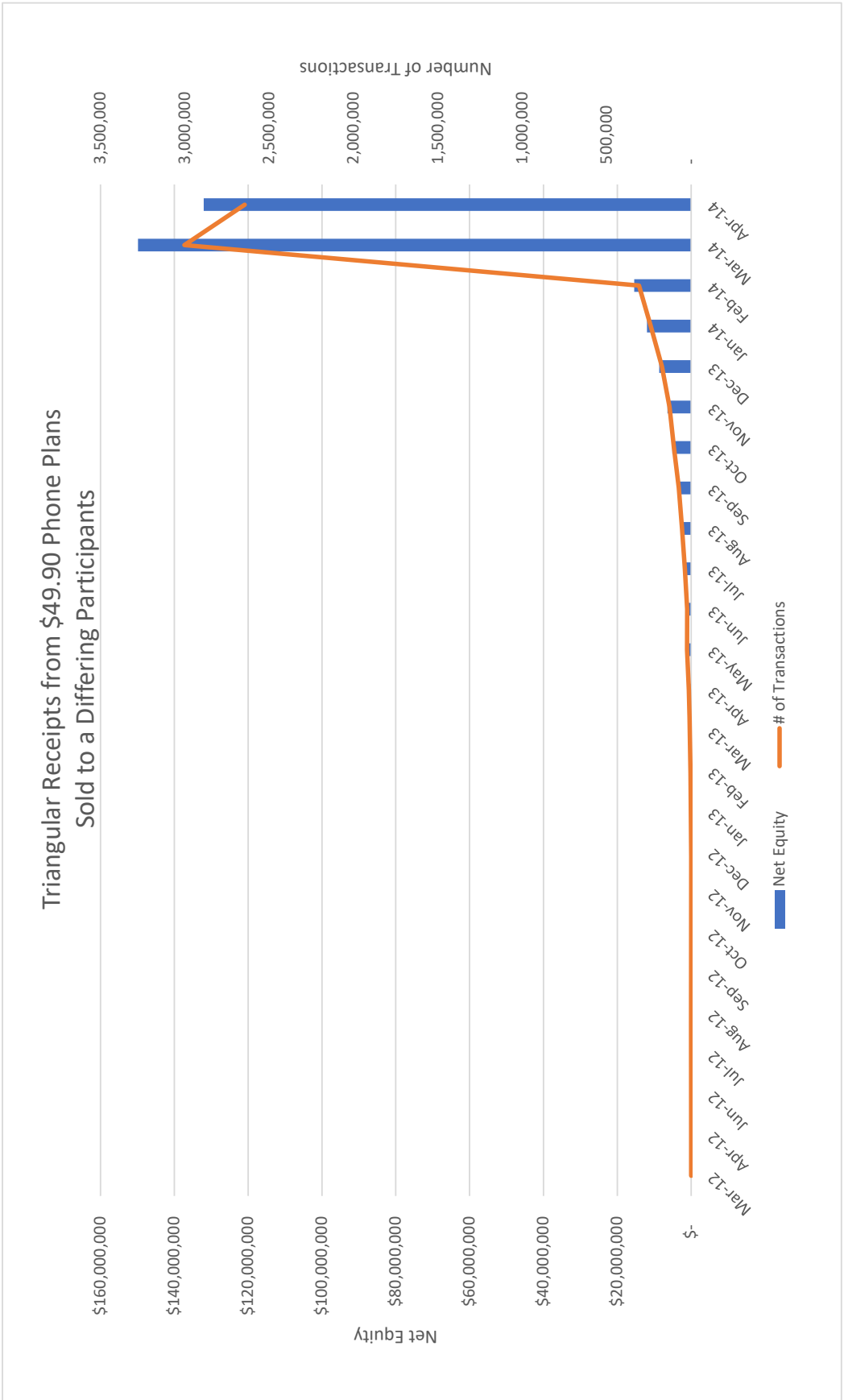

Triangular Receipts from $49.90 Phone Plans
Sold to a Differing Participants
Net Equity
Number of Transactions
Net Equity
# of Transactions
$160,000,000
$140,000,000
$120,000,000
$100,000,000
$80,000,000
$60,000,000
$40,000,000
$20,000,000
$-
3,500,000
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
Mar-12
Apr-12
May-12
Jun-12
Jul-12
Aug-12
Sep-12
Oct-12
Nov-12
Dec-12
Jan-13
Feb-13
Mar-13
Apr-13
May-13
Jun-13
Jul-13
Aug-13
Sep-13
Oct-13
Nov-13
Dec-13
Jan-14
Feb-14
Mar-14
Apr-14

Exhibit 6

Exhibit 3: User Accounts and Participants by Entity and Type

SUMMARY BY ENTITY (ALL):

Entity	User Accounts	Participants	Total Net Equity
Hybrid	138	113	$8,279,788
TelexFree	10,987,607	2,137,074	($226,190,248) *[A]*
Ympactus	4,006,249	1,234,149	($425,439,930)
Total	**14,993,994**	**3,371,336**	**($643,350,389)**
Unique Count	14,993,994	3,327,992	($643,350,389)

SUMMARY BY NET WINNER AND NET LOSER (ALL):

Participant Type	User Accounts	Participants	Total Equity
Net Loser	9,196,908	2,042,022	($1,809,367,883)
Net Winner	1,790,699	95,052	$1,583,177,635
Total	10,987,607	2,137,074	($226,190,248) *[A]*

Source(s): Based on queries of the Huron Data Tables

Case 16-04007 Doc 608-5 Filed 09/11/23 Entered 09/11/23 16:36:38 Desc Exhibit 5 Page 105 of 109

Exhibit 7

Exhibit 5: Summary of TelexFree User Accounts and Participants by Tier

		Huron Data Tables (SIG)								*ePOC Data Tables (Claims)*		
Participant Type	**Tiers**	**User Accounts**	**Avg. User Accounts per Aggregation**	**Avg. Net Equity per User Account**	**Participant Account Aggregations**	**Participant %**	**Avg. Net Equity per Participant**	**Total Equity**	**Total Equity %**	**Cumulative Total Equity %**	**# of Account Aggregations Claimed** [1]	**% of Total Account Aggregations Claimed**	**TelexFree Net Equity Claimed** [1]
All	a. 0-100	**1,270,024**	1	n/a	**1,184,220**	55.4%	n/a	**$ (61,723,699)**	n/a	n/a	6,697	0.31%	n/a
All	b. 100-1,000	**1,734,497**	5	n/a	**383,695**	18.0%	n/a	**$ (147,224,989)**	n/a	n/a	25,476	1.19%	n/a
All	c. 1,000-5,000	**3,296,955**	7	n/a	**472,790**	22.1%	n/a	**$ (776,074,847)**	n/a	n/a	69,839	3.27%	n/a
All	d. 5,000-10,000	**1,517,999**	30	n/a	**50,862**	2.4%	n/a	**$ (197,135,876)**	n/a	n/a	11,146	0.52%	n/a
All	e. 10,000-25,000	**1,638,142**	53	n/a	**30,777**	1.4%	n/a	**$ (125,779,939)**	n/a	n/a	7,301	0.34%	n/a
All	f. 25,000-50,000	**648,612**	84	n/a	**7,758**	0.4%	n/a	**$ 77,905,194**	n/a	n/a	1,602	0.07%	n/a
All	g. 50,000-100,000	**384,286**	105	n/a	**3,674**	0.2%	n/a	**$ 162,929,893**	n/a	n/a	616	0.03%	n/a
All	h. 100,000-250,000	**304,835**	138	n/a	**2,211**	0.1%	n/a	**$ 289,990,232**	n/a	n/a	341	0.02%	n/a
All	i. 250,000-1,000,000	**163,147**	162	n/a	**1,010**	0.0%	n/a	**$ 427,578,150**	n/a	n/a	114	0.01%	n/a
All	j. 1,000,000-2,000,000	**20,662**	339	n/a	**61**	0.0%	n/a	**$ 81,200,318**	n/a	n/a	6	0.00%	n/a
All	k. 2,000,000-5,000,000	**8,453**	528	n/a	**16**	0.0%	n/a	**$ 42,145,315**	n/a	n/a	2	0.00%	n/a
All	l. 5,000,000+	**-**	-	n/a	**-**	0.0%	n/a	**$ -**	n/a	n/a	-	0.00%	n/a
Total All		**10,987,612**	**5**	n/a	**2,137,074**	**100.0%**	n/a	**$ (226,190,248)**	n/a	n/a	**123,140**	**5.76%**	n/a

Total Number of Unique User with Claims [1] **85,616**
Est. Percentage of Participants that Submitted Claims **4.01%**

		Huron Data Tables (SIG)								*ePOC Data Tables (Claims)*		
Participant Type	**Tiers**	**User Accounts**	**Avg. User Accounts per Aggregation**	**Avg. Net Equity per User Account**	**Participant Account Aggregations**	**Participant %**	**Avg. Net Equity per Participant**	**Total Equity**	**Total Equity %**	**Cumulative Total Equity %**	**# of Account Aggregations Claimed** [1]	**% of Total Account Aggregations Claimed**	**TelexFree Net Equity Claimed** [1]
Net Loser	a. 0-100	**1,243,515**	1	$ (50)	**1,175,480**	57.6%	$ (53)	**$ (62,133,114)**	3.4%	3.4%	6,031	0.28%	$ (286,630)
Net Loser	b. 100-1,000	**1,608,799**	4	$ (98)	**360,150**	17.6%	$ (438)	**$ (157,629,250)**	8.7%	12.1%	23,430	1.10%	$ (12,906,466)
Net Loser	c. 1,000-5,000	**2,986,680**	7	$ (285)	**442,457**	21.7%	$ (1,924)	**$ (851,345,539)**	47.1%	59.2%	66,628	3.12%	$ (140,408,781)
Net Loser	d. 5,000-10,000	**1,305,076**	33	$ (210)	**39,983**	2.0%	$ (6,866)	**$ (274,527,424)**	15.2%	74.4%	9,822	0.46%	$ (68,014,478)
Net Loser	e. 10,000-25,000	**1,302,038**	64	$ (224)	**20,233**	1.0%	$ (14,444)	**$ (292,252,263)**	16.2%	90.5%	5,948	0.28%	$ (86,456,229)
Net Loser	f. 25,000-50,000	**401,127**	141	$ (236)	**2,853**	0.1%	$ (33,229)	**$ (94,801,934)**	5.2%	95.8%	921	0.04%	$ (30,528,785)
Net Loser	g. 50,000-100,000	**192,764**	283	$ (235)	**682**	0.0%	$ (66,331)	**$ (45,237,836)**	2.5%	98.3%	223	0.01%	$ (14,791,439)
Net Loser	h. 100,000-250,000	**117,007**	705	$ (196)	**166**	0.0%	$ (137,988)	**$ (22,906,058)**	1.3%	99.5%	44	0.00%	$ (6,058,446)
Net Loser	i. 250,000-1,000,000	**38,671**	2,275	$ (195)	**17**	0.0%	$ (442,573)	**$ (7,523,739)**	0.4%	99.9%	6	0.00%	$ (2,454,033)
Net Loser	j. 1,000,000-2,000,000	**1,232**	1,232	$ (820)	**1**	0.0%	$ (1,010,727)	**$ (1,010,727)**	0.1%	100.0%	-	0.00%	$ -
Net Loser	k. 2,000,000-5,000,000	**-**	-	$ -	**-**	0.0%	$ -	**$ -**	0.0%	100.0%	-	0.00%	$ -
Net Loser	l. 5,000,000+	**-**	-	$ -	**-**	0.0%	$ -	**$ -**	0.0%	100.0%	-	0.00%	$ -
Total Net Losers		**9,196,909**	**5**	$ (197)	**2,042,022**	**100.0%**	$ (886)	**$ (1,809,367,883)**	**100.0%**	**100.0%**	**113,053**	**5.29%**	**$ (361,905,287)**

		Huron Data Tables (SIG)								*ePOC Data Tables (Claims)*		
Participant Type	**Tiers**	**User Accounts**	**Avg. User Accounts per Aggregation**	**Avg. Net Equity per User Account**	**Participant Account Aggregations**	**Participant %**	**Avg. Net Equity per Participant**	**Total Equity**	**Total Equity %**	**Cumulative Total Equity %**	**# of Account Aggregations Claimed** [1]	**% of Total Account Aggregations Claimed**	**TelexFree Net Equity Claimed** [1]
Net Winner	a. 0-100	**26,509**	3	$ 15	**8,740**	9.2%	$ 47	**$ 409,415**	0.0%	0.0%	666	0.03%	$ 34,594
Net Winner	b. 100-1,000	**125,698**	5	$ 83	**23,545**	24.8%	$ 442	**$ 10,404,261**	0.7%	0.7%	2,046	0.10%	$ 944,918
Net Winner	c. 1,000-5,000	**310,275**	10	$ 243	**30,333**	31.9%	$ 2,481	**$ 75,270,692**	4.8%	5.4%	3,211	0.15%	$ 8,051,163
Net Winner	d. 5,000-10,000	**212,923**	20	$ 363	**10,879**	11.4%	$ 7,114	**$ 77,391,548**	4.9%	10.3%	1,324	0.06%	$ 9,472,582
Net Winner	e. 10,000-25,000	**336,104**	32	$ 495	**10,544**	11.1%	$ 15,788	**$ 166,472,324**	10.5%	20.8%	1,353	0.06%	$ 21,391,857
Net Winner	f. 25,000-50,000	**247,485**	50	$ 698	**4,905**	5.2%	$ 35,210	**$ 172,707,128**	10.9%	31.7%	681	0.03%	$ 23,711,122
Net Winner	g. 50,000-100,000	**191,522**	64	$ 1,087	**2,992**	3.1%	$ 69,575	**$ 208,167,729**	13.1%	44.9%	393	0.02%	$ 26,847,712
Net Winner	h. 100,000-250,000	**187,828**	92	$ 1,666	**2,045**	2.2%	$ 153,006	**$ 312,896,290**	19.8%	64.7%	297	0.01%	$ 45,959,512
Net Winner	i. 250,000-1,000,000	**124,476**	125	$ 3,495	**993**	1.0%	$ 438,169	**$ 435,101,889**	27.5%	92.1%	108	0.01%	$ 43,464,757
Net Winner	j. 1,000,000-2,000,000	**19,430**	324	$ 4,231	**60**	0.1%	$ 1,370,184	**$ 82,211,045**	5.2%	97.3%	6	0.00%	$ 8,316,422
Net Winner	k. 2,000,000-5,000,000	**8,453**	528	$ 4,986	**16**	0.0%	$ 2,634,082	**$ 42,145,315**	2.7%	100.0%	2	0.00%	$ 2,034,063
Net Winner	l. 5,000,000+	**-**	-	$ -	**-**	0.0%	$ -	**$ -**	0.0%	100.0%	-	0.00%	$ -
Total Net Winners		**1,790,703**	**19**	$ 884	**95,052**	**100.0%**	$ 16,656	**$ 1,583,177,635**	**100.0%**	**100.0%**	**10,087**	**0.47%**	**$ 190,228,702**

[1] Includes Participant Account Aggregations where at least one underlying User Account matched to an Account Information Set, and which was included as part of an accepted and agreed claim within ePOC.

Source(s): Based on queries of the Huron Data Tables and ePOC Data Tables.

Case 16-04007 Doc 608-5 Filed 09/11/23 Entered 09/11/23 16:36:38 Desc Exhibit 5 Page 107 of 109

Exhibit 8

Exhibit 4: Summary of TelexFree Net Equity by Tier and Transaction Type

				Included in Net Equity						*Excluded from Net Equity*		
Participant Type	**Tiers**	**Total Equity, Including Transfers**	**Total Net Equity**	**1 - Direct Payment**	**2 - Direct Receipt**	**3 - Triangular Payment**	**4 - Triangular Receipt**	**5 - Purchase of Credits**	**8 - Chargebacks**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**5 - Purchase of Credits (Excluded)**
All	a. 0-100	$ (55,883,729)	**$ (61,723,699)**	$ (6,350,040)	$ 1,771,231	$ (87,483,965)	$ 28,860,552	$ (72,529)	$ 1,551,051	$ (22,021,085)	$ 27,861,055	$ (45,500)
All	b. 100-1,000	$ (85,204,419)	**$ (147,224,989)**	$ (29,396,673)	$ 13,221,676	$ (261,771,166)	$ 129,250,382	$ (448,168)	$ 1,918,960	$ (86,034,990)	$ 148,055,560	$ -
All	c. 1,000-5,000	$ (322,041,264)	**$ (776,074,847)**	$ (149,438,063)	$ 33,008,696	$ (1,048,338,884)	$ 384,575,001	$ (1,953,492)	$ 6,071,896	$ (354,649,674)	$ 808,683,257	$ (32,184,929)
All	d. 5,000-10,000	$ (112,840,158)	**$ (197,135,876)**	$ (53,831,028)	$ 20,589,709	$ (405,641,383)	$ 239,935,281	$ (1,718,329)	$ 3,529,873	$ (252,761,851)	$ 337,057,570	$ (1,550,948)
All	e. 10,000-25,000	$ 123,694,778	**$ (125,779,939)**	$ (63,862,445)	$ 26,980,870	$ (455,389,382)	$ 366,849,804	$ (3,670,979)	$ 3,312,194	$ (413,379,390)	$ 662,854,107	$ (237,321)
All	f. 25,000-50,000	$ 42,326,042	**$ 77,905,194**	$ (23,646,642)	$ 18,121,810	$ (186,143,877)	$ 271,167,415	$ (2,967,847)	$ 1,374,335	$ (312,261,698)	$ 276,682,546	$ (2,946,374)
All	g. 50,000-100,000	$ 87,101,808	**$ 162,929,893**	$ (16,085,950)	$ 13,965,856	$ (114,955,642)	$ 282,418,175	$ (3,394,710)	$ 982,165	$ (337,639,337)	$ 261,811,252	$ (14,326,425)
All	h. 100,000-250,000	$ 120,253,309	**$ 289,990,232**	$ (9,034,426)	$ 10,572,174	$ (84,739,775)	$ 378,188,218	$ (5,588,973)	$ 593,014	$ (429,514,497)	$ 259,777,574	$ (7,650,026)
All	i. 250,000-1,000,000	$ 211,394,198	**$ 427,578,150**	$ (6,529,505)	$ 11,200,997	$ (53,645,439)	$ 482,274,368	$ (6,023,635)	$ 301,365	$ (622,120,110)	$ 405,936,158	$ (29,957,640)
All	j. 1,000,000-2,000,000	$ 48,968,394	**$ 81,200,318**	$ (1,349,415)	$ 2,608,874	$ (9,849,404)	$ 92,296,947	$ (2,517,278)	$ 10,595	$ (141,665,881)	$ 109,433,957	$ (17,317,435)
All	k. 2,000,000-5,000,000	$ 19,248,445	**$ 42,145,315**	$ (260,818)	$ 540,114	$ (5,719,817)	$ 49,405,425	$ (1,827,871)	$ 8,283	$ (64,010,771)	$ 41,113,901	$ (93,375)
All	l. 5,000,000+	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Total All		**$ 77,017,406**	**$ (226,190,248)**	**$ (359,785,006)**	**$ 152,582,007**	**$ (2,713,678,735)**	**$ 2,705,221,567**	**$ (30,183,810)**	**$ 19,653,730**	**$ (3,036,059,284)**	**$ 3,339,266,938**	**$ (106,309,974)**
		Number of Transactions		**440,969**	**126,599**	**13,333,462**	**13,320,684**	**5,281**	**5,064**	**4,461,778**	**4,461,866**	**747**

				Included in Net Equity						*Excluded from Net Equity*		
Participant Type	**Tiers**	**Total Net Equity, Including Transfers**	**Total Net Equity**	**1 - Direct Payment**	**2 - Direct Receipt**	**3 - Triangular Payment**	**4 - Triangular Receipt**	**5 - Purchase of Credits**	**8 - Chargebacks**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**5 - Purchase of Credits (Excluded)**
Net Loser	a. 0-100	$ (59,177,305)	**$ (62,133,114)**	$ (3,731,914)	$ 824,395	$ (76,423,539)	$ 16,781,894	$ (45,325)	$ 461,374	$ (12,932,413)	$ 15,888,222	$ -
Net Loser	b. 100-1,000	$ (108,488,471)	**$ (157,629,250)**	$ (20,492,949)	$ 6,559,807	$ (215,180,908)	$ 71,001,627	$ (264,470)	$ 747,644	$ (41,868,816)	$ 91,009,594	$ -
Net Loser	c. 1,000-5,000	$ (487,578,348)	**$ (851,345,539)**	$ (129,357,455)	$ 13,410,325	$ (939,344,760)	$ 202,855,660	$ (954,985)	$ 2,045,676	$ (186,812,864)	$ 550,580,055	$ (32,184,929)
Net Loser	d. 5,000-10,000	$ (167,756,527)	**$ (274,527,424)**	$ (42,171,409)	$ 6,659,952	$ (338,497,888)	$ 99,662,327	$ (830,344)	$ 649,938	$ (111,859,512)	$ 218,630,409	$ (30,407)
Net Loser	e. 10,000-25,000	$ 30,792,460	**$ (292,252,263)**	$ (47,989,538)	$ 5,567,997	$ (352,476,595)	$ 104,116,395	$ (1,882,920)	$ 412,398	$ (142,218,490)	$ 465,263,213	$ (51,719)
Net Loser	f. 25,000-50,000	$ (50,272,007)	**$ (94,801,934)**	$ (13,452,862)	$ 1,842,080	$ (115,070,596)	$ 32,900,855	$ (1,071,633)	$ 50,223	$ (54,352,327)	$ 98,882,254	$ (101,800)
Net Loser	g. 50,000-100,000	$ (14,601,445)	**$ (45,237,836)**	$ (7,721,437)	$ 1,012,269	$ (55,633,454)	$ 18,424,444	$ (1,364,233)	$ 44,575	$ (32,823,309)	$ 63,459,700	$ (5,360,627)
Net Loser	h. 100,000-250,000	$ (9,640,360)	**$ (22,906,058)**	$ (2,114,009)	$ 227,488	$ (28,399,952)	$ 8,053,996	$ (673,580)	$ -	$ (19,047,886)	$ 32,313,584	$ -
Net Loser	i. 250,000-1,000,000	$ (2,133,904)	**$ (7,523,739)**	$ (808,729)	$ 56,583	$ (7,898,568)	$ 1,553,800	$ (426,825)	$ -	$ (5,288,800)	$ 10,678,635	$ -
Net Loser	j. 1,000,000-2,000,000	$ (755,634)	**$ (1,010,727)**	$ -	$ 1,462	$ (1,457,939)	$ 445,750	$ -	$ -	$ (576,124)	$ 831,217	$ -
Net Loser	k. 2,000,000-5,000,000	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Net Loser	l. 5,000,000+	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Total Net Losers		**$ (869,611,541)**	**$ (1,809,367,883)**	**$ (267,840,302)**	**$ 36,162,357**	**$ (2,130,384,199)**	**$ 555,796,749**	**$ (7,514,316)**	**$ 4,411,828**	**$ (607,780,542)**	**$ 1,547,536,883**	**$ (37,729,482)**

				Included in Net Equity						*Excluded from Net Equity*		
Participant Type	**Tiers**	**Total Net Equity, Including Transfers**	**Total Net Equity**	**1 - Direct Payment**	**2 - Direct Receipt**	**3 - Triangular Payment**	**4 - Triangular Receipt**	**5 - Purchase of Credits**	**8 - Chargebacks**	**6 - Credit Transfers In**	**7 - Credit Transfers Out**	**5 - Purchase of Credits (Excluded)**
Net Winner	a. 0-100	$ 3,293,577	**$ 409,415**	$ (2,618,127)	$ 946,837	$ (11,060,427)	$ 12,078,658	$ (27,204)	$ 1,089,677	$ (9,088,671)	$ 11,972,833	$ (45,500)
Net Winner	b. 100-1,000	$ 23,284,052	**$ 10,404,261**	$ (8,903,724)	$ 6,661,869	$ (46,590,258)	$ 58,248,755	$ (183,697)	$ 1,171,316	$ (44,166,174)	$ 57,045,966	$ -
Net Winner	c. 1,000-5,000	$ 165,537,084	**$ 75,270,692**	$ (20,080,608)	$ 19,598,370	$ (108,994,124)	$ 181,719,341	$ (998,507)	$ 4,026,220	$ (167,836,810)	$ 258,103,202	$ -
Net Winner	d. 5,000-10,000	$ 54,916,370	**$ 77,391,548**	$ (11,659,619)	$ 13,929,757	$ (67,143,495)	$ 140,272,954	$ (887,984)	$ 2,879,935	$ (140,902,339)	$ 118,427,161	$ (1,520,542)
Net Winner	e. 10,000-25,000	$ 92,902,319	**$ 166,472,324**	$ (15,872,907)	$ 21,412,873	$ (102,912,788)	$ 262,733,409	$ (1,788,059)	$ 2,899,796	$ (271,160,900)	$ 197,590,894	$ (185,602)
Net Winner	f. 25,000-50,000	$ 92,598,049	**$ 172,707,128**	$ (10,193,780)	$ 16,279,731	$ (71,073,281)	$ 238,266,561	$ (1,896,214)	$ 1,324,112	$ (257,909,371)	$ 177,800,292	$ (2,844,574)
Net Winner	g. 50,000-100,000	$ 101,703,254	**$ 208,167,729**	$ (8,364,513)	$ 12,953,587	$ (59,322,188)	$ 263,993,731	$ (2,030,477)	$ 937,589	$ (304,816,028)	$ 198,351,552	$ (8,965,798)
Net Winner	h. 100,000-250,000	$ 129,893,669	**$ 312,896,290**	$ (6,920,417)	$ 10,344,686	$ (56,339,822)	$ 370,134,222	$ (4,915,393)	$ 593,014	$ (410,466,611)	$ 227,463,990	$ (7,650,026)
Net Winner	i. 250,000-1,000,000	$ 213,528,102	**$ 435,101,889**	$ (5,720,776)	$ 11,144,414	$ (45,746,872)	$ 480,720,568	$ (5,596,810)	$ 301,365	$ (616,831,310)	$ 395,257,523	$ (29,957,640)
Net Winner	j. 1,000,000-2,000,000	$ 49,724,028	**$ 82,211,045**	$ (1,349,415)	$ 2,607,412	$ (8,391,465)	$ 91,851,197	$ (2,517,278)	$ 10,595	$ (141,089,757)	$ 108,602,740	$ (17,317,435)
Net Winner	k. 2,000,000-5,000,000	$ 19,248,445	**$ 42,145,315**	$ (260,818)	$ 540,114	$ (5,719,817)	$ 49,405,425	$ (1,827,871)	$ 8,283	$ (64,010,771)	$ 41,113,901	$ (93,375)
Net Winner	l. 5,000,000+	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Total Net Winners		**$ 946,628,948**	**$ 1,583,177,635**	**$ (91,944,704)**	**$ 116,419,649**	**$ (583,294,536)**	**$ 2,149,424,818**	**$ (22,669,494)**	**$ 15,241,902**	**$ (2,428,278,743)**	**$ 1,791,730,055**	**$ (68,580,492)**

Source(s): Based on queries of the Huron Data Tables

Case 16-04007 Doc 608-5 Filed 09/11/23 Entered 09/11/23 16:36:38 Desc Exhibit 5 Page 109 of 109

EXHIBIT 6

UNITED STATES BANKRUPTCY COURT
DISTRICT OF MASSACHUSETTS

IN THE MATTER OF:	:	Case No. 14-40987 14-40988, 14-40989
TELEXFREE, LLC, TELEXFREE INC., and TELEXFREE FINANCIAL, INC.,	:	(Jointly Administered) Boston, Massachusetts **November 24, 2020**
Debtors,	:	10:00:34 a.m.
: : :

was said that was bought or sold, that we knew it was the security that was bought or sold.

Q And when you say came out of financial systems, what do you mean?

A In that particular instance, one was a company's actual financial system that, that records the purchase and sales. The other was a, was, actually, a FINRA dataset, which is the submission of, of financial records to, to FINRA.

Q Okay. Meaning were these system-generated transaction datasets?

A They were.

Q Okay. Not -- were they user-entered datasets?

A No.

Q Okay.

What is -- could you describe a, what a probabilistic record linkage is?

A Sure.

So, so a probabilistic record linkage uses math and statistics to compare instances where there aren't high quality unique identifiers and does that by creating weights that suggest whether a single variable is, is, is indicative or not indicative in the population of a match.

So what it allows you to do is precisely what we have here where there is, are small, there can be small details in the data like misspellings and other things like that. It can

account for those kind of discrepancies and also account for issues when there, some combination of things isn't inherently distinct, which is where deterministic is, is most helpful.

Q And, and when you refer to non, well, slight differences, is, is that what is sometimes referred to as "fuzzy" matching?

A It can be. So, for example, if I'm trying to, if in one hand I have the name StoneTurn and in another instance I have this, the name StoneBurn -- so there's a one-letter difference -- you can use algorithms like Levenshtein distance or it's called edit distance to look at the number of insertions, deletions, or substitutions you need to make to get from Name A to Name B. And what it does is create a score from 0 to 1 to suggest how close those two strings are.

So you can, what you can see is it's not a binary yes/no as it would be in the form of a deterministic approach, but put things on a spectrum that you can then look at.

Q And what level -- you said there's a score from 0 to 1. What level of subjectivity is that score, meaning is someone looking at something and assigning it or is that determined some other way?

A No, no. That, that's a score that is calculated.

Q Okay. And, and how is it calculated, meaning what principles are used to make that calculation?

A So in, in that case you're using, actually, statistical -- well, as I noted, you're using the number, it's actually a

probabilistic methodology to apply to the TelexFree case?
A No, I didn't.
Q And I take it from that, then, we can't -- we -- you -- you made, drew no conclusion, then, as to whether, if a probabilistic report [sic] had been used, how much it may have varied from Mr. Martin's findings using the deterministic, the basic deterministic approach that he used?
A I'm sorry. What -- you're saying I -- I -- can I compare the probabilistic method linkage that I didn't do to Mr. Martin's?
Q What I'm saying is, you can't compare the two results 'cause you didn't do one of them?
A That, that's right. My assignment --
Q That's true?
A -- was not to re-create a new methodology and a new approach from scratch.
Q Right. You didn't --
A Just identify issues in the existing methodology.
Q So you can't tell us whether, in fact, if Mr. Martin used a probabilistic methodology, it would have any material impact upon his findings of the various linking of user accounts?
A I don't know that that's true. I think I very much identified specific case examples that would fit the conditions of a probabilistic record linkage and would be helped under that, that approach.

EXHIBIT 7

BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION
Statistical Research Report Series
No. RR2000/05

Using the EM Algorithm for Weight Computation in the Fellegi-Sunter Model of Record Linkage

William E. Winkler
Statistical Research Division
Methodology and Standards Directorate
U.S. Bureau of the Census
Washington D.C. 20233

Report Issued: October 4, 2000

USING THE EM ALGORITHM FOR WEIGHT COMPUTATION IN THE FELLEGI-SUNTER MODEL OF RECORD LINKAGE

William E. Winkler, william.e.winkler@census.gov
Bureau of the Census

ABSTRACT

Let **A**×**B** be the product space of two sets **A** and **B** which is divided into a (pairs representing the same entity) and *nonmatches* (pairs representing different entities). Linkage rules are those that divide **A**×**B** into *links* (designated matches), *possible links* (pairs for which we delay a decision), and *nonlinks* (designated nonmatches). Under fixed bounds on the error rates, Fellegi and Sunter (1969) provided a linkage rule that is optimal in the sense that it minimizes the set of possible links. The optimality is dependent on knowledge of certain joint inclusion probabilities that are used in a crucial likelihood ratio. In applying the record linkage model, assumptions are often made that allow estimation of weights that are a function of the joint inclusion probabilities. If the assumptions are not met, then the linkage procedure using estimates computed under the assumptions may not be optimal. This paper describes a method for estimating weights using the EM Algorithm under less restrictive assumptions. The weight computation automatically incorporates a Bayesian adjustment based on file characteristics.

Keywords and phrases. Decision rule, error rate.

1. INTRODUCTION

The paper describes a method for using the EM Algorithm (Dempster, Laird, and Rubin 1977, Wu 1983) to improve computational procedures in applications of the Fellegi-Sunter model of record linkage.

Let **A**×**B** be the product space of two sets **A** and **B** which is divided into *matches* (pairs representing the same entity) and *nonmatches* (pairs representing different entities). Linkage rules are those that divide **A**×**B** into *links* (designated matches), possible links (pairs for which we delay a decision), and *nonlinks* (designated nonmatches). Under fixed bounds on the error rates, Fellegi and Sunter (1969, hereafter denoted FS) provided a linkage rule that is optimal in the sense that it minimizes the set of possible links. The optimality is dependent on knowledge of certain joint inclusion probabilities that are used in a crucial likelihood ratio.

In applications, an independence assumption is made that allows estimation of joint inclusion probabilities. If the independence assumption is not valid (Winkler 1985, Kelley 1986), then linkage rules based on the estimated probabilities may not be optimal.

The remainder of this paper contains a methodology for estimating weights under a less restrictive assumption. Section two is divided into four parts. The first part provides a summary of the FS Model of record linkage. The second describes the Conditional Independence Assumption and how computation is simplified under it. The third part introduces a more general class of distributions than those satisfying the Conditional Independence Assumption. For the new class it shows how simple agreement/disagreement weights are computed using the EM Algorithm. The fourth part presents a procedure for deriving frequency-based weights when an additional assumption is met.

The assumption is weaker than the assumption of FS (pp. 1207-1210).

The discussion in the third section comprises four components. The first describes convergence properties of the EM Algorithm. The second describes how the EM Algorithm automatically makes as Bayesian adjustment based on file characteristics. The third discusses the computation of frequency-based weights. The fourth points out how the EM Algorithm, with possible enormous increase in computation, can be extended to parameter estimation for a reasonably general class of distributions. The final section is a summary.

2. MODEL AND COMPUTATIONAL PROCEDURES

2.1. Fellegi-Sunter Model

The FS Model uses an decision-theoretic approach embodying principles first used in practice by Newcombe (Newcombe et al. 1959). To give an overview, we describe the model in terms of ordered pairs in a product space. The presentation closely follows FS (pp. 1184-1187).

There are two populations **A** and **B** whose elements will be denoted by a and b. We assume that some elements are common to **A** and **B**. Consequently the set of ordered pairs

$$\mathbf{A}\times\mathbf{B} = \{(a,b): a\varepsilon\mathbf{A}, b\varepsilon\mathbf{B}\}$$

is the union of two disjoint sets of *matches*

$$M = \{(a,b): a=b, a\varepsilon\mathbf{A}, b\varepsilon\mathbf{B}\}$$

and *nonmatches*

$$U = \{(a,b): a\neq b, a\varepsilon\mathbf{A}, b\varepsilon\mathbf{B}\}.$$

The records corresponding to **A** and **B** are denoted by $\alpha(a)$ and $ß(b)$, respectively. The *comparison vector* τ associated with the records is defined by:

$$\tau[(\alpha(a),ß(b))] \equiv \{\tau^1[(\alpha(a),ß(b))],\tau^2[(\alpha(a),ß(b))],\cdots,\tau^K[(\alpha(a),ß(b))]\}.$$

Where confusion does not arise, the function τ on $\mathbf{A}\times\mathbf{B}$ will be denoted by $\tau(\alpha,ß)$, $\tau(a,b)$, or τ. The set of all possible realizations of τ is denoted by Γ.

The conditional probability of $\tau(a,b)$ if $(a,b)\varepsilon M$ is given by

$$m(\tau) \equiv P\{\tau[\alpha(a),ß(b)]|(a,b)\varepsilon M\}$$

$$= \sum_{(a,b)\varepsilon M} P\{\tau[\alpha(a),ß(b)]\}\cdot P[(a,b)|M].$$

Similarly we denote the conditional probability of τ if $(a,b)\varepsilon U$ by $u(\tau)$.

We observe a vector of information $\tau(a,b)$ associated with pair (a,b) and wish to designate a pair as a link (in set A_1), a possible link (in set A_2), or a nonlink (in set A_3). We let L denote a linkage rule that divides $\mathbf{A}\times\mathbf{B}$ into A_1, A_2, and A_3. We say that a *Type I error* has occurred if rule L places $m\varepsilon M$ in A_3,

$$P(A_3|M) = \sum_{\tau\varepsilon\Gamma} m(\tau)\cdot P(A_3|\tau),$$

and a *Type II error* if L places $u\varepsilon U$ in A_1,

$$P(A_1|U) = \sum_{\tau\varepsilon\Gamma} u(\tau)\cdot P(A_1|\tau).$$

FS define a linkage rule L_0 with associated sets A_1, A_2, and A_3 that is optimal in the following sense:

THEOREM (Fellegi and Sunter 1969). Let L' be a linkage rule with associated sets A_1', A_2', and A_3' such that $P(A_3'|M) = P(A_3|M)$ and $P(A_1'|U) = P(A_1|U)$. Then $P(A_2|U) \leq P(A_2'|U)$ and $P(A_2|M) \leq P(A_2'|M)$.

In other words, if L' is any competitor of L_0 having the same Type I and Type II error rates (which are both conditional probabilities), then the conditional probabilities (either on set U or M) of not making a decision under rule L' is always greater than under L_0. The FS linkage rule is actually optimal with respect to any set Q of ordered pairs in $\mathbf{A}\times\mathbf{B}$ if we define error probabilities P_Q and a linkage rule L_Q conditional on Q. Thus, it may be possible to define subsets of $\mathbf{A}\times\mathbf{B}$ on which we make use of differing amounts and types of available information.

2.2. Computational Procedures

The subsection is divided into four parts. The first describes the general form of the linkage rule. The second presents a simplification of the computational procedures under the Conditional Independence Assumption. The third contains a weaker assumption and the estimation of parameters using the EM Algorithm. The fourth extends parameter estimation to frequency-based weights.

2.2.1. General Form of Linkage Rule

To provide a background for understanding why specific computational procedures are used, we consider the following likelihood ratio

$$R \equiv R[\tau(a,b)] = m(\tau)/u(\tau). \qquad (2.1)$$

If the numerator is positive and the denominator is zero in (2.1), we assign a fixed very large number to the ratio. The FS linkage rule takes the form:

If $R > T_\mu$, then denote (a,b) as a link.

If $T_\lambda \leq R \leq T_\mu$, then denote (a,b) as a possible link. (2.2)

If $R < T_\lambda$, then denote (a,b) as a nonlink.

The cutoffs T_λ and T_μ are determined by the desired error rate bounds μ and λ on the false match rates and false nonmatch rates, respectively.

2.2.2. Simplification Under Conditional Independence Assumption

In practice, computation is simplified two ways. The first is by the Conditional Independence Assumption of FS:

$$m(\tau) = m_1(\tau^1)\cdot m_2(\tau^2)\cdots m_K(\tau^K) \text{ and}$$

$$u(\tau) = u_1(\tau^1)\cdot u_2(\tau^2)\cdots u_K(\tau^K)$$

where for $i = 1, 2, \cdots, K$

$$m_i(\tau^i) = P(\tau^i \mid (a,b)\varepsilon M) \text{ and}$$

$$u_i(\tau^i) = P(\tau^i \mid (a,b)\varepsilon U).$$

The second is to use a computationally convenient function of the ratio in (2.1). Log_2 is used. We then have

$$W \equiv W(\tau) = \text{Log}_2[m(\tau)/u(\tau)] = W^1 + W^2 + \cdots + W^K, \tag{2.3}$$

where $W^i \equiv \text{Log}_2[m_i(\tau^i)/u_i(\tau^i)]$ for $i = 1, 2, \cdots, K$. We call W the *total comparison weight* associated with a pair and W^i, $i = 1, 2, \cdots, K$, the *individual comparison weights*.

2.2.3. Weaker Assumption and EM Algorithm

To describe the computational assumption that is weaker than the Conditional Independence Assumption, we need some additional background. For the remainder of the paper, unless otherwise stated, we will assume that each component τ^i, $i = 1, 2, \cdots K$, in τ represents a two state comparison (e.g., agree/disagree) and define the marginal comparison events by

$$B^i \equiv \{(a,b) \mid \tau^i(a,b) = \tau^i_o\}$$

for one fixed state of τ^i_o.

Let $C_1, C_2, \cdots, C_K$ be any reordering of $B_1, B_2, \cdots, B_K$. On any set of pairs Q in $\mathbf{A}\times\mathbf{B}$

$$P(\tau\varepsilon C_1\cap C_2\cap\cdots\cap C_K \mid Q) =$$

$$P(\tau\varepsilon C_1 \mid Q)\cdot P(\tau\varepsilon C_2 \mid C_1, Q)\cdots P(\tau\varepsilon C_K \mid C_1,\cdots,C_{K-1},Q). \tag{2.4}$$

For $\tau\varepsilon\Gamma$ we can consider each $P(\tau\varepsilon C_i|C_1,\cdots,C_{i-1},Q)$ as the successive incremental discriminating power of C_i in Q, $i = 1, 2, \cdots, K$. The discriminating power is dependent on the ordering $C_1, C_2, \cdots, C_K$, and the pairs in $C_1\cap C_2\cap\cdots\cap C_K\cap Q$. For record pairs r_j, $j = 1, 2, \cdots, N$, from Q, index the comparison vectors τ_j^i as follows:

$$\tau_j^i = \begin{cases} = 1 & \text{if field } i \text{ agrees for record pair } r_j \\ = 0 & \text{if field } i \text{ disagrees for record pair } r_j. \end{cases}$$

The elements in $Q = (Q\cap M)\cup(A\cap U)$ are distributed according to a finite mixture with the

unknown parameters $\Phi = (m, u, p)$ where p is the proportion of matched pairs in Q. Let $\mathbf{x}$ be the complete data vector $g = \langle \tau_j, g_j \rangle$ where

$g_j = (1,0)$ if $r_j \in M \cap Q$ and

$g_j = (0,1)$ if $r_j \in U \cap Q$.

Then the complete data log-likelihood (Dempster, Laird, and Rubin 1977, pp. 15-16) is given by

$$\ln f(\mathbf{x} \mid \Phi) = \sum_{j=1}^{N} g_j \cdot \langle \ln P(\tau_j \mid M \cap Q), \ln P(\tau_j \mid U \cap Q) \rangle$$

$$+ \sum_{j=1}^{N} g_j \cdot \langle \ln p, \ln(1-p) \rangle.$$

Fitting using the EM Algorithm will be performed under the following assumption: There exist vector constants $m \equiv (m_1, m_2, \cdots, m_K)$ and $u \equiv (u_1, u_2, \cdots, u_K)$ such that, for all $\tau \in \Gamma$,

$$P(\tau \mid M \cap Q) = \prod_{i=1}^{K} m_i^{\tau^i} (1-m_i)^{(1-\tau^i)}$$

and (2.5)

$$P(\tau \mid U \cap Q) = \prod_{i=1}^{K} u_i^{\tau^i} (1-u_i)^{(1-\tau^i)}.$$

Probabilities m_i and u_i, $i = 1, 2, \cdots, K$, are constant for all representations τ of pairs in Q. The set of probabilities of form (2.5) includes those obtained by reorderings as in (2.4), provided the reordering is fixed for all $\tau \in \Gamma$. To avoid trivialities, we assume that $0 < m_i, u_i < 1$, $i = 1, 2, \cdots, K$.

If the Conditional Independence Assumption holds, then the m_i and u_i, $i = 1, 2, \cdots, K$, are the usual marginal probabilities as given in FS (pp. 1194-1195). They necessarily are independent of any reordering of $B_1, B_2, \cdots, B_K$. When the Conditional Independence Assumption does not hold probability distributions of form (2.5) constitute a more general class than those obtained under the Conditional Independence Assumption.

We begin the EM Algorithm with estimates of the unknown parameter $\langle \hat{m}, \hat{u}, \hat{p} \rangle$. For the E-step under (2.5), replace g_j with

$\langle P(M \cap Q \mid \tau_j), P(U \cap Q \mid \tau_j) \rangle$ where

$$P(M\cap Q|\tau_j) \equiv \frac{\hat{p}\prod_{i=1}^{K}\hat{m}_i^{\tau_j^i}(1-\hat{m}_i)^{(1-\tau_j^i)}}{\hat{p}\prod_{i=1}^{K}\hat{m}_i^{\tau_j^i}(1-\hat{m}_i)^{(1-\tau_j^i)} + (1-\hat{p})\prod_{i=1}^{K}\hat{u}_i^{\tau_j^i}(1-\hat{u}_i)^{(1-\tau_j^i)}}$$

and (2.6)

$$P(U\cap Q|\tau_j) \equiv \frac{(1-\hat{p})\prod_{i=1}^{K}\hat{u}_i^{\tau_j^i}(1-\hat{u}_i)^{(1-\tau_j^i)}}{\hat{p}\prod_{i=1}^{K}\hat{m}_i^{\tau_j^i}(1-\hat{m}_i)^{(1-\tau_j^i)} + (1-\hat{p})\prod_{i=1}^{K}\hat{u}_i^{\tau_j^i}(1-\hat{u}_i)^{(1-\tau_j^i)}}.$$

For the M step, the complete data log-likelihood can be separated into three maximization problems. Setting the partial derivatives equal to zero and solving for $\hat{m}_i$, i = 1, 2, ···, K, yields:

$$\hat{m}_i = \frac{\sum_{j=1}^{N} P(M\cap Q \mid \tau_j)\cdot\tau_j^i}{\sum_{j=1}^{N} P(M\cap Q \mid \tau_j)}. \quad (2.7)$$

Estimates $\hat{u}_i$, i = 1, 2, ···, K, are derived in a similar manner. The matrix of second partial derivatives can be shown to be negative-definite. The estimate of the proportion of matched pairs is given by

$$\hat{p} = \frac{\sum_{j=1}^{N} P(M\cap Q \mid \tau^j)}{N}.$$

2.2.4. Extension to Frequency-Based Weights

This section considers a procedure for extending simple agreement/disagreement weights to weights that account for frequency. We call such a procedure a *dispersion*. When the more stringent assumptions of FS (pp. 1207-1210) are satisfied our dispersion procedure agrees with theirs. If the agreement/disagreement weights found via the EM Algorithm coincide with the agreement/disagreement weights found via the FS procedures, then the frequency-based weights also coincide.

Frequency-based weights are useful because they can account for the fact that a specific surname pair such as (Zabrinksy, Zabrinsky) occurs less often than a surname pair such as (Smith, Smith).

We need some background material before presenting the computational procedures for frequency-based weights.

We observe that if, for some i and k,

$$m_i = P(\tau^k = 1 \mid M \cap Q)$$

and (2.8)

$$u_i = P(\tau^k = 1 \mid U \cap Q),$$

then the kth comparison is independent of the other $K-1$ comparisons. The right hand sides of (2.8) are just the appropriate marginal inclusion probabilities. Note that m_i and u_i, $i = 1, 2, \cdots, K$, of this paper generally differ from the m_1, m_2, m_3, u_1, u_2, and u_3 in FS (pp. 1194-1195, 1207-1210).

We define a random variable $\tau^{\wedge k}$ by

$$\tau^{\wedge k} = \mu_j^k \text{ if the } k\text{th comparison pair takes value } \mu_j^k$$

where μ_j^k, $j = 1, \cdots, L^k$, is an enumeration of the specific values of the kth comparison. We make two assumptions:

A1. Agreement/disagreement in the kth comparison is independent of the other $K-1$ comparisons.

A2. There exists a comparison k' such that the specific realizations of $\tau^{\wedge k}$ are pairwise independent of agreement/disagreement in the k'th comparison.

If we consider one comparison, say of agreement/disagreement in surname, then we can perform EM fitting under a restricted version of (2.5) by specifying that one of the (m_i, u_i) must converge to the marginal probabilities (as in (2.8)) associated with surname. We can, thus, always find a comparison satisfying assumption A1 for the restricted class of distributions.

Assumption A2 is a weaker form of independence assumption than the one considered by FS (p. 1208). It allows dispersion of the agreement/disagreement weight obtained under assumption A1 to frequency-based weights.

In a manner similar to the dispersion of FS (pp. 1207-1210), we define

$$N_k(\mu_i^k) = P(\tau^{\wedge k} = \mu_i^k, \tau^{k'}=1),$$

$$V_k(\mu_i^k) = P(\tau^{\wedge k} = \mu_i^k),$$

c = # pairs in Q, and

N = # pairs in $M \cap Q$.

Then, for $i = 1, 2, \cdots L^k$,

$$c \cdot N_k(\mu_i^k) = N \cdot P(\tau^{\wedge k} = \mu_i^k \mid M \cap Q) \cdot P(\tau^{k'}=1 \mid M \cap Q) + \quad (2.9)$$

$$(c - N) \cdot P(\tau^{\wedge k} = \mu_i^k \mid U \cap Q) \cdot P(\tau^{k'}=1 \mid U \cap Q)$$

and

$$c \cdot V_k(\mu_i^k) = N \cdot P(\tau^{\wedge k} = \mu_i^k \mid M \cap Q) + \quad (2.10)$$

$$(c - N) \cdot P(\tau^{\wedge k} = \mu_i^k \mid U \cap Q).$$

In (2.9) and (2.10) c, $N_k(\mu_i^k)$, $V_k(\mu_i^k)$, $i = 1, 2, \cdots L^k$, can be computed directly because they are based on observed file characteristics. The marginal probabilities $P(\tau^k=1 \mid M \cap Q)$ and $P(\tau^k=1 \mid U \cap Q)$ and the number of matches N in $M \cap Q$ can be computed using the estimated parameters of (2.4) that are obtained by the EM Algorithm. Equations (2.9) and (2.10), thus, consist of two equations to be solved for the two unknowns $P(\tau^{\wedge k} = \mu_i^k \mid M \cap Q)$ and $P(\tau^{\wedge k} = \mu_i^k \mid U \cap Q)$, $i = 1, 2, \cdots L^k$.

3. **DISCUSSION**

This section is divided into four parts. The first discusses the convergence properties of the EM Algorithm. The second describes how the EM Algorithm automatically incorporates a Bayesian adjustment of weights for file characteristics. The third considers the extension to frequency-based weights. The fourth deals with the extension of the EM Algorithm to parameter estimation for a reasonably general class of distributions.

3.1. Convergence Properties of EM Algorithm

This paper's application of the EM Algorithm most closely resembles the more general approach of Hasselblad (1969). Hasselblad noted the increase in likelihood on successive steps but was unable to prove convergence. Haberman (1977) proved convergence in a substantially more general setting. He observed that the limiting value was dependent of the initial values of the parameters and, thus, not necessarily unique.

Wu (1983) noted that limiting values of the EM Algorithm are stationary points that can either be saddle points or local maxima. He made the conjecture that there is unlikely to be any general condition that assures convergence to a unique maximum. Wu did observe, however, that if the likelihood is unimodal, if the estimated parameters have at most one stationary limiting point, and if a technical condition holds (which it does for the distributions of this paper), then the estimated parameters converge to the unique maximizer of the likelihood. The implication is that, while the EM algorithm of this paper is of value in accounting for failures of the Conditional Independence Assumption, several starting points for the EM Algorithm should be used. The estimated parameters associated with the largest local maximum are the ones that are used. If we can show that there is at most one stationary limiting point, then the parameter estimates will necessarily converge to it.

3.2. Bayesian Adjustment of Weights

If weights are computed via Method II of FS (p. 1194), then they correspond to observed characteristics of the files. Similarly, weights computed using the application of the EM Algorithm of this paper will also correspond to file characteristics.

If, during a file updating project, we use as initial weights those weights obtained from an earlier project using different files, then the EM Algorithm will automatically adjust the weights to the characteristics of the new files. 3.3. Extension to Frequency-Based Weights

Under the more stringent assumptions of FS (pp. 1207-1210) frequency-based weights computed using the techniques of this paper agree with those computed using Method II of FS. This follows because the dispersion method of this paper is identical to the dispersion method of FS and the there can exist at most one local maximum of the likelihood.

The chief value of assumptions like Assumptions A1 and A2 of this paper is that they allow dispersal of agreement/disagreement weights with little increase in computation. Although the EM Algorithm might be extended to allow direct computation of frequency-based weights, such an extension will generally require enormous increases in computation.

3.4. Extension of Computational Procedures

Distributions of form (2.5) are not sufficiently general to deal with the joint relationship between A_1 and A_2. Form (2.5) yields distributions having essentially independent effects. Generally, use of such distributions does not allow effective modeling of events of the form

$$P(\tau\varepsilon A_1^x \cap A_2^x \mid A_3 \cap M \cap Q) \text{ and } P(\tau\varepsilon A_1^x \cap A_2^x \mid A_3 \cap U \cap Q)$$

where A_i^x represents the comparison event or its complement. A_3 represents any other comparison or set of comparisons.

Rather than give a fully rigorous development, we will consider a relatively simple case. There are two comparisons that are independent of other comparisons but may not be independent of each other. Due to the form of the independence, we can suppress any parameters associated with comparisons other than the first two.

There exist vector constants $m \equiv (m_1, m_2, m_{12})$ and $u \equiv (u_1, u_2, u_{12})$ such that, for all $\tau\varepsilon\Gamma$

$$P(\tau \mid M \cap Q) = m_1^{\tau^1\cdot\tau^2} \cdot (1-m_1)^{(1-\tau^1)\cdot\tau^2} \cdot m_2^{\tau^2} \cdot (1-m_2)^{(1-\tau^2)} \cdot m_{12}^{\tau^1\cdot(1-\tau^2)} \cdot (1-m_{12})^{(1-\tau^1)\cdot(1-\tau^2)}$$

and (3.1)

$$P(\tau \mid U \cap Q) = u_1^{\tau^1\cdot\tau^2} \cdot (1-u_1)^{(1-\tau^1)\cdot\tau^2} \cdot u_2^{\tau^2} \cdot (1-u_2)^{(1-\tau^2)} \cdot u_{12}^{\tau^1\cdot(1-\tau^2)} \cdot (1-u_{12})^{(1-\tau^1)\cdot(1-\tau^2)}$$

Representations of form (3.1) with additional parameters m_{12} and u_{12} can account for the interaction of events A_1 and A_2.

In special cases, form (2.5) might allow modelling such pairwise interactions. If, in (3.1), m_1 =

m_{12} and $u_1 = u_{12}$, we have independence and form (3.1) agrees with form (2.5). Formal extension to general cases is straightforward.

Use of the EM Algorithm with distributions of form (3.1) is straightforward. Probability distributions of form (2.5) in the E step (2.6) are replaced with distributions of form (3.1). The M-step estimates still take form (2.7). We note that two additional parameters m_{12} and u_{12} must be estimated. The reason the extension is straightforward is that the complete data log-likelihood of the M step takes essentially the same form it did for the simpler class of distributions (2.5).

To create classes of distributions that more effectively model multiple interactions of events necessitates straightforward extension of the probability distributions of form (3.1). In practice, we would likely select just a few interactions to model in order to minimize the increase in computation involved with the EM Algorithm.

4. SUMMARY

This paper provides a method of estimating matching weights in the Fellegi-Sunter model of record linkage. Under an assumption weaker than the usual Conditional Independence Assumption, the estimates are obtained via the EM Algorithm. The procedure automatically incorporates a Bayesian adjustment for file characteristics.

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion The shorter version of this paper was presented at the American Statistical Association Annual Meeting and appeared in the 1988 Proceedings of the Section on Survey Research Methods (pp. 667-671).

REFERENCES

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Society*, B, **39**, 1-38.

Haberman, S. J. (1977), "Product Models for Frequency Tables Involving Indirect Observation: Maximum Likelihood Equations," *Annals of Statistics*, **5**, 1124-1147.

Haselblad, V. (1969), "Estimation of Finite Mixtures of Distributions from Exponential Family," *Journal of the American Statistical Association*, **64**,1459-1471.

Kelley, R. P. (1986), "Robustness of the Census Bureau's Record Linkage System, American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 620-624.

Newcombe, H.B., Kennedy, J.M., Axford, S.J., and James, A. P. (1959), "Automatic Linkage of Vital Records," *Science*, **130**, 954-959.

Winkler, W. E. (1987a), "An Application of the Fellegi-Sunter Model of Record Linkage to Lists of Businesses," Energy Information Administration Technical Report.

Winkler, W. E. (1987b), "Computational Aspects of Applying the Fellegi-Sunter Model of Record Linkage to Lists of Businesses," paper presented at the Symposium on Statistical Uses of Administrative Data.

Wu, C. F. J. (1983) "On the Convergence Properties of the EM Algorithm," Annals of Statistics, **11**, 95-103.

EXHIBIT 8

AN APPLICATION OF THE FELLEGI-SUNTER MODEL OF RECORD LINKAGE TO THE 1990 U.S. DECENNIAL CENSUS

William E. Winkler and Yves Thibaudeau
U.S. Bureau of the Census

ABSTRACT

This paper describes a methodology for computer matching the Post Enumeration Survey with the Census. Computer matching is the first stage of a process for producing adjusted Census counts. All crucial matching parameters are computed solely using characteristics of the files being matched. No a priori knowledge of truth of matches is assumed. No previously created lookup tables are needed. The methods are illustrated with numerical results using files from the 1988 Dress Rehearsal Census for which the truth of matches is known.

Key words and phrases. EM Algorithm; String Comparator Metric; LP Algorithm; Decision Rule; Error Rate.

1. INTRODUCTION

This paper describes a particular application of the Fellegi-Sunter (1969) model of record linkage. New computational methods are used for computer matching the Post Enumeration Survey (PES) with the Census. The PES is used to produce adjusted Census counts. Computer matching is the first stage of PES processing.

All crucial matching parameters associated with comparisons of individual fields are computed automatically. The parameters are generally based on characteristics of the files being matched. No a priori knowledge of truth of matches is assumed. Lookup tables that account for the relative frequency of occurrence of different strings are computed using the files being matched. The paper is divided into a number of sections.

The second section consists of five parts. The first part gives background on the Fellegi-Sunter model. The second part describes PES and Census files from the 1988 Dress Rehearsal Census and overall matching procedures. Truth and falsehood of matches is known for the Dress Rehearsal files.

The third part provides details of a modified Expectation-Maximization (EM) Algorithm for estimating probability distributions used in a crucial likelihood ratio (see e.g., Winkler 1988, 1989a; Thibaudeau 1989). In the fourth part, new computational methods for automatically creating frequency tables accounting for the relative distinguishing power of strings such as 'Smith' and 'Zabrinsky' are given. The methods are a special case of Winkler (1989b).

The fifth part describes new string comparator metrics that allow comparison of strings that do not agree on a character-by-character basis. The metrics generalize Damerau-Levenstein and Jaro metrics (see e.g., Winkler 1985, 1989c, 1990b). Methods for modeling how the metric adjusts matching weights between pure agreement and pure disagreement are covered.

A new linear sum assignment algorithm that forces one-to-one assignments is described in the

sixth part. It is substantially faster than the algorithm of Burkard and Derigs used by Jaro (1989).

Results are presented in the third section. Two commonly used matching methods are compared with the method used in the 1988 computer matching system and the enhanced methods adopted for 1990.

The fourth section consists of a six part discussion. The first describes how matching accuracy may fall when files are of relatively lower quality than those available in the 1990 Census and PES. In the second part, we cover how the string comparator metrics and weight adjustment methods fit in with the general Fellegi-Sunter model. The third part presents general limitations of the parameters produced via the EM Algorithm.

In the fourth part of the fourth section, we explain why the specific adjustments of the frequency-table building procedures are suitable in the applications. The fifth part covers how the linear sum assignment procedure improves matching efficacy and why decision rules using it are still optimal in the sense given by Fellegi and Sunter (1969, Theorem). In the sixth part, we discuss different ways of determining cutoff weights.

The fifth section consists of a summary, conclusions, and future work.

2. BACKGROUND AND DESCRIPTION OF METHODS

2.1. Fellegi-Sunter Model of Record Linkage

The Fellegi-Sunter Model uses a decision-theoretic approach establishing the validity of principles first used in practice by Newcombe (Newcombe et al. 1959). To give an overview, we describe the model in terms of ordered pairs in a product space. The description closely follows Fellegi and Sunter (1969, pp. 1184-1187).

There are two populations **A** and **B** whose elements will be denoted by a and b. We assume that some elements are common to **A** and **B**.
Consequently the set of ordered pairs

$$\mathbf{A}\mathrm{X}\mathbf{B} = \{(a,b): a \in \mathbf{A},\ b \in \mathbf{B}\}$$

is the union of two disjoint sets of matches

$$M = \{(a,b): a = b,\ a \in \mathbf{A},\ b \in \mathbf{B}\}$$

and nonmatches

$$U = \{(a,b): a \neq b,\ a \in \mathbf{A},\ b \in \mathbf{B}\}.$$

The records corresponding to members of **A** and **B** are denoted by $\alpha(a)$ and $ß(b)$, respectively. The comparison vector γ associated with the records is defined by:

$$\gamma[\alpha(a),ß(b)] \equiv \{\gamma^{1}[\alpha(a),ß(b)],\gamma^{2}[\alpha(a),ß(b)],\cdots,\gamma^{K}[\alpha(a),ß(b)]\}.$$

Each of the γ^{i}, i = 1, ···, K, represents a specific comparison. For instance, γ^{1} could represent agreement/disagreement on sex. γ^{2} could represent the comparison that two surnames agree

and take a specific value or that they disagree.

Where confusion does not arise, the function γ on $\mathbf{A}\mathbf{X}\mathbf{B}$ will be denoted by $\gamma(\alpha,ß)$, $\gamma(a,b)$, or γ. The set of all possible realizations of γ is denoted by Γ.

The conditional probability of $\gamma(a,b)$ if $(a,b)\epsilon M$ is given by

$$m(\gamma) \equiv P\{\gamma[\alpha(a),ß(b)]|(a,b)\epsilon M\}$$

$$= \sum_{(a,b)\epsilon M} P\{\gamma[\alpha(a),ß(b)]\}\cdot P[(a,b)|M].$$

Similarly we denote the conditional probability of γ if $(a,b)\epsilon U$ by $u(\gamma)$.

We observe a vector of information $\gamma(a,b)$ associated with pair (a,b) and wish to designate a pair as a link (denote the decision by A_1), a possible link (decision A_2), or a nonlink (decision A_3). A <u>linkage rule</u> L is defined a mapping from Γ, the comparison space, onto a set of random decision functions $D = \{d(\gamma)\}$ where

$$d(\gamma) = \{P(A_1|\gamma),P(A_2|\gamma),P(A_3|\gamma)\};\ \gamma\epsilon\Gamma$$

and

$$\sum_{i=1}^{3} P(A_i|\gamma) = 1.$$

There are two types of error associated with a linkage rule. A <u>Type I error</u> occurs if an unmatched comparison is erroneously linked. It has probability

$$P(A_1|U) = \sum_{\gamma\epsilon\Gamma} u(\gamma)\cdot P(A_1|\gamma)$$

A <u>Type II error</u> occurs if a matched comparison is erroneously not linked. It has probability

$$P(A_3|U) = \sum_{\gamma\epsilon\Gamma} m(\gamma)\cdot P(A_3|\gamma)$$

Fellegi and Sunter (1969) define a linkage rule L_0, with associated decisions A_1, A_2, and A_3, that is optimal in the following sense:

Theorem (Fellegi-Sunter 1969). Let L' be a linkage rule with associated decisions A_1', A_2', and A_3' such that it has the same error probabilities $P(A_3'|M) = P(A_3|M)$ and $P(A_1'|U) = P(A_1|U)$ as L_0. Then L_0 is optimal in that $P(A_2|U) \le P(A_2'|U)$ and $P(A_2|M) \le P(A_2'|M)$.

In other words, if L' is any competitor of L_0 having the same Type I and Type II error rates

(which are both conditional probabilities), then the conditional probabilities (either on set U or M) of not making a decision under rule L' are always greater than under L_0.

To describe rule L_0, we need the following likelihood ratio

$$R \equiv R[\gamma(a,b)] = m(\gamma)/u(\gamma). \qquad (2.1)$$

We observe that, if γ represents a comparison of K fields, then there are at least 2^K probabilities of form $m(\gamma)$. If γ represents agreements of K fields, we would expect this to occur more often for matches M than for nonmatches U. The ratio R would then be large. Alternatively, if γ consists of disagreements, the ratio R would be small.

If the numerator is positive and the denominator is zero in (2.1), we assign an arbitrary very large number to the ratio. The Fellegi-Sunter linkage rule L_0 takes the form:

If $R > UPPER$, then denote (a,b) as a link.

If $LOWER \leq R \leq UPPER$, then denote (a,b) as a possible link.

If $R < LOWER$, then denote (a,b) as a nonlink.

The cutoffs LOWER and UPPER are determined by the desired error rate bounds.

The Fellegi-Sunter linkage rule is actually optimal with respect to any set Q of ordered pairs in **A**X**B** if we define error probabilities P_Q and a linkage rule L_Q conditional on Q. Thus, it may be possible to define subsets of **A**X**B** on which we make use of differing amounts and types of available information.

For instance, if we have a set of pairs in which telephone number is present, we might use telephone number and a few characters from the name to designate links. With other pairs, we may additionally have to utilize information from the street address and the city name.

Sets of ordered pairs Q on which the Fellegi-Sunter linkage rule is applied are often obtained by blocking criteria. Blocking criteria are sort keys that are used to reduce the number of pairs that are considered. Rather than consider all pairs in **A**X**B**, we might only consider pairs that agree on the first three digits of the ZIP code or on a suitable abbreviation of surname.

2.2. Empirical Data Bases and Description of Matching Procedures

Files for three geographic regions, (1) St Louis, MO, (2) Columbia, MO and vicinity, and (3) rural WA, were computer matched during the 1988 Dress Rehearsal Census and Post Enumeration Survey. The PES is a reenumeration of all individuals in all households in a sample of Census subregions. Each subregion, referred to as a block cluster, may be a Census block, a subsample of a Census block, or a group of Census blocks. The Census file consists of the individuals in the Census for the same sampled block clusters and for individuals in blocks or geographic subregions surrounding the PES blocks.

On the first computer matching pass, individuals in the PES are matched only with those individuals in the Census that are in the same block cluster. As almost all true matches that are ultimately found are identified during the first computer matching pass and associated clerical operations, the results of this paper only consider the first pass. Subsequent matching passes

against the entire Census file (which is typically 25 to 30 times the size of the PES) are primarily intended to identify individuals who have been counted in the wrong location (i.e., near-by blocks). File sizes for within-block matching are given in

Table 1. File Sizes from the 1988 Dress Rehearsal Census Nonmovers for Within Block Cluster Matching

City	Census	PES
St Louis	15048	12072
Columbia	9794	7649
Washington	3030	2214

.

Matching within block clusters is mandated by the fact that PES-estimation procedures produce adjusted counts at the block-cluster level and aggregate. For matching within block cluster, we additionally use first character of the surname as a logical blocking criteria. Use of first character of surname reduces computation by a factor of 150 but generally causes 3-5 percent of true matches to be missed. Approximately 2 percent of the missed true matches (about half of the 3-5 percent) are in households for which at least one person is correctly matched by computer procedures and, thus, quickly located during clerical review.

The printouts used for the initial clerical review bring together all individuals in a Census household with all individual in a PES household if at least one individual is designated as a computer match or possible match. This, for instance, allows matching of children that are listed under different last names if at least one of the parents is matched.

The remaining 1-2 percent that are not quickly identified as a result of the computer matching operations are located during additional clerical review procedures. This involves manual comparison of PES and Census lists that have been alphabetically sorted by last name and also by address.

Fields available for matching are first name, middle initial, last name, house number, street name, rural route number, postal box number, conglomerated address, unit, telephone number, age, sex, marital status, relationship to head of household, and race. The last name field receives minor processing to separate out words such as Junior, Senior, and III. The first name field receives processing to change nicknames to standard roots of their given first names. A modified version of the Census Geography Division address standardizer is used for delineating address components.

Generally, if house number and street name are not missing, then rural route and box number are missing and vice-versa. The conglomerated address is used as a default identifier for those addresses failing standardization. As the PES reenumerates hard-to-count regions, many addresses such as those associated with trailer courts, migrant worker camps, and various institutional group quarters such as retirement homes and university dormitories are difficult to standardize.

For each PES individual, their true match status and, where appropriate, to whom they were matched is known. The quality of the matching status is very high. The quality is covered in the discussion.

Telephone number, although generally a very good identifier in many applications, is only a somewhat good identifier for the applications in this paper. For instance, in St Louis, only 66 percent of those individuals that were eligible for computer matching had a nonblank telephone number agreeing with the telephone number of the individual to whom matched. For some individuals who were ultimately correctly matched, say those living in trailer courts, telephone numbers sometimes disagree.

2.3. Expectation-Maximization Algorithm

Applying the Fellegi-Sunter model involves determining estimates of the conditional probabilities $m(\gamma)$ and $u(\gamma)$. To obtain maximum likelihood estimates we use the EM Algorithm.

For record pairs r_j, $j = 1, 2, \cdots, N$, from Q, index the comparison vectors γ_j^i as follows:

$\gamma_j^i = 1$ if field i agrees for record pair r_j

$= 0$ if field i disagrees for record pair r_j.

The elements in $Q = (Q \cap M) \cup (A \cap U)$ are distributed according to a finite mixture with the unknown parameters $\Phi = (m, u, p)$ where p is the proportion of matched pairs in Q. Let $\mathbf{x}$ be the complete data vector
$g = \langle \gamma_j, g_j \rangle$ where

$g_j = (1,0)$ if $r_j \in M \cap Q$ and

$g_j = (0,1)$ if $r_j \in U \cap Q$.

Then the complete data log-likelihood (Dempster, Laird, and Rubin 1977, pp. 15-16) is given by

$$\ln f(\mathbf{x} \mid \Phi) = \sum_{j=1}^{N} g_j \cdot \langle \ln P(\gamma_j \mid M \cap Q), \ln P(\gamma_j \mid U \cap Q) \rangle$$

$$+ \sum_{j=1}^{N} g_j \cdot \langle \ln p, \ln(1-p) \rangle.$$

Fitting using the EM Algorithm will be performed under the following conditional independence assumption: There exist vector constants $m \equiv (m_1, m_2, \cdots, m_K)$ and $u \equiv (u_1, u_2, \cdots, u_K)$ such that, for all $\gamma \in \Gamma$,

$$P(\gamma \mid M \cap Q) = \prod_{i=1}^{K} m_i^{\gamma^i} (1-m_i)^{(1-\gamma^i)}$$

and (2.2)

$$P(\gamma \mid U \cap Q) = \prod_{i=1}^{K} u_i^{\gamma^i} (1-u_i)^{(1-\gamma^i)}.$$

Probabilities m_i and u_i, $i = 1, 2, \cdots, K$, are constant for all representations γ of pairs in Q. To avoid trivialities, we assume that $0 < m_i, u_i < 1$, $i = 1, 2, \cdots, K$.

We begin the EM Algorithm with estimates of the unknown parameter

$<\hat{m}, \hat{u}, \hat{p}>$. For the E-step under (2.2), replace g_j with $<P(M \cap Q|\gamma_j), P(U \cap Q|\gamma_j)>$ where

$$P(M \cap Q|\gamma_j) \equiv \frac{\hat{p} \prod_{i=1}^{K} \hat{m}_i^{\gamma_j^i} (1-\hat{m}_i)^{(1-\gamma_j^i)}}{\hat{p} \prod_{i=1}^{K} \hat{m}_i^{\gamma_j^i} (1-\hat{m}_i)^{(1-\gamma_j^i)} + (1-\hat{p}) \prod_{i=1}^{K} \hat{u}_i^{\gamma_j^i} (1-\hat{u}_i)^{(1-\gamma_j^i)}}$$

and

$$P(U \cap Q|\gamma_j) \equiv \frac{(1-\hat{p}) \prod_{i=1}^{K} \hat{u}_i^{\gamma_j^i} (1-\hat{u}_i)^{(1-\gamma_j^i)}}{\hat{p} \prod_{i=1}^{K} \hat{m}_i^{\gamma_j^i} (1-\hat{m}_i)^{(1-\gamma_j^i)} + (1-\hat{p}) \prod_{i=1}^{K} \hat{u}_i^{\gamma_j^i} (1-\hat{u}_i)^{(1-\gamma_j^i)}}.$$

For the M step, the complete data log-likelihood can be separated into three maximization problems. Setting the partial derivatives equal to zero and solving for $\hat{m}_i$,

$i = 1, 2, \cdots, K$, yields:

$$\hat{m}_i = \frac{\sum_{j=1}^{N} P(M \cap Q \mid \gamma_j) \cdot \gamma_j{}^i}{\sum_{j=1}^{N} P(M \cap Q \mid \gamma_j)}.$$

Estimates $\hat{u}_i$, i = 1, 2, ⋯, K, are derived in a similar manner. The matrix of second partial derivatives can be shown to be negative-definite. The estimate of the proportion of matched pairs is given by

$$\hat{p} = \frac{\sum_{j=1}^{N} P(M \cap Q \mid \gamma^j)}{N}.$$

The m- and u-probabilities obtained from the EM Algorithm do not generally work well in the decision rule of Fellegi-Sunter because the fitting procedures tend to force m-probabilities away from 1 and u-probabilities away from 0.

To get parameters that yield better decision rules (i.e., both lower rates of false matches and smaller regions of clerical pairs), we use the following procedure:

1. Compute probabilities of random agreement for those pairs agreeing on geocode and on a characteristic such as first name. If the random agreement probability is less than the EM-derived u-probability, substitute. If the resultant u-probability is less than 0.005, replace it with 0.005.
2. Replace the m-probabilities for last and first names with 0.9997 and 0.9999, respectively.

The resultant m- and u-probabilities are still obtained automatically. The effect of 2. is negative disagreement weights of greater absolute value. That the substitution 0.9999 for the m-probability of first name improves distinguishing power is intuitive. As family members, in particular, agree on last name, street name, house number, telephone number, and some demographic characteristics, the more negative disagreement weight induced by the substitution helps delineate different individuals in the same family.

2.4. Frequency-Based Weighting

In this section, we also closely follow the terminology of Fellegi and Sunter (1969, section 3.3.1). Let the true frequencies of occurrence of a specified string in files **A** and **B**, respectively, be

$$f_1, f_2, \cdots, f_m ; \sum_{j=1}^{m} f_j = N_A$$

and

$$g_1, g_2, \cdots, g_m ; \sum_{j=1}^{m} g_j = N_B.$$

Let the corresponding true frequencies in $\mathbf{A} \cap \mathbf{B}$ be

$$h_1, h_2, \cdots, h_m ; \sum_{j=1}^{m} h_j = N_{AB}.$$

We note that $h_j \leq \min(f_j, g_j)$, $j = 1, 2, \cdots, m$. For the empirical examples of section 3.2, for $j = 1, 2, \cdots, m$, we will generally use

$h_j = \min(f_j, g_j)$ if $f_j > 1$ or $g_j > 1$

$h_j = 2/3$ otherwise.

The latter part of the definition implicitly means that, if we observe only one pair agreeing on a specific string, the pair has 2/3 chance of being a match and 1/3 chance of being a nonmatch.

The following additional notation is needed

e_A or e_B the respective probabilities of a name being misreported in **A** or **B** (assumed independent of a particular name);

e_{A0} or e_{B0} the respective probabilities of a name not being reported in **A** or **B** (name independent);

e_T the probability that a name is differently (but correctly) reported in the two files.

Then we have the following representations:

P(string agrees & jth string| M)

$$= h_j (1-e_A)(1-e_B)(1-e_T)(1-e_{A0})(1-e_{B0})/N_{AB}$$

$\approx h_j\,(1-e_A-e_B-e_T-e_{A0}-e_{B0})/N_{AB}$;

P(string disagrees | M)

$= [1-(1-e_A)(1-e_B)(1-e_T)](1-e_{A0})(1-e_{B0}) \approx e_A + e_B + e_T$;

P(string missing on either file | M)

$= 1 - (1-e_{A0})(1-e_{B0}) \approx e_{A0} + e_{B0}$;

P(string agrees & jth string| U)

$= (f_j \cdot g_j - h_j)(1-e_A)(1-e_B)(1-e_T)(1-e_{A0})(1-e_{B0})/(N_A \cdot N_B - N_{AB})$

$\approx (f_j \cdot g_j - h_j)(1-e_A-e_B-e_T-e_{A0}-e_{B0})/(N_A \cdot N_B - N_{AB})$;

P(string disagrees | U)

$= [1-(1-e_A)(1-e_B)(1-e_T) \sum_{J=1}^{m} (f_j \cdot g_j - h_j)/(N_A \cdot N_B - N_{AB})](1-e_{A0})(1-e_{B0})$

$\approx [1-(1-e_A-e_B-e_T) \sum_{J=1}^{m} (f_j \cdot g_j - h_j)/(N_A \cdot N_B - N_{AB})](1-e_{A0}-e_{B0})$; and

P(string missing on either file | U)

$= 1 - (1-e_{A0})(1-e_{B0}) \approx e_{A0} + e_{B0}$.

We define the weight for agreement on the jth specific string, j = 1, 2, ⋯, m, by

$wgt(j) = h_j \cdot (N_A \cdot N_B - N_{AB}) / ((f_j \cdot g_j - h_j) \cdot N_{AB})$,

if either $f_j>1$ or $g_j>1$ and by

$wgt(j) = 2 \cdot (N_A \cdot N_B - N_{AB})/N_{AB}$

if $f_j=1$ and $g_j=1$. The weight represents the probability of agreement over the entire product space.

We observe that e_{A0} and e_{B0} can be estimated directly using file characteristics. To estimate e_T, e_A and e_B, we need to know $\mathbf{A} \cap \mathbf{B}$. They cannot be estimated directly. In practice, guesses based on past experience are often used.

Use of the EM-Algorithm, within the limitations of the previous section, allows direct estimation of P(string disagrees|M) and, thus, approximate estimation of the sum $e_A+e_B+e_T$.

To assure that frequency weights do not overwhelm simple agree/disagree weights given by the EM procedure, we use for the agreement weights

$c_1 \cdot \ln(c_2 \cdot P(\text{jth string agrees} \mid M) / P(\text{jth string agrees} \mid U))$.

For last name c_1 and c_2 are 0.03125 and 0.5, respectively. For first name c_1 and c_2 are 0.0625 and 0.7, respectively.

General methods of directly computing c_1 and c_2 are given in Winkler (1989b) but do not work well for all types of files. For the narrow classes of files we consider in the applications of this paper, we can choose c_1 and c_2 a priori based on modeling files with similar characteristics.

2.5. String Comparator Metrics

When human beings review a pair, they often do not make simple yes/no decisions related to the weight of agreement associated with a fixed string. Typically, they will assign various degrees of maybe to a pair of strings exhibiting typographical variation (Table2).

Table 2. Hypothetical Human Decisions Based on Typographical Errors
Full Agreement = 10, Full Disagreement = -10

Pair	Jonathon Johathon	Bellieu Baliew	Ovid Ouid	Wilansk Wolansky	Doret Doris
Weight	10	4	1	-3	-10

Jaro (see e.g., Winkler 1985, 1989c, 1990b) introduced a string comparator measure that gives values of partial agreement between two strings. The string comparator accounts for length of strings and partially accounts for the types of errors typically made in alphanumeric strings by human beings. It is used in adjusting exact agreement weights when two strings do not agree on a character-by-character basis.

Specifically, if $c > 0$, the Jaro string comparator is

$$\Phi = W_1 \cdot c/d + W_2 \cdot c/r + W_t \cdot (c-\tau)/c,$$

where

W_1 = weight associated with characters in the first of two files,
W_2 = weight associated with characters in the second of two files,
W_t = weight associated with transpositions,
d = length of string in first file,
r = length of string in second file,
τ = number of transpositions of characters, and
c = number of characters in common in pair of strings.

If $c = 0$, then $\Phi = 0$.

Two characters are considered in common only if they are no further apart than (m/2 - 1) where m = max(d,r). Characters in common from two strings are assigned; remaining characters unassigned. Each string has the same number of assigned characters.

The number of transpositions is computed as follows: The first assigned character on one string is compared to the first assigned character on the other string. If the characters are not the same, half of a transposition has occurred. Then the second assigned character on one string is compared to the second assigned character on the other string, etc. The number of mismatched characters is divided by two to yield the number of transpositions.

If two strings agree on a character-by-character basis, then the Jaro string comparator Φ is set to $W_1+W_2+W_t$, which is the maximum value that Φ can assume. The minimum value that the Φ can assume is 0, which occurs when the two strings have no characters in common (subject to the above definition of common).

For present matching applications, W_1, W_2, and W_t are arbitrarily set to 1/3. The new string comparator metric basically modifies the basic string comparator according to whether the first few characters in the strings being compared agree. Specifically, for i = 1, 2, 3, 4,

$$\Phi_n = \Phi + i \cdot 0.1 \cdot (1 - \Phi) \quad \text{if the first } i \text{ characters agree.}$$

If w_a and w_d are the estimated agreement and disagreement weights for a specific field, respectively, then the Jaro adjusted matching weight w_{am} used in the total weight calculation is given by

$$w_{am} = \begin{cases} w_a & \text{if } \Phi = 1, \text{ and} \\ \max\{w_a-(w_a-w_d)\cdot(1-\Phi)\cdot(9/2), w_d\} & \text{if } 0 \le \Phi < 1. \end{cases}$$

The constant 9/2 controls how quickly decreases in partial agreement values force the adjusted weight to the disagreement weight.

Instead of assuming that the same adjustment procedure works for different fields such as first name, last name, and house number, procedures for modeling the weight adjustment as a piecewise linear function were developed. The procedures necessitate having representative sets of pairs for which the truth of matches is known. The new adjusted weights w_{na} take the form

$$w_{na} = \begin{cases} w_a & \text{if } \Phi_n \ge b_1 \\ \max\{w_a-(w_a-w_d)\cdot(1-\Phi_n)\cdot(a_1), w_d\} & \text{if } b_2 \le \Phi_n < b_1. \\ \max\{w_a-(w_a-w_d)\cdot(1-\Phi_n)\cdot(a_2), w_d\} & \text{if } \Phi_n < b_2. \end{cases}$$

The constants a_1, a_2, b_1, and b_2 depend on the specific type of string (such as first name) to which the weight adjustment is applied. Generally, $a_1 < a_2$. The specific constants used are given in Table 3.

Table 3. Constants used in piecewise linear weight adjustments

Field	a_1	a_2	b_1	b_2
first	1.5	3.0	.92	.75
last	3.0	4.5	.96	.88
house #	4.5	7.5	.98	.83

Table 4 provides examples of string comparator values for pairs of last names and for pairs of first names. The abroms-abrams example with string comparator value .9333 in contrast to the lampley-campley with value .9048 shows that the string comparator gives a higher value to the

Table 4. Examples of String Comparator Values for Various Pairs

shackleford	shackelford	.9848
cunningham	cunnigham	.9833
campell	campbell	.9792
nichleson	nichulson	.9630
massey	massie	.9444
abroms	abrams	.9333
galloway	calloway	.9167
lampley	campley	.9048
dixon	dickson	.8533
frederick	fredrick	.9815
michele	michelle	.9792
jesse	jessie	.9722
marhta	martha	.9667
jonathon	jonathan	.9583
julies	juluis	.9333
jeraldine	geraldine	.9246
yvette	yevett	.9111
tanya	tonya	.8933
dwayne	duane	.8578

pair that differs by a single character further from the first position. The martha-marhta example with value .9667 in contrast to the jonathon-jonathan example with value .9583 shows that transposition of two characters causes less of a downweighting than differing by one

2.6. Linear Sum Assignment Algorithm

A linear sum assignment procedure similar to the one introduced by Jaro (1989) is used to force one-to-one assignments. The actual algorithm, due to Errol Rowe (1987), requires less storage than the one of Burkard and Derig (see e.g., Jaro 1989, page 418). Basically, the Rowe algorithm works with rectangular arrays while the Burkard-Derig algorithm only works with square arrays. The algorithm also makes use of some of the characteristics of the arrays of weights. For the arrays of weights occurring when the set of PES records remaining from within-block matching are matched against Census files covering extended areas, the new algorithm is more than 100 times as fast as the previously used algorithm.

3. RESULTS

The section presents a comparison of matching results. The basic strategies are the 1988 strategy used during the Dress Rehearsal Census and the 1990 computer matching strategy. The differences and similarities between the two strategies are presented in Table 5.

Table 5. Comparison of 1988 and 1990 Matching Strategies

	1988	1990
logical blocking	cluster+ soundex surname	cluster+ first char surname
pairs for parameter estimation	cluster+ soundex surname	cluster+ soundex surname
multi string comparison adjust	no	yes
frequency-based	no	yes

. This section also presents comparisons with several other strategies which will be described.

A comparison of matching results is given in Tables 6, 7, and 8 for St Louis, Columbia, and Washington, respectively. To understand the tables, we need describe the types of matching procedures. The simplest procedure, crude, merely uses an ad hoc guess for matching parameters and does not use string comparators. The ad hoc guess consists of assigning matching weights of ±2 to agreement/disagreement on fields such as first name, house number, and age and assigning matching weights of ±1 to less important fields such as marital status and sex.

The next, param, does not use string comparators but does estimate the probabilities $m(\gamma)$ and

u(γ). Such probabilities are often estimated through an iterative procedure that involves manual review of matching results and successive reuse of the reestimated parameters. The third type, param2, uses the same probabilities as param and the basic string comparators.

The fourth type, em, uses an EM-Algorithm for estimating matching parameters (see e.g., Winkler 1988, Thibaudeau 1990) and uses the basic Jaro string comparator. The fourth type is the 1988 matching strategy. The fifth type, em2, uses the EM-derived weights and the new string comparator and new weight adjustments. The final type, freq, replaces simple agree/disagree weights for first name and last name with frequency-based weights (see e.g., Winkler 1989) and also makes adjustments for joint dependencies of agreements on first name, sex, and age. The final type is the 1990 matching strategy.

In each table, the number of matches is determined by a false match rate of 0.002. The crude and param types are allowed to rise slightly above the 0.002 level because they generally have higher error levels.

Table 6. True Matches in Computer Categories
Various Procedures, St Louis
10291 True Matches, 12072 Records
Pairs Agreeing on Cluster and First Character Surname 1/

	--computer designation-- match		clerical	
truth->	match	non-match	match	non-match
crude	310/	1	9344/	794
param	7899/	16	1863/	198
param2	9276/	23	545/	191
em (1988)	9587/	23	271/	192
em2	9639/	24	215/	189
freq (1990)	9801/	24	52/	94

1/ Approximately 400 true matches disagree on first character of surname and are not eligible for inclusion in the table. Of the 400, approximately 200 reside in households where some is correctly matched.

By examining the tables we observe that a dramatic improvement in matches can occur when string comparators are first used (from param to param2). The basic reason is that disagreements (on a character-by-character basis) are replaced by partial agreements. The improvements due to the new string comparators and weighting adjustments (from em to em2) are quite minor.

The main matching results are given in Tables 6, 7, and 8 for St Louis, Columbia, and Washington files, respectively. For St Louis files, with the 1990 strategy, 9801 of 10291 true

nonmover matches are designated as matches by the computer with an error rate of 0.2 percent. An additional 55 true matches are in the set of designated clerical pairs and also quickly obtained. Thus, the computer matcher allows 96 percent (= 9856/10291) of the true nonmover matches to be obtained quickly. For Columbia and Washington, the corresponding percentages are 97 (= 6798/6984) and 97 (= 1899/1950), respectively.

Table 7. True Matches in Computer Categories Various Procedures, Columbia 6984 True Matches, 7649 Records Pairs Agreeing on Cluster and First Character Surname 1/

	--computer designation-- match		clerical	
truth->	match	non-match	match	non-match
crude	2429/ 7		4327/119	
param	6449/ 22		327/ 92	
param2	6655/ 13		135/ 35	
em (1988)	6719/ 13		78/ 22	
em2	6762/ 13		37/ 20	
freq (1990)	6792/ 11		6/ 9	

1/ Approximately 180 true matches disagree on first character of surname and are not eligible for inclusion in the table. Of the 180, approximately 80 reside in households where some is correctly matched.

Because printouts for clerical review and by PES household and some of the individuals not designated as matches or clerical pairs reside in households having individuals containing matching individuals, the additional individuals are also found during clerical review. The

Table 8. True Matches in Computer Categories
Various Procedures, Washington
950 True Matches, 2214 Records
Pairs Agreeing on Cluster and First Character Surname 1/

	--computer designation-- match		clerical	
truth->	match	non-match	match	non-match
crude	1307/	3	564/	98
param	1250/	5	614/	88
param2	1765/	4	134/	41
em (1988)	1749/	4	149/	29
em2	1795/	3	107/	29
freq (1990)	1892/	4	7/	9

1/ Approximately 40 true matches disagree on first character of surname and are not eligible for inclusion in the table. Of the 40, approximately 20 reside in households where some is correctly matched.

percentages of these individuals are 2, 1-2, and 1-2 percent in St Louis, Columbia, and Washington, respectively. Thus, individuals quickly found via computer match/printout procedures are 98, 98-99, and 98-99 percent in St Louis, Columbia, and Washington, respectively.

4. DISCUSSION

4.1. How the Quality of Files Effects Matching Results

The overall success of the procedures introduced in this paper is highly dependent on the facts that individual fields contain relatively few typographical variations and that there are redundant fields. Comparing strings having relatively severe typographical variations yields disagreement weights rather than the agreement weights that would occur without the typographical variation. If fields were without typographical variation, then having three crucial fields such as last name, first name, and house number might be sufficient to identify uniquely an individual. Having additional matching fields such as street name, age, and telephone number would be redundant. For those records having typographical variation or missing fields, the redundancy provides the extra information needed so that the computerized procedures can delineate many more pairs as matches.

Winkler (1989b, section 3.5) provides an example of matching techniques that are quite similar to the techniques of this paper for files of administrative records. On an absolute basis, overall matching results are much worse than the results of this paper. The reasons are that there is

relatively less information for matching and available information is often missing or inaccurate. For these files, matching procedures that use EM-derived parameters, frequency-based techniques, and string comparator metrics also yield better results than those procedures that do not.

The data of this paper are suitable for evaluating matching procedures becauses essentially all matches were found and correctly identified. The identification is with codes specifying to which record a record is matched. All basic identifying information was carefully checked and rechecked. In particular, virtually no matches were found among the set of code-identified nonmatches using a variety of procedures.

4.2. String Comparison and Weight Adjustment in the Fellegi-Sunter Model

For matching applications of files having significantly different characteristics (i.e., matching fields) from those of the files of this paper, string comparator weighting adjustments may have to be remodeled.

In all matching situations, it seems likely that modeling partial agreement should improve matching efficacy because the proportions of exact agreement on key matching fields can be quite low. For the files of this paper, the proportions of true matches agreeing on a character-by-character basis (Φ_n=1.0) are approximately 76 percent for first name and approximately 86 percent for last name (Table 9).

Table 9. Proportional Agreement by String Comparator Values Key Fields by Geography

	StL	Col	Wash
First			
M_n=1.0	0.75	0.82	0.75
$M_n \geq$0.6	0.93	0.94	0.93
Last			
M_n=1.0	0.85	0.88	0.86
$M_n \geq$0.6	0.95	0.96	0.96

Most of the reason that the crude and param methods (Tables 6, 7, and 8) perform relatively poorly is that they only assign the full agreement weight for first name and surname to the 86 and 76 percent, respectively, of the pairs that are truly matches and agree on a character-by-character basis. The remaining 24 and 14 percent, respectively, of the true matches get the full disagreement even though a number of them differ by only one or two characters.

4.3. Weights Produced by the EM Algorithm

The obvious reason that probabilities and associated agreement and disagreement weights produced by the EM Algorithm do not work that well is the failure of the independence assumption. Decision rules that incorporate interaction effects into the weighting on either an ad

hoc or formal basis may yield improved matching efficacy and/or automation.

4.3.1. Weights From General Statistical Fitting Procedures

Work by Winkler (1989a) and Thibaudeau (1989) using the formal model of three-way interactions (Bishop, Fienberg, and Holland 1975) and six matching variables have yielded normalized chi-square fits that are generally 200 times as accurate as fits under the independence model. Matching software, however, that uses three-way weights generally has not performed any better than the software of section 3 that utilizes simple agree/disagree weights.

Only with highly expert adjustments, does the 3-way matching software make slight improvements over the em results of section 3. The computation of 3-way probabilities, using theoretical and computational results generalizing results of Haberman (1977, 1976), are excessively slow to converge. There are two chief difficulties. The first is that the maximization step of the EM Algorithm for 3-way interactions is via an iterative fitting procedure whereas for independent situations the maximization step is closed form. The second is that different starting points typically yields different limiting solutions in the 3-way model whereas the limiting solutions are typically unique in the independent case.

Recently, Thibaudeau (1990) developed a scoring algorithm approach for fitting 3-way interaction models that converges much more rapidly than the previous fitting methods. In terms of chi-square statistics, the fits are only 20 times as good as the independent fits and would be rejected by any reasonable hypothesis test. They do, however, yield decision rules that are better than those using parameters produced by the independent EM Algorithm. While the resultant parameters do not yield decision rules that work as well as the best of this paper, they do yield improvements without the expert, file-specific adjustments that are used for the results of section 3. The reason for the improved decision rules is that the fitting procedures are better are finding local maxima of the likelihood and at rejecting unacceptable starting points.

4.3.2. Comparison With Existing Weighting Methods

All other existing computer matching systems of which we are aware use time-consuming, trial-and-error procedures for estimating matching parameters. Generally, initial guesses of matching parameters are used, matched pairs are reviewed, new matching parameters guessed at and the matching is repeated until satisfactory matching results obtained. In the hands of experts, such procedures are generally quite robust.

As the Generalized Iterative Record Linkage System of Statistics Canada (1983) is the only system to systematize the iterative review process and, thus, make it relatively usable by nonexperts, we use it as a point of comparison. Because we know truth and falsehood of matches, we merely assume that the known marginals P(agree specific field | Match) are the result of the iterative procedure. The Statistics Canada iterative procedures assume that convergence is generally to P(agree specific field | Match). Statistics Canada uses random agreement weights for P(agree specific field | Nonmatch) just as we do.

The matching results (Tables 6, 7, and 8) for the param procedure correspond roughly to a Statistics Canada procedure that does not use frequency-based weights and the param2 procedure adds the string comparator metrics. The comparisons between the em procedure and the param2 show that the EM-produced weights perform better. The reason is that, while both set of parameters are associated with independent distributions, the EM-produced weights are not constrained to satisfy additional marginal constraints P(agree specific field | Match).

4.4. Frequency-Based Weights

For the matching applications of this paper frequency-based weights make a noticeable improvement (e.g., section 3) over simple agree/disagree weights. The improvement is basically due to pairs having relatively rare first names and surnames that are designated as possible matches using simple agree/disagree weights but are designated as matches using frequency-based matches.

For applications to files having relatively poor distinguishing information (e.g., Winkler 1989b, section 3.5) relative improvements can be even greater. In such applications, only first and last name are generally available and they have relatively greater typographical variation. With simple agree/disagree weights very few pairs can be reliably designated as matches or clerical pairs. With frequency-based weighting a moderate percentage of pairs change from nonmatch to clerical pairs.

4.5. Linear Sum Assignment Procedure

The reason the assignment works well for the particular applications considered in this paper is the nature of the data for individuals in households that are brought together. Generally, all the individuals will have good information or they will all have substantial missing or erroneous data (e.g., as a result of proxy information). In both cases clerical review sets are necessarily much larger.

In the first case, say, the assignment procedure eliminates all pairs (such as brother-sister or husband-wife) that might have moderately high weights. The UPPER cutoff would have to be raised to force them into the set of clerical pairs. If such pairs were retained, additional pairs based having only moderately good identifying information (e.g., with some typographical differences) would no longer be in the set of computer-designated matches.

We observe that if one reasonable assumption is made, then the decision rules that incorporates the linear sum assignment procedure are still optimal. The assumption is that the basic weighting procedure (prior to the linear sum assignment) always makes the best assignment (either to the true match or the more closely agreeing nonmatch). Then the decision rule modified with the linear sum assignment procedure is still optimal. Indeed, under the assumption, the second through nth best assignments will always be associated with true nonmatches and the likelihood ratio associated with them should be forced to 0.

4.6. Determination of Cutoff Weights

Fellegi and Sunter (1969, section 3.3.2 and 3.7; see also Jaro 1989, section 3.4) provide methods for calculating the cutoff weights (or threshold values) UPPER and LOWER provided the probabilities m( ) and u( ) can be given a strict probabilistic interpretation. In such a situation, if we are given an upper bound on the false match rate, we should be able to use m( ) and u( ) directly to computer UPPER and LOWER.

Our experience has been that the probabilities m( ) and u( ) for each of the methods of section 3 (e.g., param, em, etc.) are never suitable for direct estimation of the cutoffs because the estimated m( ) and u( ) deviate much too severely from the true underlying probabilities. Part of the deviation is due to the failure of the independence assumption that is typically used. Other deviations can be caused by having a small subpopulation that has characteristics significantly different the the population as a whole.

In practice, we determine UPPER and LOWER by reviewing a set of pairs that are printed by

decreasing matching weight. Based on experience, it is quite rapid to determine the cutoffs with a rough a priori bound on the error rates.

Belin and Rubin (1990) have introduced a method for directly determining the cutoffs at desired error levels for any weighting curve. Their method necessitates having a training set of weights associated with matches and nonmatches. Under the assumption that the weighting curve is a mixture of normals that can be transformed to normality via Box-Cox techniques, they provide an EM-type Algorithm to estimate the curves for new data sets. The a priori information is used primarily to get the shapes of the curves. Their method yields estimates of the error rate and estimates of the corresponding confidence intervals.

5. SUMMARY AND CONCLUSIONS

5.1. Summary

This paper provides a methodology for computer matching that generally falls under the formal decision-theoretic procedures given by Fellegi and Sunter. The two crucial ideas for improving matching efficacy are having parameters estimated automatically via a modified version of the Expectation-Maximization Algorithm and using string comparator metrics for assigning adjusted weights to pairs of strings that agree almost exactly rather than using full disagreement weights.

5.2. Conclusion

At present, the very high quality of the matching results appears very dependent on having files similar to those that will be available during 1990 Post Enumeration Survey processing.

5.3. Future Work

New research on more general, theoretically valid string comparator metrics and associated weighting adjustments is just beginning. Research into decision rules that allow for general interactions rather than independent iteractions continues and shows promise.

Applications to general administrative lists and to lists of businesses is just beginning. Matching of business lists is highly dependent on the quality of name and address parsing software.

REFERENCES

Belin, T. R. and Rubin, D. B. (1990), "Calibration of Errors in Computer Matching for Census Undercount," *American Statistical Association, Proceedings of the Section on Government Statistics*, to appear.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis*, MIT Press, Cambridge, MA.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, **B**, **39,** 1-38.

Fellegi, I. P., and Sunter, A. B. (1969), "A Theory for Record Linkage," *Journal of the American Statistical Association*, **40**, 1183-1210.

Haberman, S. J. (1976), "Iterative Scaling for Log-Linear Model for Frequency Tables Derived by Indirect Observation, *American Statistical Association, Proceedings of the Section on Statistical Computing*, 45-50.

Haberman, S. J. (1977), "Product Models for Frequency Tables Involving Indirect Observation," Annals of Statistics, **5** 1124-1147.

Haberman, S. (1979), *Analysis of Qualitative Data*, Academic Press. New York.

Jaro, M. A. (1989), "Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida,"*Journal of the American Statistical Association*, **89**, 414-420.

Newcombe, H. B. (1988) *Handbook of Record Linkage: Methods for Health and Statistical Studies, Administration, and Business*. Oxford: Oxford University Press.

Newcombe, H. B., Kennedy, J. M., Axford, S. J., and James, A. P. (1959), "Automatic Linkage of Vital Records," *Science*, **130** 954-959.

Rowe, E. (1987), private communication of FORTRAN and Pascal versions of linear sum assignment algorithm.

Statistics Canada (1983) "Generalized Iterative Record Linkage System", Systems Development Division.

Thibaudeau, Y. (1989), "Fitting Log-Linear Models When Some Dichotomous Variables are Unobservable,"*American Statistical Association, Proceedings of the Section on Biometrics*, to appear.

Thibaudeau, Y. (1990), "Improving the Performance of Computer Matching Algorithms Through Better Choices of Parameters," paper presented at Annual *American Statistical Association* meeting in Anaheim, California.

Winkler, W. E. (1985), "Preprocessing of Lists and String Comparison," in *Record Linkage Techniques- 1985*, edited by W. Alvey and B. Kilss, U.S. Internal Revenue Service, Publication 1299 (2-86), 181-187.

Winkler, W. E. (1988), "Using the EM Algorithm for Weight Computation in the Fellegi-Sunter Model of Record Linkage," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 667-671.

Winkler, W. E. (1989a), "Near Automatic Weight Computation in the Fellegi-Sunter Model of Record Linkage," *Proceedings of the Fifth Census Bureau Annual Research Conference*, 145-155.

Winkler, W. E. (1989b), "Frequency-based Matching in the Fellegi-Sunter Model of Record Linkage," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 788-793.

Winkler, W. E. (1989c, "The Interaction of Record Linkage Theory and Practice," Ottawa, Ontario, Canada, August 1989, *Proceedings of the Record Linkage Sessions and Workshop*, 139-148.

Winkler, W. E. (1990a), "Bootstrap Evaluation of Calibration Procedures used for Decision Rules in the Fellegi-Sunter Model of Record Linkage," *Computing Science and Statistics, Proceedings of Interface '90*, edited by LePage, R. and Billard, L., to appear

Winkler, W. E. (1990b), "String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, to appear.

EXHIBIT 9

American Political Science Review (2019) 113, 2, 353–371
doi:10.1017/S0003055418000783
© American Political Science Association 2019

Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records

TED ENAMORADO *Princeton University*
BENJAMIN FIFIELD *Princeton University*
KOSUKE IMAI *Harvard University*

Since most social science research relies on multiple data sources, merging data sets is an essential part of researchers' workflow. Unfortunately, a unique identifier that unambiguously links records is often unavailable, and data may contain missing and inaccurate information. These problems are severe especially when merging large-scale administrative records. We develop a fast and scalable algorithm to implement a canonical model of probabilistic record linkage that has many advantages over deterministic methods frequently used by social scientists. The proposed methodology efficiently handles millions of observations while accounting for missing data and measurement error, incorporating auxiliary information, and adjusting for uncertainty about merging in post-merge analyses. We conduct comprehensive simulation studies to evaluate the performance of our algorithm in realistic scenarios. We also apply our methodology to merging campaign contribution records, survey data, and nationwide voter files. An open-source software package is available for implementing the proposed methodology.

INTRODUCTION

As the amount and diversity of available data sets rapidly increase, social scientists often harness multiple data sources to answer substantive questions. Indeed, merging data sets, in particular large-scale administrative records, is an essential part of cutting-edge empirical research in many disciplines (e.g., Ansolabehere and Hersh 2012; Einav and Levin 2014; Jutte, Roos, and Browne 2011). Data merging can be consequential. For example, the American National Election Studies (ANES) and Cooperative Congressional Election Study (CCES) validate self-reported turnout by merging their survey data with a nationwide voter file where only the matched respondents are treated as registered voters. Although Ansolabehere and Hersh (2012) advocate the use of such a validation procedure, Berent, Krosnick, and Lupia (2016) argue that the discrepancy between self-reported and validated turnout is due to the failure of the merge procedure rather than social desirability and nonresponse bias.

Merging data sets is straightforward if there exists a unique identifier that unambiguously links records from different data sets. Unfortunately, such a unique identifier is often unavailable. Under these circumstances, some researchers have used a deterministic algorithm to automate the merge process (e.g., Adena et al. 2015; Ansolabehere and Hersh 2017; Berent, Krosnick, and Lupia 2016; Bolsen, Ferraro, and Miranda 2014; Cesarini et al. 2016; Figlio and Guryan 2014; Giraud-Carrier et al. 2015; Hill 2017; Meredith and Morse 2014) whereas others have relied on a proprietary algorithm (e.g., Ansolabehere and Hersh 2012; Engbom and Moser 2017; Figlio and Guryan 2014; Hersh 2015; Hill and Huber 2017; Richman, Chattha, and Earnest 2014). However, these methods are not robust to measurement error (e.g., misspelling) and missing data, which are common to social science data. Furthermore, deterministic merge methods cannot quantify the uncertainty of the merging procedure and instead typically rely on arbitrary thresholds to determine the degree of similarity sufficient for matches.[1] This means that post-merge data analyses fail to account for the uncertainty of the merging procedure, yielding a bias due to measurement error. These methodological

Ted Enamorado, Ph.D. Candidate, Department of Politics, Princeton University, tede@princeton.edu, http://www.tedenamorado.com.
Benjamin Fifield, Ph.D. Candidate, Department of Politics, Princeton University, bfifield@princeton.edu, http://www.benfifield.com.
Kosuke Imai, Professor, Department of Government and Department of Statistics, Harvard University. imai@harvard.edu, https://imai.fas.harvard.edu.

The proposed methodology is implemented through an open-source R package, `fastLink`: Fast Probabilistic Record Linkage, which is freely available for download at the Comprehensive R Archive Network (CRAN; https://CRAN.R-project.org/package=fastLink). We thank Bruce Willsie of L2 and Steffen Weiss of YouGov for data and technical assistance, Jake Bowers, Seth Hill, Johan Lim, Marc Ratkovic, Mauricio Sadinle, five anonymous reviewers, and audiences at the 2017 Annual Meeting of the American Political Science Association, Columbia University (Political Science), Fifth Asian Political Methodology Meeting, Gakusyuin University (Law), Hong Kong University of Science and Technology, the Institute for Quantitative Social Science (IQSS) at Harvard University, the Quantitative Social Science (QSS) colloquium at Princeton University, Universidad de Chile (Economics), Universidad del Desarrollo, Chile (Government), the 2017 Summer Meeting of the Society for Political Methodology, the Center for Statistics and the Social Sciences (CSSS) at the University of Washington for useful comments and suggestions. Replication materials can be found on Dataverse at: https://doi.org/10.7910/DVN/YGUHTD.

Received: August 8, 2017; revised: May 13, 2018; accepted: October 17, 2018. First published online: January 2, 2019.

[1] These thresholds are highly dependent on data. For example, Ansolabehere and Hersh (2017) find that using three fields with exact matches as the threshold works well for the Texas voter file, but the same threshold may not work for other data. In contrast, probabilistic methods can automatically weight observations.

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

challenges are amplified especially when merging large-scale administrative records.

We demonstrate that social scientists should use probabilistic models rather than deterministic methods when merging large data sets. Probabilistic models can quantify the uncertainty inherent in many merge procedures, offering a principled way to calibrate and account for false positives and false negatives. Unfortunately, although there exists a well-known statistics literature on probabilistic record linkage (e.g., Harron, Goldstein, and Dibben 2015; Herzog, Scheuren, and Winkler 2007; Winkler 2006b), the current open-source implementation does not scale to large data sets commonly used in today's social science research. We address this challenge by developing a fast and scalable implementation of the canonical probabilistic record linkage model originally proposed by Fellegi and Sunter (1969). Together with parallelization, this algorithm, which we call `fastLink`, can be used to merge data sets with millions of records in a reasonable amount of time using one's laptop computer. Additionally, building on the previous methodological literature (e.g., Lahiri and Larsen 2005), we show (1) how to incorporate auxiliary information such as population name frequency and migration rates into the merge procedure and (2) how to conduct post-merge analyses while accounting for the uncertainty about the merge process. We describe these methodological developments in the following section.

We then describe the comprehensive simulation studies to evaluate the robustness of `fastLink` to several factors including the size of data sets, the proportion of true matches, measurement error, and missing data proportion and mechanisms. A total of 270 simulation settings consistently show that `fastLink` significantly outperforms the deterministic methods. Although the proposed methodology produces high-quality matches in most situations, the lack of overlap between two data sets often leads to large error rates, suggesting that effective blocking is essential when the expected number of matches is relatively small. Furthermore, `fastLink` appears to perform at least as well as recently proposed probabilistic approaches (Sadinle 2017; Steorts 2015). Importantly, our merge method is faster and scales to larger data sets than these state-of-art methods.

Next, we present two empirical applications. First, we revisit Hill and Huber (2017) who examine the ideological differences between donors and nondonors by merging the CCES data of more than 50,000 survey respondents, with the a campaign contribution database of over five million donor records (Bonica 2013). We find that the matches identified by `fastLink` are at least as high quality as those identified by the proprietary method, which was used by the original authors. We also improve the original analysis by incorporating the uncertainty of the merge process in the post-merge analysis. We show that although the overall conclusion remains unchanged, the magnitude of the estimated effects is substantially smaller.

As the second application, we merge two nationwide voter files of over 160 million voter records each, representing one of the largest data merges ever conducted in social science research.[2] By merging voter files over time, scholars can study the causes and consequences of partisan residential segregation (e.g., Tam Cho, Gimpel, and Hui 2013; Mummolo and Nall 2016) and political analytics professionals can develop effective microtargeting strategies (e.g., Hersh 2015). We show how to incorporate available within-state and across-state migration rates in the merge process. Given the enormous size of the data sets, we propose a two-step procedure where we first conduct a within-state merge for each state followed by across-state merges for every pair of states. The proposed methodology is able to match about 95% of voters, which is about 30-percentage points greater than the exact matching method. Although it is more difficult to find across-state movers, we are able to find 20 times as many such voters than the existing matching method.

Finally, we give concluding remarks. We provide an open-source R software package `fastLink`: Fast Probabilistic Record Linkage, which is freely available at the Comprehensive R Archive Network (CRAN; https://CRAN.R-project.org/package=fastLink) for implementing our methodology so that other researchers can effectively merge data sets in their own projects.

THE PROPOSED METHODOLOGY

In this section, we first introduce the canonical model of probabilistic record linkage originally proposed by Fellegi and Sunter (1969). We describe several improvements we make to this model, including a fast and scalable implementation, the use of auxiliary information to inform parameter estimation, and the incorporation of uncertainty about the merge process in post-merge analyses.

The Setup

Suppose that we wish to merge two data sets, $\mathcal{A}$ and $\mathcal{B}$, which have sample sizes of N_A and N_B, respectively. We use K variables, which are common to both data sets, to conduct the merge. We consider all possible pair-wise comparisons between these two data sets. For each of these $N_A \times N_B$ distinct pairs, we define an agreement vector of length K, denoted by $\gamma(i,j)$, whose kth element $\gamma_k(i, j)$ represents the discrete level of within-pair similarity for the kth variable between the ith observation of data set $\mathcal{A}$ and the jth observation of data set $\mathcal{B}$. Specifically, if we have a total of L_k similarity levels for the kth variable, then the corresponding element of the agreement vector can be defined as,

[2] Although Hersh (2015) conducted a large-scale data merge, he relied on a proprietary algorithm. Others such as Ansolabehere and Hersh (2017) and Tam Cho, Gimpel, and Hui (2013) match data sets of several million voters each, but neither of these studies approaches the scale of our applications. Note that the US Census Bureau routinely conducts large-scale data merges for decennial census (Winkler, Yancey, and Porter 2010).

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

TABLE 1. An Illustrative Example of Agreement Patterns.

	Name				Address	
	First	Middle	Last	Date of birth	House	Street
Data set $\mathcal{A}$						
1	James	V	Smith	12-12-1927	780	Devereux St.
2	Robert	NA	Martines	01-15-1942	60	16th St.
Data set $\mathcal{B}$						
1	Michael	F	Martinez	02-03-1956	4	16th St.
2	James	D	Smithson	12-12-1927	780	Dvereuux St.
Agreement patterns						
$\mathcal{A}.1 - \mathcal{B}.1$	Different	Different	Different	Different	Different	Different
$\mathcal{A}.1 - \mathcal{B}.2$	Identical	Different	Similar	Identical	Identical	Similar
$\mathcal{A}.2 - \mathcal{B}.1$	Different	NA	Similar	Different	Different	Identical
$\mathcal{A}.2 - \mathcal{B}.2$	Different	NA	Different	Different	Different	Different

The top panel of the table shows two artificial data sets, $\mathcal{A}$ and $\mathcal{B}$, each of which has two records. The bottom panel shows the agreement patterns for all possible pairs of these records. For example, the second line of the agreement patterns compares the first record of the data set $\mathcal{A}$ with the second record of the data set $\mathcal{B}$. These two records have an identical information for first name, date of birth, and house number; similar information for last name and street name; and different information for middle name. A comparison involving at least one missing value is indicated by NA.

$$\gamma_k(i,j) = \left\{ \begin{array}{c} 0 \\ 1 \\ \vdots \\ L_k - 2 \\ L_k - 1 \end{array} \right\} \begin{array}{l} \text{different} \\ \\ \text{similar} \\ \\ \text{identical} \end{array} \quad \textbf{(1)}$$

The proposed methodology allows for the existence of missing data. We define a missingness vector of length K, denoted by $\delta(i,j)$, for each pair (i,j) where its kth element $\delta_k(i,j)$ equals 1 if at least one record in the pair has a missing value in the kth variable and is equal to 0 otherwise.

Table 1 presents an illustrative example of agreement patterns based on two artificial data sets, $\mathcal{A}$ and $\mathcal{B}$, each of which has two records. In this example, we consider three possible values of $\gamma_k(i,j)$ for first name, last name, and street name, i.e., $L_k = 3$ (different, similar, nearly identical), whereas a binary variable is used for the other fields, i.e., $L_k = 2$ (different, nearly identical). The former set of variables requires a similarity measure and threshold values. We use the Jaro–Winkler string distance (Jaro 1989; Winkler 1990), which is a commonly used measure in the literature (e.g., Cohen, Ravikumar, and Fienberg 2003; Yancey 2005).[3] Because the Jaro–Winkler distance is a continuous measure whose values range from 0 (different) to 1 (identical), we discretize it so that $\gamma_k(i,j)$ takes an integer value between 0 and $L_k - 1$ as defined in equation (1). Suppose that we use three levels (i.e., different, similar, and nearly identical) based on the threshold values of 0.88 and 0.94 as recommended by Winkler (1990). Then, when comparing the last names in Table 1, we find that, for example, Smith and Smithson are similar (a Jaro–Winkler distance of 0.88) whereas Smith and Martinez are different (a Jaro–Winkler distance of 0.55).[4]

The above setup implies a total of $N_{\mathcal{A}} \times N_{\mathcal{B}}$ comparisons for each of K fields. Thus, the number of comparisons grows quickly as the size of data sets increases. One solution is to use blocking and avoid comparisons that should not be made. For example, we may make comparisons within gender group only. While it is appealing because of computational efficiency gains, Winkler (2005) notes that blocking often involves ad hoc decisions by researchers and faces difficulties when variables have missing values and measurement error. Here, we focus on the data merge within a block and refer interested readers to Christen (2012) and Steorts et al. (2014) for comprehensive reviews of blocking techniques.[5] We also note a related technique, called filtering, which has the potential to overcome the weaknesses of traditional blocking methods by discarding pairs that are unlikely to be matches when fitting a probabilistic model (Murray 2016).

The Canonical Model of Probabilistic Record Linkage

The Model and Assumptions

We first describe the most commonly used probabilistic model of record linkage (Fellegi and Sunter 1969). Let a latent mixing variable M_{ij} indicate whether a pair of records (the ith record in the data set $\mathcal{A}$ and the jth record in the data set $\mathcal{B}$) represents a match. The model has the following simple finite mixture structure (e.g., Imai and Tingley 2012; McLaughlan and Peel 2000):

$$\gamma_k(i,j) \mid M_{ij} = m \overset{\text{indep.}}{\sim} \text{Discrete}(\boldsymbol{\pi}_{km}), \quad \textbf{(2)}$$

$$M_{ij} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\lambda), \quad \textbf{(3)}$$

[3] Online Supplementary Information (SI) S1 describes how the Jaro–Winkler string distance is calculated.

[4] As shown in the subsection *Computational Efficiency* and Appendix A, the discretization of the distance measure leads to substantial computational efficiency when making pairwise comparison for each linkage field.

[5] The parameters of record linkage models must be interpreted separately for each block (Murray 2016).

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

where $\boldsymbol{\pi}_{km}$ is a vector of length L_k, containing the probability of each agreement level for the kth variable given that the pair is a match ($m = 1$) or a nonmatch ($m = 0$), and λ represents the probability of a match across all pairwise comparisons. Through $\boldsymbol{\pi}_{k0}$, the model allows for the possibility that two records can have identical values for some variables even when they do not represent a match.

This model is based on two key independence assumptions. First, the latent variable M_{ij} is assumed to be independently and identically distributed. Such an assumption is necessarily violated if, for example, each record in the data set $\mathcal{A}$ should be matched with no more than one record in the data set $\mathcal{B}$. In theory, this assumption can be relaxed (e.g., Sadinle 2017) but doing so makes the estimation significantly more complex and reduces its scalability (see Online SI S8). Later in the paper, we discuss how to impose such a constraint without sacrificing computational efficiency. Second, the conditional independence among linkage variables is assumed given the match status. Some studies find that the violation of this assumption leads to unsatisfactory performance (e.g., Belin and Rubin 1995; Herzog, Scheuren, and Winkler 2010; Larsen and Rubin 2001; Thibaudeau 1993; Winkler and Yancey 2006). In Online SI S4, we show how to relax the conditional independence assumption while keeping our scalable implementation.

In the literature, researchers often treat missing data as disagreements, i.e., $\gamma_k(i,j) = 0$ if $\delta_k(i,j) = 1$ (e.g., Goldstein and Harron 2015; Ong et al. 2014; Sariyar, Borg, and Pommerening 2012). This procedure is problematic because a true match can contain missing values. Other imputation procedures also exist but none of them has a theoretical justification or appears to perform well in practice.[6] To address this problem, following Sadinle (2014, 2017), we assume that data are missing at random (MAR) conditional on the latent variable M_{ij},

$$\delta_k(i,j) \perp\!\!\!\perp \gamma_k(i,j) \,|\, M_{ij},$$

for each $i = 1, 2, \ldots, N_{\mathcal{A}}$, $j = 1, 2, \ldots, N_{\mathcal{B}}$, and $k = 1, 2, \ldots, K$. Under this MAR assumption, we can simply ignore missing data. The observed-data likelihood function of the model defined in equations (2) and (3) is given by,

$$\mathcal{L}_{\text{obs}}(\boldsymbol{\lambda}, \boldsymbol{\pi} \,|\, \boldsymbol{\delta}, \boldsymbol{\gamma}) \propto \prod_{i=1}^{N_{\mathcal{A}}} \prod_{j=1}^{N_{\mathcal{B}}} \left\{ \sum_{m=0}^{1} \lambda^m (1-\lambda)^{1-m} \prod_{k=1}^{K} \left(\prod_{\ell=0}^{L_k - 1} \pi_{km\ell}^{1\{\gamma_k(i,j)=\ell\}} \right)^{1-\delta_k(i,j)} \right\},$$

where $\pi_{km\ell}$ represents the ℓth element of probability vector $\boldsymbol{\pi}_{km}$, i.e., $\pi_{km\ell} = \Pr\big(\gamma_k(i,j) = \ell \,|\, M_{ij} = m\big)$. Because the direct maximization of the observed-data log-likelihood function is difficult, we estimate the model parameters using the Expectation-Maximization (EM) algorithm (see Online SI S2).

[6] For example, although Goldstein and Harron (2015) suggest the possibility of treating a comparison that involves a missing value as a separate agreement value, but Sariyar, Borg, and Pommerening (2012) find that this approach does not outperform the standard method of treating missing values as disagreements.

The Uncertainty of the Merge Process

The advantage of probabilistic models is their ability to quantify the uncertainty inherent in merging. Once the model parameters are estimated, we can compute the match probability for each pair using Bayes rule,[7]

$$\xi_{ij} = \Pr\big(M_{ij} = 1 \mid \boldsymbol{\delta}(i,j), \boldsymbol{\gamma}(i,j)\big) = \frac{\lambda \prod_{k=1}^{K} \left(\prod_{\ell=0}^{L_k - 1} \pi_{k1\ell}^{1\{\gamma_k(i,j)=\ell\}} \right)^{1-\delta_k(i,j)}}{\sum_{m=0}^{1} \lambda^m (1-\lambda)^{1-m} \prod_{k=1}^{K} \left(\prod_{\ell=0}^{L_k - 1} \pi_{km\ell}^{1\{\gamma_k(i,j)=\ell\}} \right)^{1-\delta_k(i,j)}}. \quad (4)$$

In the subsection *Post-merge Analysis*, we show how to incorporate this match probability into post-merge regression analysis to account for the uncertainty of the merge process.

Although in theory a post-merge analysis can use all pairs with nonzero match probabilities, it is often more convenient to determine a threshold S when creating a merged data set. Such an approach is useful especially when the data sets are large. Specifically, we call a pair (i,j) a match if the match probability ξ_{ij} exceeds S. There is a clear trade-off in the choice of this threshold value. A large value of S will ensure that most of the selected pairs are correct matches but may fail to identify many true matches. In contrast, if we lower S too much, we will select more pairs but many of them may be false matches. Therefore, it is important to quantify the degree of these matching errors in the merging process.

One advantage of probabilistic models over deterministic methods is that we can estimate the false discovery rate (FDR) and the false negative rate (FNR). The FDR represents the proportion of false matches among the selected pairs whose matching probability is greater than or equal to the threshold. We estimate the FDR using our model parameters as follows:

$$\Pr\Big(M_{ij} = 0 \mid \xi_{ij} \geq S\Big) = \frac{\sum_{i=1}^{N_{\mathcal{A}}} \sum_{j=1}^{N_{\mathcal{B}}} \mathbf{1}\Big\{\xi_{ij} \geq S\Big\}\Big(1 - \xi_{ij}\Big)}{\sum_{i=1}^{N_{\mathcal{A}}} \sum_{j=1}^{N_{\mathcal{B}}} \mathbf{1}\Big\{\xi_{ij} \geq S\Big\}}, \quad (5)$$

whereas the FNR, which represents the proportion of true matches that are not selected, is estimated as

$$\Pr\Big(M_{ij} = 1 \mid \xi_{ij} < S\Big) = \frac{\sum_{i=1}^{N_{\mathcal{A}}} \sum_{j=1}^{N_{\mathcal{B}}} \xi_{ij} \mathbf{1}\Big\{\xi_{ij} < S\Big\}}{\lambda N_{\mathcal{A}} N_{\mathcal{B}}}. \quad (6)$$

Researchers typically select, at their own discretion, the value of S such that the FDR is sufficiently small. But, we also emphasize the FNR because a strict threshold can lead to many false negatives.[8] In our simulations and empirical studies, we find that the

[7] This is known as the maximum a posteriori estimate.

[8] A more principled solution to the threshold S selection problem would require data for which the true matching status $M(i, j)$ is known—so that one can select the value of S to minimize the classification error. However, in record linkage problems, only in rare occasions do labeled data sets exist. See Larsen and Rubin (2001), Feigenbaum (2016), and Enamorado (2018) for approaches that directly incorporate labeled data.

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

results are not particularly sensitive to the choice of threshold value, although in other applications, scholars found ex-post adjustments are necessary for obtaining good estimates of error rates (e.g., Belin and Rubin 1995; Larsen and Rubin 2001; Murray 2016; Thibaudeau 1993; Winkler 1993; Winkler 2006a).

In the merging process, for a given record in the data set $\mathcal{A}$, it is possible to find multiple records in the data set $\mathcal{B}$ that have high match probabilities. In some cases, multiple observations have an identical value of match probability, i.e., $\xi_{ij} = \xi_{ij'}$ with $j \neq j'$. Following the literature (e.g., McVeigh and Murray 2017; Sadinle 2017; Tancredi and Liseo 2011), we recommend that researchers analyze all matched observations by weighting them according to the matching probability (see the subsection *Post-Merge Analysis*). If researchers wish to enforce a constraint that each record in one data set is only matched at most with one record in the other data set, they may follow a procedure described in Online SI S5.

Incorporating Auxiliary Information

Another advantage of the probabilistic model introduced above is that we can incorporate auxiliary information in parameter estimation. This point has not been emphasized enough in the literature. Here, we briefly discuss two adjustments using auxiliary data—first, how to adjust for the fact that some names are more common than others, and second, how to incorporate aggregate information about migration. More details can be found in Online SI S6.

Because some first names are more common than others, they may be more likely to be false matches. To adjust for this possibility without increasing the computational burden, we formalize the conditions under which the ex-post correction originally proposed by Winkler (2000) is well-suited for this purpose. Briefly, the probability of being a match will be up-weighted or down-weighted given the true frequencies of different first names (obtained, for instance, from Census data) or observed frequencies of each unique first name in the data (see Online SI S6.3.1).

Furthermore, we may know *a priori* how many matches we should find in two data sets because of the knowledge and data on over-time migration. For instance, the Internal Revenue Service (IRS) publishes detailed information on migration in the United States from tax records (see https://www.irs.gov/uac/soi-tax-stats-migration-data). An estimate of the share of individuals who moved out of a state or who moved in-state can be easily reformulated as a prior on relevant parameters in the Fellegi–Sunter model and incorporated into parameter estimation (see Online SI S6.3.2).

Post-Merge Analysis

Finally, we discuss how to conduct a statistical analysis once merging is complete. One advantage of probabilistic models is that we can directly incorporate the uncertainty inherent to the merging process in the post-merge analysis. This is important because researchers often use the merged variable either as the outcome or as the explanatory variable in the post-merge analysis. For example, when the ANES validates self-reported turnout by merging the survey data with a nationwide voter file, respondents who are unable to be merged are coded as nonregistered voters. Given the uncertainty inherent to the merging process, it is possible that a merging algorithm fails to find some respondents in the voter file even though they are actually registered voters. Similarly, we may incorrectly merge survey respondents with other registered voters. These mismatches, if ignored, can adversely affect the properties of post-match analyses (e.g., Neter, Maynes, and Ramanathan 1965; Scheuren and Winkler 1993).

Unfortunately, most of the record linkage literature has focused on the linkage process itself without considering how to conduct subsequent statistical analyses after merging data sets.[9] Here, we build on a small literature about post-merge regression analysis, the goal of which is to eliminate possible biases due to the linkage process within the Fellegi–Sunter framework (e.g., Hof and Zwinderman 2012; Kim and Chambers 2012; Lahiri and Larsen 2005; Scheuren and Winkler 1993, 1997). We also clarify the assumptions under which a valid post-merge analysis can be conducted.

The Merged Variable as an Outcome Variable

We first consider the scenario, in which researchers wish to use the variable Z merged from the data set $\mathcal{B}$ as a proxy for the outcome variable in a regression analysis. We assume that this regression analysis is applied to all observations of the data set $\mathcal{A}$ and uses a set of explanatory variables $\mathbf{X}$ taken from this data set. These explanatory variables may or may not include the variables used for merging. In the ANES application mentioned above, for example, we may be interested in regressing the validated turnout measure merged from the nationwide voter file on a variety of demographic variables measured in the survey.

For each observation i in the data set $\mathcal{A}$, we obtain the mean of the merged variable, i.e., $\zeta_i = \mathbb{E}(Z_i^* \mid \gamma, \boldsymbol{\delta})$ where Z_i^* represents the true value of the merged variable. This quantity can be computed as the weighted average of the variable Z merged from the data set $\mathcal{B}$ where the weights are proportional to the match probabilities, i.e., $\zeta_i = \sum_{j=1}^{N_\mathcal{B}} \xi_{ij} Z_j / \sum_{j=1}^{N_\mathcal{B}} \xi_{ij}$. In the ANES application, for example, ζ_i represents the probability of turnout for survey respondent i in the data set $\mathcal{A}$ and can be computed as the weighted average of turnout among the registered voters in the voter file merged with respondent i. If we use thresholding and one-to-one match assignment so that each record in the data set $\mathcal{A}$ is matched with at most one record in the data set $\mathcal{B}$ (see the subsection *The Canonical Model of Probabilistic*

[9] An important exception includes a fully Bayesian approach outside of the Fellegi–Sunter framework, which we do not pursue here because of its limited scalability (see Tancredi and Liseo 2011; Gutman, Afendulis, and Zaslavsky 2013; Gutman et al. 2016; Dalzell and Reiter 2018).

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

Record Linkage), then we compute the mean of the merged variable as $\zeta_i = \sum_{j=1}^{N_{\mathcal{B}}} M_{ij}^* \xi_{ij} Z_j$ where M_{ij}^* is a binary variable indicating whether record i in the data set $\mathcal{A}$ is matched with record j in the data set $\mathcal{B}$ subject to the constraint $\sum_{j=1}^{N_{\mathcal{B}}} M_{ij}^* \leq 1$.

Under this setting, we assume that the true value of the outcome variable is independent of the explanatory variables in the regression conditional on the information used for merging, i.e.,

$$Z_i^* \perp\!\!\!\perp \mathbf{X}_i \mid (\boldsymbol{\delta}, \gamma), \tag{7}$$

for each $i = 1, 2, \ldots, N_{\mathcal{A}}$. The assumption implies that the merging process is based on all relevant information. Specifically, within an agreement pattern, the true value of the merged variable Z_i^* is not correlated with the explanatory variables $\mathbf{X}_i$. Under this assumption, the law of iterated expectation implies that regressing ζ_i on $\mathbf{X}_i$ gives the results equivalent to the ones based on the regression of Z_i^* on $\mathbf{X}_i$ in expectation.

$$\mathbb{E}(Z_i^* \mid \mathbf{X}_i) = \mathbb{E}\{\mathbb{E}(Z_i^* \mid \gamma, \boldsymbol{\delta}, \mathbf{X}_i) \mid \mathbf{X}_i\} = \mathbb{E}(\zeta_i \mid \mathbf{X}_i). \tag{8}$$

The conditional independence assumption may be violated if, for example, within the same agreement pattern, a variable correlated with explanatory variables is associated with merging error. Without this assumption, however, only the bounds can be identified (Cross and Manski 2002). Thus, alternative assumptions such as parametric assumptions and exclusion restrictions are needed to achieve identification (see Ridder and Moffitt 2007, for a review).

The Merged Variable as an Explanatory Variable

The second scenario we consider is the case where we use the merged variable as an explanatory variable. Suppose that we are interested in fitting the following linear regression model:

$$Y_i = \alpha + \beta Z_i^* + \boldsymbol{\eta}^\top \mathbf{X}_i + \varepsilon_i, \tag{9}$$

where Y_i is a scalar outcome variable and the strict exogeneity is assumed, i.e., $\mathbb{E}(\varepsilon_i \mid Z^*, \mathbf{X}) = 0$ for all i. We follow the analysis strategy first proposed by Lahiri and Larsen (2005) but clarify the assumptions required for their approach to be valid (see also Hof and Zwinderman 2012). Specifically, we maintain the assumption of no omitted variable for merging given in equation (7). Additionally, we assume that the merging variables are independent of the outcome variable conditional on the explanatory variables $\mathbf{Z}^*$ and $\mathbf{X}$, i.e.,

$$Y_i \perp\!\!\!\perp (\gamma, \boldsymbol{\delta}) \mid \mathbf{Z}^*, \mathbf{X}. \tag{10}$$

Under these two assumptions, we can consistently estimate the coefficients by regressing Y_i on ζ_i and $\mathbf{X}_i$,

$$\begin{aligned} \mathbb{E}(Y_i \mid \gamma, \boldsymbol{\delta}, \mathbf{X}_i) &= \alpha + \beta \mathbb{E}(Z_i^* \mid \gamma, \boldsymbol{\delta}, \mathbf{X}_i) + \boldsymbol{\eta}^\top \mathbf{X}_i + \mathbb{E}(\varepsilon_i \mid \gamma, \boldsymbol{\delta}, \mathbf{X}_i) \\ &= \alpha + \beta \zeta_i + \boldsymbol{\eta}^\top \mathbf{X}_i, \end{aligned} \tag{11}$$

where the second equality follows from the assumptions and the law of iterated expectation.

We generalize this strategy to the maximum likelihood (ML) estimation, which, to the best of our knowledge, has not been considered in the literature (but see Kim and Chambers (2012) for an estimating equations approach),

$$Y_i \mid Z_i^*, \mathbf{X}_i \overset{\text{indep.}}{\sim} P_\theta(Y_i \mid Z_i^*, \mathbf{X}_i), \tag{12}$$

where θ is a vector of model parameters. To estimate the parameters of this model, we maximize the following weighted log-likelihood function:

$$\hat{\boldsymbol{\theta}} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{N_{\mathcal{A}}} \sum_{j=1}^{N_{\mathcal{B}}} \xi_{ij}^* \log P_\theta(Y_i \mid Z_i^* = Z_j, \mathbf{X}_i), \tag{13}$$

where $\xi_{ij}^* = \xi_{ij} / \sum_{j'=1}^{N_{\mathcal{B}}} \xi_{ij'}$. Online SI S7 shows that under the two assumptions described earlier and mild regularity conditions, the weighted ML estimator given in equation (13) is consistent and asymptotically normal. Note that because we are considering large data sets, we ignore the uncertainty about ξ_{ij}^*.

SIMULATION STUDIES

We conduct a comprehensive set of simulation studies to evaluate the statistical accuracy and computational efficiency of our probabilistic modeling approach and compare them with those of deterministic methods. Specifically, we assess the ability of the proposed methodology to control estimation error, false positives and false negatives, and its robustness to missing values and noise in the linkage fields, as well as the degree of overlap between two data sets to be merged. We do so by systematically varying the amount and structure of missing data and measurement error.

The Setup

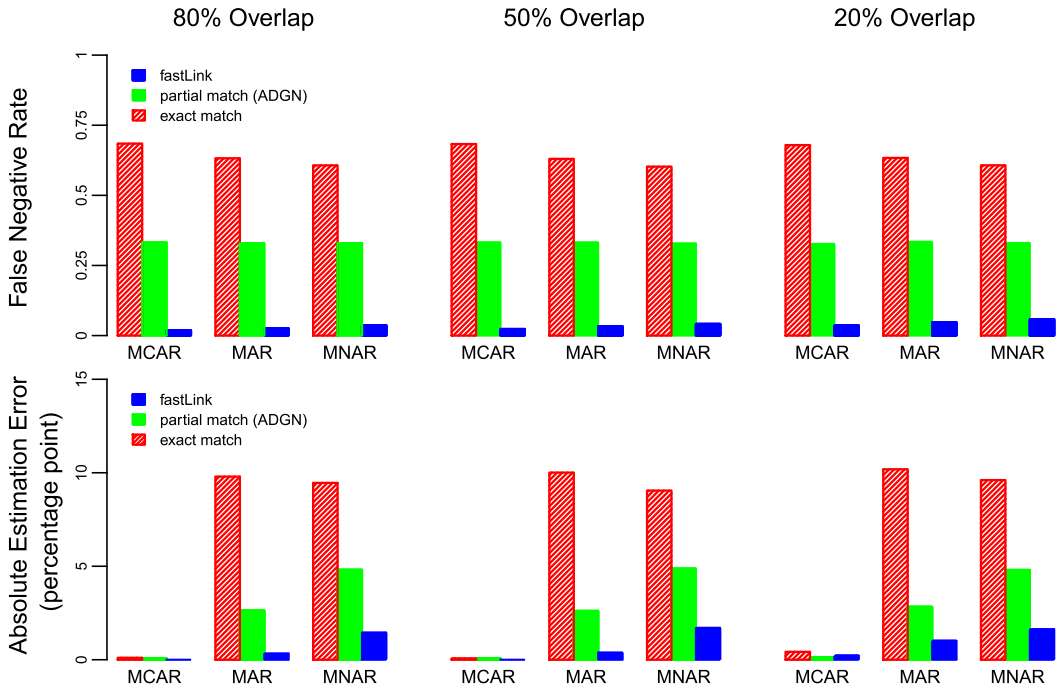
To make our simulation studies realistic, we use a data set taken from the 2006 California voter file. Because merging voter files is often done by blocking on gender, we subset the data set to extract the information about female voters only, reducing the number of observation to approximately 17 million voters to 8.3 million observations. To create a base data set for simulations, we further subset the data set by removing all observations that have at least one missing value in the following variables: first name, middle initial, last name, date of birth, registration date, address, zip code, and turnout in the 2004 Presidential election. After listwise deletion, we obtain the final data set of 341,160 voters, from which we generate two data sets of various characteristics to be merged. From this data set, we independently and randomly select two subsamples to be merged under a variety of scenarios.

We design our simulation studies by varying the values of the five parameters as summarized below. Online SI S9.1 describes in detail the precise setups of these simulations.

1. **Degree of overlap**: Proportion of records in the smaller data set that are also in the larger data set. We consider three scenarios—20% (small), 50% (medium), and 80% (large).

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

FIGURE 1. Accuracy of Data Merge



The top and bottom panels present the false negative rate (FNR) and the absolute estimation error (for estimating the turnout rate), respectively, when merging datasets of 100,000 records each across with different levels of overlap (measured as a percentage of a data set). Three missing data mechanisms are studied with the missing data proportion of 10% for each linkage field other than year of birth: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Classical measurement error is introduced to several linkage fields. The proposed probabilistic methodology (`fastLink`; blue solid bars) significantly outperforms the two deterministic algorithms, i.e., exact match (red shaded bars) and partial match (ADGN; light green solid bars), across simulation settings.

2. **Size balance**: Balance of sample sizes between the two data sets to be merged. We consider three ratios—1:1 (equally sized), 1:10 (imbalanced), and 1:100 (lopsided).
3. **Missing data**: We consider five different mechanisms, missing completely at random (MCAR), MAR, and not missing at random (NMAR). For MAR and NMAR, we consider independent and dependent missingness patterns across linkage fields.
4. **Amount of missing data**: Proportion of missing values in each linkage variable other than year of birth. We consider three scenarios—5% (small), 10% (medium), and 15% (large).
5. **Measurement error**: Proportion of records (6%) for which the first name, last name, and street name contain a classical measurement error.

Together, we conduct a total of 135 ($=3^3 \times 5$) simulation studies where missing data are of main concern. We also conduct another set of 135 simulations with various types of nonclassical measurement errors, while keeping the amount of missing values fixed (see Online SI S9.2).

Results

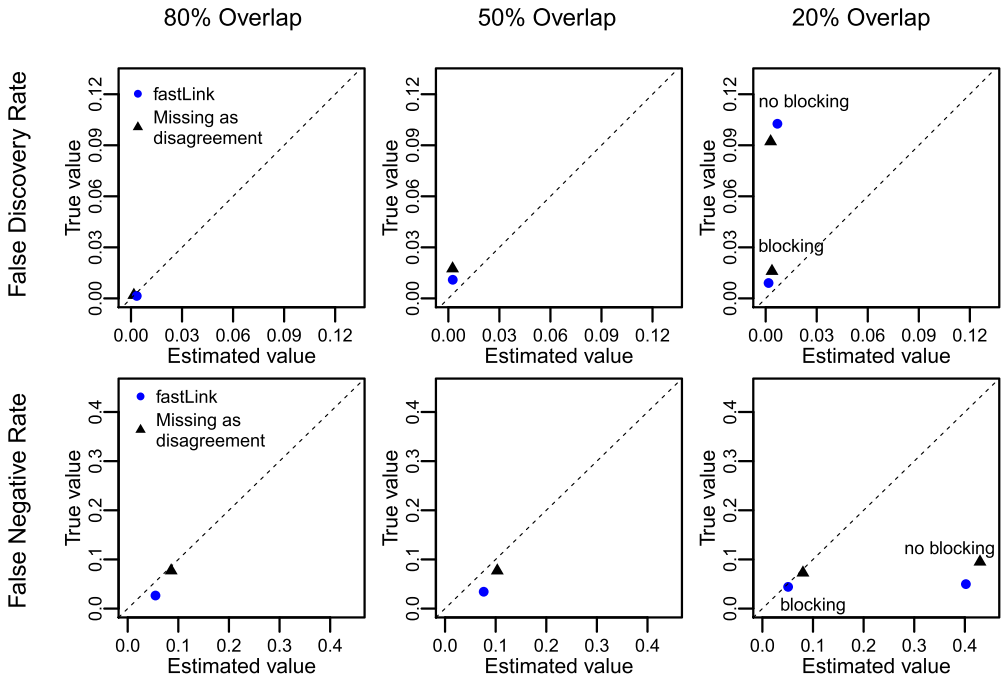
Figure 1 compares the performance of `fastLink` (blue solid bars) to the two deterministic methods often used by social scientists. The first is the merging method based on exact matches (red shaded bars), whereas the second is the recently proposed partial match algorithm (ADGN; light green solid bars) that considers two records as a match if at least three fields of their address, date of birth, gender, and name are identical (Ansolabehere and Hersh 2017). The top panel of Figure 1 presents the FNR whereas the bottom panel presents the absolute error for estimating the 2004 turnout rate. We merge two data sets of equal size (100,000 records each) after introducing the classical measurement error and the medium amount of missing data as explained above. For `fastLink`, only pairs with a match probability ≥ 0.85 are considered to be matches, but the results remain qualitatively similar if we change the threshold to 0.75 or 0.95.

We find that `fastLink` significantly outperforms the two deterministic methods.[10] Although all three methods are designed to control the FDR, only `fastLink` is able to keep the FNR low (less than five percentage in all cases considered here). The deterministic algorithms are not robust to missing data and measurement error, yielding a FNR of much greater magnitude. Additionally, we observe that the deterministic methods yield a substantially greater estimation

[10] In Online SI S8, we compare `fastLink` to the state-of-the-art probabilistic methods and find that `fastLink` performs as well as these methods.

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

FIGURE 2. Accuracy of FDR and FNR Estimates



The top panel compares the estimated FDR (*x*-axis) with its true value (*y*-axis) whereas the bottom panel compares the estimated FNR against its true value. We consider the medium amount of missing data generated under MAR as a missingness mechanism and add measurement error to some linkage fields. The blue solid circles represent the estimates based on `fastLink` whereas the black solid triangles represent the estimates obtained by treating missing data as disagreements. The FDR and FNR estimates are accurate when the overlap is high. Additionally, `fastLink` gives lower FDR and FNR than the same algorithm that treats missing values as a disagreement. Note that in cases where the overlap is small (20%), blocking improves the precision of our estimates.

bias than `fastLink` unless the data are MCAR. Under the other two missing data mechanisms, the magnitude of the bias is substantially greater than that of `fastLink`. Although `fastLink` has an absolute estimation error of less than 1.5 percentage points even under MNAR, the other two methods have an absolute estimation error of more than 7.5 percentage points under both MAR and MNAR. Finally, the performance of `fastLink` worsens as the size of overlap reduces and the missing data mechanism becomes less random.

We next evaluate the accuracy of FDR and FNR estimates in the top and bottom panels, respectively. Because the deterministic methods do not give such error estimates, we compare the performance of the proposed methodology (indicated by blue solid circles) with that of the same probabilistic modeling approach, which treats missing values as disagreements following a common practice in the literature (indicated by solid triangles). Figure 2 presents the results of merging two data sets of equal size where the medium amount of data are assumed to be MAR and some noise are added as described earlier. In the top panel of the figure, we find that the true FDR is low and its estimate is accurate unless the degree of overlap is small. With a small degree of overlap, both methods significantly underestimate the FDR. A similar finding is obtained for the FNR in the bottom panel of the figure where estimated FNR is biased upward.

One way to address the problem of having small overlap would be to use blocking based on a set of fully observed covariates. For example, in our simulations, because the year of birth is observed for each record in both data sets, we block the data by making comparisons only across individuals within a window of ± 1 year around each birth year.[11] Then, we apply `fastLink` to each block separately. As shown in the right most column of Figure 2, blocking significantly improves the estimation accuracy for the FDR and FNR estimates as well as their true values although the bias is not eliminated. The reason for this improvement is that traditional blocking increases the degree of overlap. For example, in this simulation setting for each of the 94 blocks under consideration, the ratio of true matches to all possible pairs is at least 8×10^{-5}, which is more than 15 times as large as the corresponding ratio for no blocking and is comparable to the case of overlap of 50%.

We present the results of the remaining simulation studies in the Online Simulation Appendix. Two major

[11] In Online SI S12, we also present results using a clustering method, i.e., *k*-means, to group similar observations.

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

patterns discussed above are also found under these other simulation scenarios. First, regardless of the missing data mechanisms and the amount of missing observations, `fastLink` controls FDR, FNR, and estimation error well. Second, a greater degree of overlap between data sets leads to better merging results in terms of FDR and FNR as well as the accuracy of their estimates. Blocking can ameliorate these problems caused by small overlap to some extent. These empirical patterns are consistently found across simulations even when two data sets have unequal sizes.

Computational Efficiency

We compare the computational performance of `fastLink` with that of the RecordLinkage package in R (Sariyar and Borg 2016) and the Record Linkage package in Python (de Bruin 2017) in terms of running time. The latter two are the only other open source packages in R and Python that implement a probabilistic model of record linkage under the Fellegi–Sunter framework. To mimic a standard computing environment of applied researchers, all the calculations are performed in a Macintosh laptop computer with a 2.8 GHz Intel Core i7 processor and 8 GB of RAM.

Although `fastLink` takes advantage of a multicore machine via the OpenMP-based parallelization (the other two packages do not have a parallelization feature), we perform the comparison on a single-core computing environment so that we can assess the computational efficiency of our algorithm itself. Additionally, we include runtime results where we parallelize computation across eight cores. For all implementations, we set the convergence threshold to 1×10^{-5}.[12]

We consider the setup in which we merge two data sets of equal size with 50% overlap, 10% missing proportion under MCAR, and no measurement error. Our linkage variables are first name, middle initial, last name, house number, street name, and year of birth. We vary the size of each data set from 1,000 records to 300,000 observations. As in the earlier simulations, each data set is based on the sample of 341,160 female registered voters in California, for whom we have complete information in each linkage field. To build the agreement patterns, we use the Jaro–Winkler string distance with a cutoff of 0.94 for first name, last name, and street name. For the remaining fields, we only consider exact matches as agreements.

Figure 3 presents the results of this running time comparison. We find that although all three packages take a similar amount of time for data sets of 1,000 records, the running time increases exponentially for the other packages in contrast to `fastLink` (black solid triangles connected by a dashed line, single core; blue solid circles connected by a solid line, eight cores), which exhibits a near linear increase. When matching data sets of 150,000 records each, `fastLink` takes less than six hours to merge using a single core (under three hours when parallelized across eight cores). In contrast, it takes more than 24 hours for Record Linkage (Python; solid purple squares connected by a dotted line), to merge two data sets of only 20,000 observations each. The performance is not as bad for Record Linkage (R; red crosses connected by a dashed line), but it still takes over six hours to merge data sets of 40,000 records each. Moreover, an approximation based on an exponential regression model suggests that Record Linkage (R) would take around 22 hours to merge two data sets of 50,000 records each, while Record Linkage (Python) would take about 900 days to accomplish the same merge. In Online SI S3.1, we further decompose the runtime comparison to provide more detail on the sources of our computational improvements. We detail the choices we make in the computational implementation that yields these substantial efficiency gains in Appendix A.

[12] Starting values differ across methods because other methods do not allow us to change their default starting values. However, the EM algorithm converges quickly regardless of the choice of starting values. In fact, it is well known that the bottleneck is a large number of required comparisons (e.g., Christen 2012; Jaro 1972), for which we use a hashing technique as described below in Appendix A.

EMPIRICAL APPLICATIONS

In this section, we present two empirical applications of the proposed methodology. First, we merge election survey data (about 55,000 observations) with political contribution data (about five million observations). The major challenge of this merge is the fact that the expected number of matches between the two data sets is small. Therefore, we utilize blocking and conduct the data merge within each block. The second application is to merge two nationwide voter files, each of which has more than 160 million records. This may, therefore, represent the largest data merge ever conducted in the social sciences. We show how to use auxiliary information about within-state and across-state migration rates to inform the match.

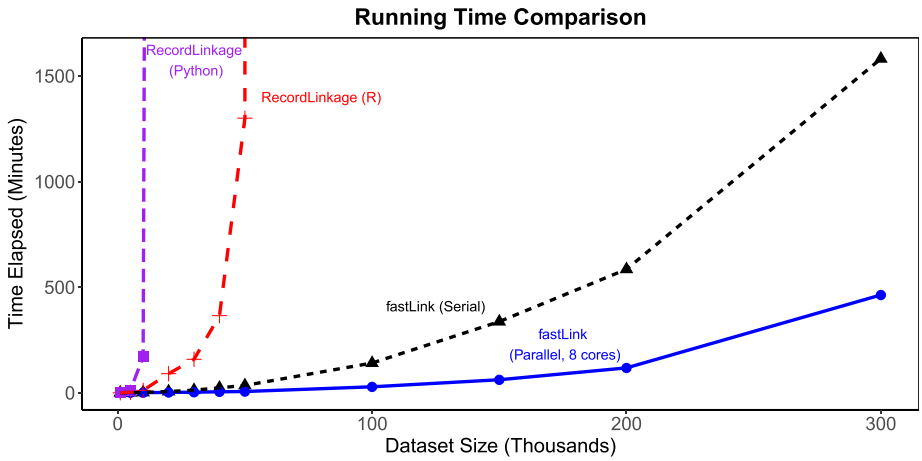
Merging Election Survey Data with Political Contribution Data

Hill and Huber (2017) study differences between donors and nondonors by merging the 2012 CCES survey with the Database on Ideology, Money in Politics, and Elections [DIME, Bonica (2013)]. The 2012 CCES is based on a nationally representative sample of 54,535 individuals recruited from the voting-age population in the United States. The DIME data, on the other hand, provide the information about individual donations to political campaigns. For the 2010 and 2012 elections, the DIME contains over five million donors.

The original authors asked YouGov, the company which conducted the survey, to merge the two data sets using a proprietary algorithm. This yielded a total of 4,432 CCES respondents matched to a donor in the DIME data. After the merge, Hill and Huber (2017) treat each matched CCES respondent as a donor and conduct various analyses by comparing these matched respondents with those who are not matched with a donor in the DIME data and hence are treated as nondonors. Below, we apply the proposed methodology to merge these two data sets and conduct a post-merge

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

FIGURE 3. Running Time Comparison



The plot presents the results of merging datasets of equal size using different implementations of the Fellegi-Sunter model. The datasets were constructed from a sample of female registered voters in California. The amount of overlap between datasets is 50%, and, for each dataset, there are 10% missing observations in each linkage variable: first name, middle initial, last name, house number, street name, and year of birth. The missing data mechanism is Missing Completely at Random (MCAR). The computation is performed on a Macintosh laptop computer with a 2.8 GHz Intel Core i7 processor and 8 GB of RAM. The proposed implementation `fastLink` (single-core runtime as black solid triangles connected by a dashed line, and parallelized over eight cores as blue solid dots connected by a solid line) is significantly faster than the other open-source packages.

analysis by incorporating the uncertainty about the merge process.

Merge Procedure

We use the name, address, and gender information to merge the two data sets. To protect the anonymity of CCES respondents, YouGov used `fastLink` to merge the data sets on our behalf. Moreover, because of contractual obligations, the merge was conducted only for 51,184 YouGov panelists, which is a subset of the 2012 CCES respondents. We block based on gender and state of residence, resulting in 102 blocks (50 states plus Washington DC $\times$ two gender categories). The size of each block ranges from 175,861 (CCES = 49, DIME = 3,589) to 790,372,071 pairs (CCES = 2,367, DIME = 333,913) with the median value of 14,048,151 pairs (CCES = 377, DIME = 37,263). Within each block, we merge the data sets using the first name, middle initial, last name, house number, street name, and postal code. As done in the simulations, we use three levels of agreement for the string-valued variables based on the Jaro–Winkler distance with 0.85 and 0.92 as the thresholds. For the remaining variables (i.e., middle initial, house number, and postal code), we utilize a binary comparison indicating whether they have an identical value.

To construct our set of matched pairs between CCES and DIME, first, we use the one-to-one matching assignment algorithm described in Online SI S5 and find the best match in the DIME data for each CCES respondent. Then, we declare as a match any pair whose matching probability is above a certain threshold. We use three thresholds, i.e., 0.75, 0.85, and 0.95, and examine the sensitivity of the empirical results to the choice of threshold value.[13] Finally, in the original study of Hill and Huber (2017), noise is added to the amount of contribution to protect the anonymity of matched CCES respondents. However, we signed a non-disclosure agreement with YouGov for our analysis so that we can make a precise comparison between the proposed methodology and the proprietary merge method used by YouGov.

Merge Results

Table 2 presents the merge results. We begin by assessing the match rates, which represent the proportion of CCES respondents who are matched with donors in the DIME data. Although the match rates are similar between the two methods, `fastLink` appears to find slightly more (less) matches for male (female) respondents than the proprietary method regardless of the threshold used. However, this does not mean that both methods find identical matches. In fact, out of 4,797 matches identified by `fastLink` (using the threshold of 0.85), the proprietary method does not identify 861 or 18% of them as matches.

As discussed in the subsection *The Canonical Model of Probabilistic Record Linkage*, one important advantage of the probabilistic modeling approach is that we can estimate the FDR and FNR, which are shown in the table. Such error rates are not available for the

[13] In Online SI S10.3, instead of a one-to-one matching restriction used here, we present the results of the weighted approach described in the subsection *The Uncertainty of the Merge Process*. As shown in Figure S7 of Online SI S10.3, there is no distinguishable difference in the results obtained from either approach.

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

TABLE 2. The Results of Merging the 2012 Cooperative Congressional Election Study (CCES) with the 2010 and 2012 Database on Ideology, Money in Politics, and Elections (DIME) Data

		fastLink			Proprietary method
		0.75	0.85	0.95	
Number of matches	All	4,948	4,797	4,576	4,534
	Female	2,198	2,156	2,067	2,210
	Male	2,750	2,641	2,524	2,324
Overlap between fastLink and proprietary method	All	3,959	3,936	3,881	
	Female	1,877	1,866	1,844	
	Male	2,082	2,070	2,037	
Match rate (%)	All	9.67	9.37	8.94	8.85
	Female	8.12	7.96	7.63	8.16
	Male	11.40	10.95	10.40	9.64
FDR (%)	All	1.24	0.65	0.21	
	Female	0.92	0.53	0.14	
	Male	1.50	0.75	0.28	
FNR (%)	All	15.25	17.35	20.81	
	Female	5.35	6.80	10.30	
	Male	21.83	24.36	27.79	

The table presents the merging results for both fastLink and the proprietary method used by YouGov. The results of fastLink are presented for one-to-one match with three different thresholds (i.e., 0.75, 0.85, 0.95) for the matching probability to declare a pair of observations as a successful match. The number of matches, the amount of overlap, and the overall match rates are similar between the two methods. The table also presents information on the estimated false discovery and false negative rates (FDR and FNR, respectively) obtained using fastLink. These statistics are not available for the proprietary method.

proprietary method. As expected, the overall estimated FDR is controlled to less than 1.5% for both male and female respondents. The estimated FNR, on the other hand, is large, illustrating the difficulty of finding some donors. In particular, we find that female donors are much more difficult to find than male donors.

Specifically, there are 12,803 CCES respondents who said they made a campaign contribution during the last 12 months before the 2012 election. Among them, 5,206 respondents claimed to have donated at least 200 dollars. Interestingly, both fastLink and the proprietary method matched an essentially identical number of self-reported donors with a contribution of over 200 dollars (2,431 and 2,434 or approximately 47%, respectively), whereas among the self-reported small donors both methods can only match approximately 16% of them.

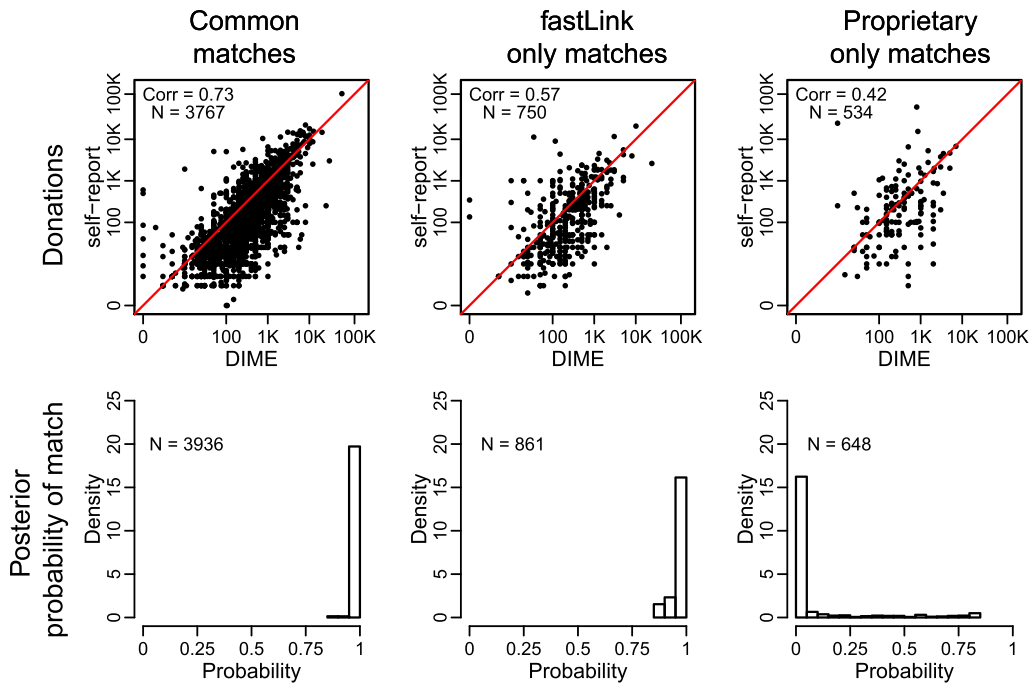
Next, we examine the quality of matches for the two methods (see also Online SI S13). We begin by comparing the self-reported donation amount of matched CCES respondents with their actual donation amount recorded in the DIME data. Although only donations greater than 200 dollars are recorded at the federal level, the DIME data include some donations of smaller amounts, if not all, at the state level. Thus, although we do not expect a perfect correlation between self-reported and actual donation amount, under the assumption that donors do not systematically under or over report the amount of campaign contributions, a high correlation between the two measures implies a more accurate merging process.

The upper panel of Figure 4 presents the results where for fastLink, we use one-to-one match with the threshold of 0.85.[14] We find that for the respondents who are matched by both methods, the correlation between the self-reported and matched donation amounts is reasonably high (0.73). In the case of respondents who are matched by fastLink only, we observe that the correlation is low (0.57) but is greater than the correlation for those matches identified by the proprietary method alone (0.42). We also examine the distribution of match probabilities for these three groups of matches. The bottom panel of the figure presents the results, which are consistent with the patterns of correlation identified in the top panel. That is, those matches identified by the two methods have the highest match probability whereas most of the matches identified only by the proprietary method have extremely low match probabilities. In Online SI S13, we also examine the quality of the agreement patterns separately for the matches identified by both methods, fastLink only, and the proprietary method only. Overall, our results indicate that fastLink produces matches whose quality is often better than that based on the proprietary method.

[14] Figures S5 and S6 in Online SI S10 present the results under two different thresholds: 0.75 and 0.95, respectively. The results under those thresholds are similar to those with the threshold of 0.85 presented here.

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

FIGURE 4. Comparison of `fastLink` and the Proprietary Method



The top panel compares the self-reported donations (*y*-axis) by matched CCES respondents with their donation amount recorded in the DIME data (*x*-axis) for the three different groups of observations: those declared as matches by both `fastLink` and the proprietary method (left), those identified by `fastLink` only (middle), and those matched by the proprietary method only (right). The bottom panel presents histograms for the match probability for each group. For `fastLink`, we use one-to-one match with the threshold of 0.85.

Post-Merge Analysis

An important advantage of the probabilistic modeling approach is its ability to account for the uncertainty of the merge process in post-merge analyses. We illustrate this feature by revisiting the post-merge analysis of Hill and Huber (2017). The original authors are interested in the comparison of donors (defined as those who are matched with records in the DIME data) and nondonors (defined as those who are not matched) among CCES respondents. Using the matches identified by a proprietary method, Hill and Huber (2017) regress policy ideology on the matching indicator variable, which is interpreted as a donation indicator variable, the turnout indicator variables for the 2012 general election and 2012 congressional primary elections, as well as several demographic variables. Policy ideology, which ranges from −1 (most liberal) to 1 (most conservative), is constructed by applying a factor analysis to a series of questions on various issues.[15] The demographic control variables include income, education, gender, household union membership, race, age in decades, and importance of religion. The same model is fitted separately for Democrats and Republicans.

To account for the uncertainty of the merge process, as explained in the subsection *Post-Merge Analysis*, we fit the same linear regression except that we use the mean of the match indicator variable as the main explanatory variable rather than the match indicator variable. Table 3 presents the estimated coefficients of the aforementioned linear regression models with the corresponding heteroskedasticity-robust standard errors in parentheses. Generally, the results of our improved analysis agree with those of the original analysis, showing that donors tend to be more ideologically extreme than nondonors.

Although the overall conclusion is similar, the estimated coefficients are smaller in magnitude when accounting for the uncertainty of merge process. In particular, according to `fastLink`, for Republican respondents, the estimated coefficient of being a donor represents only 12% of the standard deviation of their ideological positions (instead of 21% given by the proprietary method). Indeed, the difference in the estimated coefficients between `fastLink` and the proprietary method is statistically significant for both Republicans (0.035, s.e. = 0.014), and Democrats (−0.015, s.e. = 0.007). Moreover, although the original analysis find that the partisan mean ideological difference for donors (1.108, s.e. = 0.018) is 31 percentage larger than that for nondonors (0.848, s.e. = 0.001), the results based on `fastLink` shows that this difference is only 25 percentage larger for donors (1.058,

[15] They include gun control, climate change, immigration, abortion, jobs versus the environment, same-sex marriage, affirmative action, and fiscal policy.

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

TABLE 3. Predicting Policy Ideology Using Contributor Status

	Republicans		Democrats	
	Original	`fastLink`	Original	`fastLink`
Contributor	0.080***	0.046***	−0.180***	−0.165***
	(0.016)	(0.015)	(0.008)	(0.009)
Turnout for 2012 general election	0.095***	0.095***	−0.060***	−0.060***
	(0.013)	(0.013)	(0.010)	(0.010)
Turnout for 2012 primary election	0.094***	0.095***	−0.019**	−0.022***
	(0.009)	(0.009)	(0.009)	0.009)
Demographic controls	Yes	Yes	Yes	Yes
Number of observations	17,386	17,386	20,925	20,925

The estimated coefficients from the linear regression of policy ideology score on the contributor indicator variable and a set of demographic controls. Along with the original analysis, the table presents the results of the improved analysis based on `fastLink`, which accounts for the uncertainty of the merge process. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Robust standard errors in parentheses.

s.e. = 0.018). Thus, although the proprietary method suggests that the partisan gap for donors is similar to the partisan gap for those with a college degree or higher (1.100, s.e. = 0.036), `fastLink` shows that it is closer to the partisan gap for those with just some college education but without a degree (1.036, s.e. = 0.035).

Merging Two Nationwide Voter Files Over Time

Our second application is what might be the largest data merging exercise ever conducted in social sciences. Specifically, we merge the 2014 nationwide voter file to the 2015 nationwide voter file, each of which has over 160 million records. The data sets are provided by L2, Inc., a leading national non-partisan firm and the oldest organization in the United States that supplies voter data and related technology to candidates, political parties, pollsters, and consultants for use in campaigns. In addition to the sheer size of the data sets, merging these nationwide voter files is methodologically challenging because some voters change their residence over time, making addresses uninformative for matching these voters.

Merge Procedure

When merging data sets of this scale, we must drastically reduce the number of comparisons. In fact, if we examine all possible pairwise comparisons between the two voter files, the total number of such pairs exceeds 2.5×10^{16}. It is also important to incorporate auxiliary information about movers because the address variable is noninformative when matching these voters. We use the IRS Statistics of Income (SOI) to calibrate match rates for within-state and across-state movers. Details on incorporating migration rates into parameter estimation can be found in the subsection *Incorporating Auxiliary Information* and Online SI S6.2. The IRS SOI data are definitive source of migration data in the United States that tracks individual residences year-to-year across all states through their tax returns.

We develop the following two-step procedure that utilizes random sampling and blocking of voter records to reduce the computational burden of the merge (see Online SI S3.2 and S6.2). Our merge is based on first name, middle initial, last name, house number, street name, date/year/month of birth, date/year/month of registration, and gender. The first step uses each of these fields to inform the merge, whereas the second step uses only first name, middle initial, last name, date/year/month of birth, and gender. For both first name and last name, we include a partial match category based on the Jaro–Winkler string distance calculation, setting the cutoff for a full match at 0.92 and for a partial match at 0.88.

As described in Online SI S6.2, we incorporate auxiliary information into the model by moving from the likelihood framework to a fully Bayesian approach. Because of conjugacy of our priors, we can obtain the estimated parameters by maximizing the log posterior distribution via the EM algorithm. This approach allows us to maintain the computational efficiency.[16]

Step 1: Matching within-state movers and nonmovers for each state.

(a) Obtain a random sample of voter records from each state file.
(b) Fit the model to this sample using the within-state migration rates from the IRS data to specify prior parameters.
(c) Create blocks by first stratifying on gender and then applying the *k*-means algorithm to the first name.
(d) Using the estimated model parameters, conduct the data merge within each block.

Step 2: Matching across-state movers for each pair of states.

(a) Set aside voters who are identified as successful matches in Step 1.
(b) Obtain a random sample of voter records from each state file as done in Step 1(a).

[16] Specifically, we set prior parameters on the expected match rate and expected within-state movers rate using the IRS data, giving 75% weight to the prior estimate and 25% weight to the maximum likelihood estimate. For the first step, we set priors on both $\pi_{\text{address},1,0}$ (the probability of a voter's address not matching conditional on being in the matched set, which is equivalent to the share of in-state movers in the matched set) and λ. For the second step, we set a prior on λ.

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

TABLE 4. The Results of Merging the 2014 Nationwide Voter File with the 2015 Nationwide Voter File

		fastLink			
		0.75	0.85	0.95	Exact
Match count (millions)	All	135.60	129.69	128.73	91.62
	Within-state	127.38	127.12	126.80	91.36
	Across-state	8.22	2.57	1.93	0.27
Match rate (%)	All	97.25	93.67	93.04	66.24
	Within-state	92.06	91.87	91.66	66.05
	Across-state	5.19	1.80	1.38	0.19
FDR (%)	All	1.02	0.10	0.03	
	Within-state	0.08	0.04	0.01	
	Across-state	0.95	0.06	0.02	
FNR (%)	All	3.35	3.63	3.86	
	Within-state	2.63	2.83	3.05	
	Across-state	0.72	0.80	0.81	

This table presents the merging results for `fastLink` for three different thresholds (i.e., 0.75, 0.85, 0.95) for the matching probability to declare a pair of observations a successful match. Across the different thresholds, the match rates do not change substantially and are significantly greater than the corresponding match rates of the exact matching technique.

(c) Fit the model using the across-state migration rates from the IRS data to specify prior parameters.
(d) Create blocks by first stratifying on gender and then applying the *k*-means algorithm to the first name as done in Step 1(c).
(e) Using the estimated model parameters, conduct the data merge within each block as done in Step 1(e).

In Step 1, we apply random sampling, rather than blocking, strategy to use the within-state migration rates from the IRS data and fit the model to a representative sample for each state. For the same reason, we use a random sampling strategy in Step 2 to exploit the availability of IRS across-state migration rates. We obtain a random sample of 800,000 voter records for files with over 800,000 voters and use the entire state file for states with fewer than 800,000 voter records on file. Online SI S11 shows through simulation studies that for data sets as small as 100,000 records, a 5% random sample leads to parameter estimates nearly indistinguishable from those obtained using the full data set. Based on this finding, we choose 800,000 records as the size of the random samples, corresponding to a 5% of records from California, the largest state in the United States.

Second, within each step, we conduct the merge by creating blocks to reduce the number of pairs for consideration. We block based on gender, first name, and state, and we select the number of blocks so that the average size of each blocked data set is approximately 250,000 records. To block by first name, we rank ordered the first names alphabetically and ran the *k*-means algorithm on this ranking in order to create clusters of maximally similar names.[17] Finally, the entire merge procedure is computationally intensive. The reason is that we need to repeat Step 1 for each of 50 states plus Washington DC and apply Step 2 to each of 1,275 pairs. Thus, as explained in Online SI S2, we use parallelization whenever possible. All merges were run on a Linux cluster with 16 2.4-GHz Broadwell 28-core nodes with 128 GB of RAM per node.

[17] See Online SI S14 for evidence that this blocking strategy performs similarly to a blocking strategy based on age windowing.

Merge Results

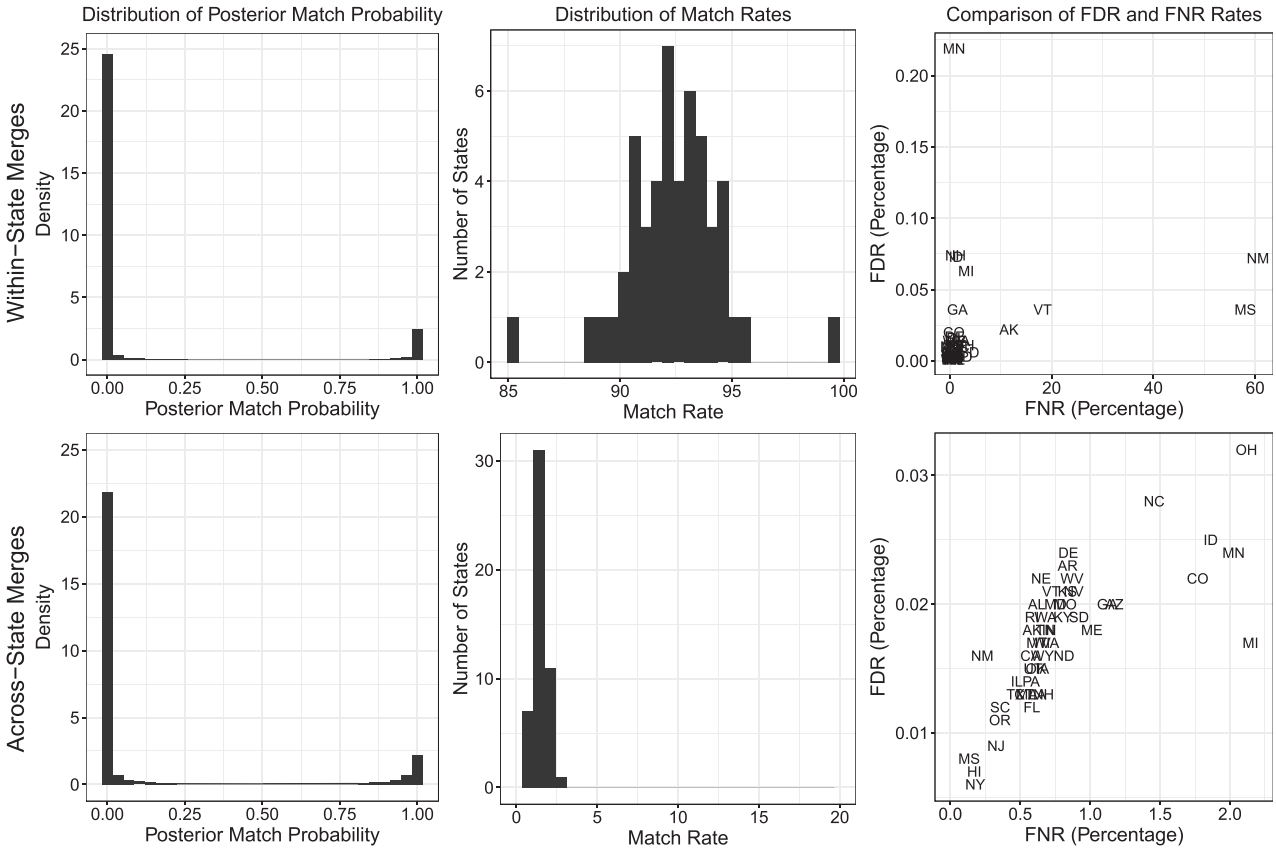
Table 4 presents the overall match rate, FDR, and FNR obtained from `fastLink`. We assess the performance of the match at three separate matching probability thresholds to declare a pair of observations a successful match: 0.75, 0.85, and 0.95. We also break out the matches by within-state matches only and across-state matches only. Across the three thresholds, the overall match rate remains very high, at 93.04% under a 95% acceptance threshold, although the estimated FDR and FNR remain controlled at 0.03% and 3.86%. All three thresholds yield match rates that are significant higher than the corresponding match rates of the exact matching technique.

In Figure 5, we examine the quality of the merge separately for the within-state merge (top panel) and across-state merge (bottom panel). The first column plots the distribution of the matching probability across all potential match pairs. For both within-state and across-state merge, we observe a clear separation between the successful matches and unsuccessful matches, with very few matches falling in the middle. This suggests that the true and false matches are identified reasonably well. In the second column, we examine the distribution of the match rate by state. Here, we see that most states are tightly clustered between 88% and 96%. Only Ohio, with a match rate of 85%, has a lower match rate. For the across-state merge, the match rate is clustered tightly between 0% and 5%.

In the third column, we plot the estimated FDR against the estimated FNR for each state. For the within-state merge, the FDR is controlled well—every state other than Minnesota has an FDR below 0.1%. Additionally, there are only two states, Mississippi and

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

FIGURE 5. Graphical Diagnostics From Merging the 2014 Nationwide Voter File with the 2015 Nationwide Voter File



This figure presents graphical diagnostics for `fastLink` for within-state matches (top panel) and across-state matches (bottom panel). The first column plots the distribution of the matching probability across all patterns. The second column plots the distribution of the match rate for each state. Lastly, the third column compares the FNR against the FDR for each state separately.

New Mexico, where `fastLink` seems to have trouble identifying true matches, as measured by the FNR. In the across-state merge, the FDR for every state is below 0.1%, suggesting that the resulting matches are of high quality. Furthermore, `fastLink` appears to be finding a high share of true movers across voter files, as the FNR for all but three states falls under 2%.

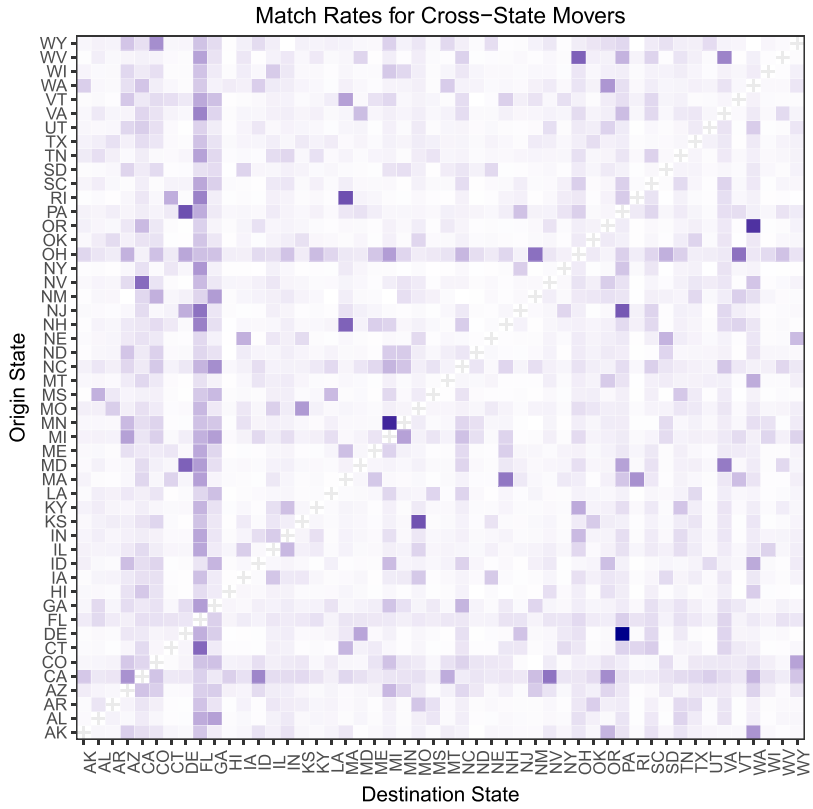
Finally, we examine the across-state migration patterns recovered from our matching procedure. Figure 6 displays a heatmap of the migration patterns obtained from `fastLink` with darker purple colors indicating a higher match rate when merging the 2014 nationwide voter file for a given state (origin state) to the 2015 nationwide voter file for a given state (destination state). We uncover several regional migration patterns. First, we find a migration cluster in New England, where voters from New Hampshire and Rhode Island migrated to Massachusetts between 2014 and 2015. Another strong migration cluster exists between New Jersey, Delaware, and Pennsylvania in the mid-Atlantic region. Both patterns suggest that most migration occurs between clusters of adjacent states and urban centers. Finally, we find a large volume of out migration to Florida from across the United States, and the out migration is particularly concentrated in states on the Eastern seaboard such as Virginia, New Hampshire, New Jersey, and Connecticut. This possibly reflects the flow of older voters and retirees to the more temperate climate.

CONCLUDING REMARKS

With the advance of the Internet, the last two decades have witnessed a "data revolution" in the social sciences where diverse and large data sets have become electronically available to researchers. Much of today's cutting-edge quantitative social science research results from researchers' creativity to link multiple data sets that are collected separately. In many cases, however, a unique identifier that can be used to merge multiple data sources does not exist. Currently, most social scientists rely on either deterministic or proprietary methods. Yet, deterministic methods are not robust to measurement errors and missing data, cannot quantify the uncertainty inherent in merge process, and often require arbitrary decisions from researchers. Proprietary methods, many of which are also deterministic, lack transparency and hence are not suitable for academic and

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

FIGURE 6. Across-State Match Rates for the 2014 Nationwide Voter File to 2015 Nationwide Voter File Merge



We plot the match rates from each across-state match pair as a heatmap, where darker colors indicate a higher match rate.

policy research where reproducibility and transparency play an essential role.

Here, we advocate the use of probabilistic modeling to assist merging large-scale data sets. The main advantage of probabilistic models is their ability to estimate false positive and false negative rates that arise when linking multiple data sets. We contribute to the statistical literature of record linkage by developing a fast and scalable implementation of the canonical model. Through simulation and empirical studies, we demonstrate that the proposed methodology can quickly and reliably merge data sets even when they have millions of records.

Like any methods, however, the proposed record linkage technology has important limitations of which researchers must be aware. Most importantly, the proposed methodology is likely to have a difficult time producing high-quality matches when the overlap between two data sets is expected to be small. As shown in our simulation studies, for these difficult merge problems, effective blocking is essential. Blocking is even more important when linking many data sets at once. Other important research questions are how to merge more than two files at the same time and how to efficiently use a small amount of hand-coded data to improve the quality of record linkage. We leave these methodological challenges to future research.

Appendix A. COMPUTATIONALLY EFFICIENT IMPLEMENTATION

In this appendix, we describe the details of our computationally efficient implementation of the canonical model of probabilistic record linkage.

Reverse Data Structures for Field Comparisons

The critical step in record linkage is to compare pairs of records across the K fields used to link two data sets, which is often regarded as the most expensive step in terms of computational time (Christen 2012). To do so, for each linkage field k, we first compare observation i of data set $\mathcal{A}$ and j from data set $\mathcal{B}$ via a predefined distance metric (e.g., Jaro–Winkler for string-valued fields) and obtain a value which we call $S_k(i, j)$. However, comparisons in the Fellegi–Sunter model are represented in terms of a discrete agreement level per linkage field, not a continuous measure of agreement as the one implied by the distance metric. In other words, we need a discrete representation of $S_k(i, j)$.

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

Specifically, if we have a total of L_k agreement levels for the kth variable, then,

$$\gamma_k(i,j) = \begin{cases} 0 & \text{if } S_k(i,j) \leq \tau_0 \\ 1 & \text{if } \tau_0 < S_k(i,j) \leq \tau_1 \\ \vdots & \\ L_k - 1 & \text{if } \tau_{L_k-2} < S_k(i,j) \leq \tau_{L_k-1} \end{cases}, \quad \textbf{(14)}$$

where $\gamma_k(i, j)$ represents the agreement level between the values for variable k for the pair (i,j) and $\boldsymbol{\tau} = \{\tau_0, \tau_1, \ldots, \tau_{L_{k-1}}\}$ the set of predetermined thresholds use to define the agreement levels. For example, to compare names and last names, some authors such as Winkler (1990) argue in favor of using the Jaro–Winkler string distance to produce S_k, where one could use $\boldsymbol{\tau} = \{0.88, 0.94\}$ to construct γ_k for three agreement levels.

Still the problem with constructing γ_k is that the number of comparisons we have to make is often large. In our proposed implementation we exploit the following characteristics of typical record linkage problems in social sciences:

- The number of unique values observed in each linkage field is often less than the number of observations in each data set. For example, consider a variable such as first name. Naively, one may compare the first name of each observation in data set $\mathcal{A}$ with that of every observation in $\mathcal{B}$. In practice, however, we can reduce the number of comparisons by considering only unique first name that appears in each data set. The same trick can be used for all linkage fields by focusing on the comparison of the unique values of each variable.
- For each comparison between two unique first names ($name_{1,\mathcal{A}}$ and $name_{1,\mathcal{B}}$), for example, we only keep the indices of the original data sets and store them using what is often referred as a reverse data structure in the literature (Christen 2012). In such an arrangement, a pair of names ($name_{1,\mathcal{A}}$, $name_{1,\mathcal{B}}$) becomes a key with two lists, one containing the indices from data set $\mathcal{A}$ that have a first name equal to $name_{1,\mathcal{A}}$, and another list that does the same for $name_{1,\mathcal{B}}$ in data set $\mathcal{B}$.
- Comparisons involving a missing value need not be made. Instead, we only need to store the indices of the observations in $\mathcal{A}$ and $\mathcal{B}$ that contain missing information for field k.
- Because the agreement levels are mutually exclusive, we use the lowest agreement level as the base category. Once a set of threshold values has been defined, then a pair of names can only be categorized in one of the L_k agreement levels. The indices for the pairs of values that can be categorized as disagreements (or nearly disagreements) do not need to be stored. For most variables, disagreement is the category that encompasses the largest number of pairs. Thus, our reverse data structure lists become quite sparse. This sparsity can be exploited by the use of sparse matrix, yielding a substantially memory-efficient implementation.

Sparse Matrix Representation of Hash Tables to Count Agreement Patterns

Next, we describe our computationally efficient implementation of the Fellegi–Sunter model via the EM algorithm (see Online SI S2 for the exact algorithm we use). First, for implementing the E step, notice that the match probability given in equation (5) takes the same value for two pairs if their agreement patterns are identical. For the sake of illustration, consider a simple example where two variables are used for merging, i.e., $K = 2$, and binary comparison is made for each variable, i.e., $L_k = 2$. Under this setting, there are a total of nine agreement patterns: $(0,0), (0,1), (1,0), (1,1), (\text{NA}, 0), (\text{NA}, 1)$, (0, NA), (1, NA), and (NA, NA), where 1 and 0 represent agreement and disagreement, respectively, whereas NA represents a missing value. Then, for instance, the match probability for (0, 1) is given by $\lambda\pi_{110}\pi_{211}/\{\lambda\pi_{110}\pi_{211} + (1-\lambda)\pi_{100}\pi_{201}\}$ whereas that for (1, NA) is equal to $\lambda\pi_{111}/\{\lambda\pi_{111} + (1-\lambda)\pi_{101}\}$. If all comparison values are missing, e.g., (NA, NA), then we set the match probability to λ. Thus, the E step can be implemented by computing the match probability for each of the *realized* agreement patterns. Often, the total number of realized agreement patterns is much smaller than the number of all *possible* agreement patterns.

Second, the M step defined in equations (S1) and (S2) requires the summation of match probabilities across all pairs or their subset. Because this probability is identical within each agreement pattern, all we have to do is to count the total number of pairs that have each agreement pattern. In other words, the number of pairs per agreement pattern becomes our sufficient statistic. We use the following hash function for efficient counting:[18]

$$\mathbf{H} = \sum_{k=1}^{K} \mathbf{H}_k \quad \text{where}$$

$$\mathbf{H}_k = \begin{bmatrix} h_k^{(1,1)} & h_k^{(1,2)} & \ldots & h_k^{(1,N_{\mathcal{B}})} \\ \vdots & \vdots & \ddots & \vdots \\ h_k^{(N_{\mathcal{A}},1)} & h_k^{(N_{\mathcal{A}},2)} & \ldots & h_k^{(N_{\mathcal{A}},N_{\mathcal{B}})} \end{bmatrix}, \quad \textbf{(15)}$$

where $h_k^{(i,j)} = \mathbf{1}\{\gamma_k(i,j) > 0\} 2^{\gamma_k(i,j) + \mathbf{1}\{k>1\} \times \sum_{e=1}^{k-1}(L_e - 1)}$. The matrix $\mathbf{H}_k$ maps each pair of records to a corresponding agreement pattern in the kth variable that is represented by a unique hash value based on the powers of 2. These hash values are chosen such that the matrix $\mathbf{H}$ links each pair to the corresponding agreement pattern across K variables.

Because an overwhelming majority of pairs do not agree in any of the linkage fields, most elements of the $\mathbf{H}_k$ matrix are zero. As a result, the $\mathbf{H}$ matrix also has many zeros. In our implementation, we utilize sparse matrices the lookup time of which is $O(P)$ where P is the number of unique agreement patterns observed. In most applications, P is much less than the total number of possible agreement patterns, i.e., $\prod_{k=1}^{K} L_k$. This hashing technique is applicable if the number of variables used for merge is moderate. If many variables are used for the merge, approximate hashing techniques such as min hashing and locally sensitive hashing are necessary.

[18] Since the work of Jaro (1972), the use of table-like objects to store agreement patterns has been recognized as an important step to improve computational efficiency. Our contribution goes beyond by tying together, under a unified framework, reverse data structures and novel use of a sparse matrix representation of a hash table to store agreement patterns.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S0003055418000783.

Replication materials can be found on Dataverse at: https://doi.org/10.7910/DVN/YGUHTD.

REFERENCES

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of the Nazis in Prewar Germany." *Quarterly Journal of Economics* 130: 1885–939.

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate." *Political Analysis* 20: 437–59.

Ansolabehere, Stephen, and Eitan Hersh. 2017. "ADGN: An Algorithm for Record Linkage Using Address, Date of Birth, Gender and Name." *Statistics and Public Policy* 4: 1–10.

Belin, Thomas R., and Donald B. Rubin. 1995. "A Method for Calibrating False-Match Rates in Record Linkage." *Journal of the American Statistical Association* 90: 694–707.

Berent, Matthew K., Jon Arthur Krosnick, and A. Lupia. 2016. "Measuring Voter Registration and Turnout in Surveys. Do Official Government Records Yield More Accurate Assessments?" *Public Opinion Quarterly.* 80: 597–621.

Bolsen, Toby, Paul J. Ferraro, and Juan Jose Miranda. 2014. "Are Voters More Likely to Contribute to Other Public Goods? Evidence from a Large-Scale Randomized Policy Experiment." *American Journal of Political Science* 58: 17–30.

Bonica, Adam. 2013. *Database on Ideology, Money in Politics, and Elections: Public Version 1.0 [Computer File]*. Stanford, CA: Stanford University Libraries.

Cesarini, David, Erik Lindqvist, Robert Ostling, and Bjorn Wallace. 2016. "Wealth, Health, and Child Development: Evidence from Administrative Data on Swedish Lottery Players." *Quarterly Journal of Economics* 131: 687–738.

Christen, Peter. 2012. *Data Matching. Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection.* Heidelberg, Germany: Springer.

Cohen, William W., Pradeep Ravikumar, and Stephen Fienberg. 2003. "A Comparison of String Distance Metrics for Name-Matching Tasks." In *International Joint Conference on Artificial Intelligence (IJCAI) 18*.

Cross, Philip J., and Charles F. Manski. 2002. "Regressions, Short and Long." *Econometrica* 70: 357–68.

Dalzell, Nicole M., and Jerome P. Reiter. 2018. "Regression Modeling and File Matching Using Possibly Erroneous Matching Variables." *Journal of Computational and Graphical Statistics* 1–11. Published online July 11, 2018.

de Bruin, Jonathan. 2017. "Record Linkage. Python library. Version 0.8.1." https://recordlinkage.readthedocs.io/.

Einav, Liran, and Jonathan Levin. 2014. "Economics in the Age of Big Data." *Science* 346 (6210): 1243089-1–6.

Enamorado, Ted. 2018. "Active Learning for Probabilisitic Record Linkage." Social Science Research Network (SSRN). URL: https://ssrn.com/abstract=3257638.

Engbom, Niklas, and Christian Moser. 2017. "Returns to Education Through Access to Higher-Paying Firms: Evidence from US Matched Employer-Employee Data." *American Economic Review: Papers and Proceedings* 107: 374–78.

Feigenbaum, James. 2016. Automated Census Record Linking: A Machine Learning Approach. Boston University. Technical Report. https://jamesfeigenbaum.github.io/research/pdf/census-link-ml.pdf

Fellegi, Ivan P., and Alan B. Sunter. 1969. "A Theory of Record Linkage." *Journal of the American Statistical Association* 64: 1183–210.

Figlio, David, and Jonathan Guryan. 2014. "The Effects of Poor Neonatal Health on Children's Cognitive Development." *American Economic Review* 104: 3921–55.

Giraud-Carrier, Christophe, Jay Goodlife, Bradley M. Jones, and Stacy Cueva. 2015. "Effective Record Linkage for Mining Campaign Contribution Data." *Knowledge and Information Systems* 45: 389–416.

Goldstein, Harvey, and Katie Harron. 2015. *Methodological Developments in Data Linkage.* John Wiley & Sons, Ltd. Chapter 6: Record Linkage: A Missing Data Problem, pp. 109–24.

Gutman, Roee, Christopher C. Afendulis, and Alan M. Zaslavsky. 2013. "A Bayesian Procedure for File Linking to End-of-Life Medical Costs." *Journal of the American Medical Informatics Association.* 103: 34–47.

Gutman, Roee, Cara J. Sammartino, Traci C. Green, and Brian T. Montague. 2016. "Error Adjustments for File Linking Methods Using Encrypted Unique Client Identifier (eUCI) with Application to Recently Released Prisoners Who Are HIV+." *Statistics in Medicine* 35: 115–29.

Harron, Katie, Harvey Goldstein, and Chris Dibben, eds. 2015. *Methodological Developments in Data Linkage.* West Sussex: John Wiley & Sons.

Hersh, Eitan D. 2015. *Hacking the Electorate: How Campaigns Perceive Voters*. Cambridge, U.K.: Cambridge University Press.

Herzog, Thomas H., Fritz Scheuren, and William E. Winkler. 2010. "Record Linkage." *Wiley Interdisciplinary Reviews: Computational Statistics* 2: 535–43.

Herzog, Thomas N., Fritz J. Scheuren, and William E. Winkler. 2007. *Data Quality and Record Linkage Techniques*. New York: Springer.

Hill, Seth. 2017. "Changing Votes or Changing Voters: How Candidates and Election Context Swing Voters and Mobilize the Base." *Electoral Studies* 48: 131–48.

Hill, Seth J., and Gregory A. Huber. 2017. "Representativeness and Motivations of the Contemporary Donorate: Results from Merged Survey and Administrative Records." *Political Behavior* 39: 3–29.

Hof, Michel H. P., and Aeilko H. Zwinderman. 2012. "Methods for Analyzing Data from Probabilistic Linkage Strategies Based on Partially Identifying Variables." *Statistics in Medicine* 31: 4231–42.

Imai, Kosuke, and Dustin Tingley. 2012. "A Statistical Method for Empirical Testing of Competing Theories." *American Journal of Political Science* 56: 218–36.

Jaro, Matthew. 1972. "UNIMATCH-A Computer System for Generalized Record Linkage Under Conditions of Uncertainty." Technical Report, Spring Joint Computer Conference.

Jaro, Matthew. 1989. "Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida." *Journal of the American Statistical Association.* 84: 414–20.

Jutte, Douglas P., Leslie L. Roos, and Marni D. Browne. 2011. "Administrative Record Linkage as a Tool for Public Health Research." *Annual Review of Public Health* 32: 91–108.

Kim, Gunky, and Raymond Chambers. 2012. "Regression Analysis under Incomplete Linkage." *Computational Statistics and Data Analysis* 56: 2756–70.

Lahiri, Partha, and Michael D. Larsen. 2005. "Regression Analysis with Linked Data." *Journal of the American Statistical Association* 100: 222–30.

Larsen, Michael D., and Donald B. Rubin. 2001. "Iterative Automated Record Linkage Using Mixture Models." *Journal of the American Statistical Association* 96: 32–41.

McLaughlan, Geoffrey, and David Peel. 2000. *Finite Mixture Models*. New York: John Wiley & Sons.

McVeigh, Brendan S., and Jared S. Murray. 2017. "Practical Bayesian Inference for Record Linkage." Technical Report, Carnegie Mellon University.

Meredith, Marc, and Michael Morse. 2014. "Do Voting Rights Notification Laws Increase Ex-Felon Turnout?" *The ANNALS of the American Academy of Political and Social Science* 651: 220–49.

Mummolo, Jonathan, and Clayton Nall. 2016. "Why Partisans Don't Sort: The Constraints on Political Segregation." *Journal of Politics* 79: 45–59.

Murray, Jared S. 2016. "Probabilistic Record Linkage and Deduplication after Indexing, Blocking, and Filtering." *Journal of Privacy and Confidentiality* 7: 3–24.

Neter, John, E. Scott Maynes, and R. Ramanathan. 1965. "The Effect of Mismatching on the Measurement of Response Errors." *Journal of the American Statistical Association* 60: 1005–27.

Ong, Toan C., Michael V. Mannino, Lisa M. Schilling, and Michael G. Kahn. 2014. "Improving Record Linkage Performance in the

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

Presence of Missing Linkage Data." *Journal of Biomedical Informatics* 52: 43–54.
Richman, Jesse T., Gulshan A. Chattha, and David C. Earnest. 2014. "Do Non-Citizens Vote in U.S. Elections?" *Electoral Studies* 36: 149–57.
Ridder, Geert, and Robert Moffitt. 2007. *Handbook of Econometrics*. Vol. 6. Elsevier Chapter The Econometrics of Data Combination, pp. 5469–547.
Sadinle, Mauricio. 2014. "Detecting Duplicates in a Homicide Registry Using a Bayesian Partitioning Approach." *Annals of Applied Statistics.* 8: 2404–34.
Sadinle, Mauricio. 2017. "Bayesian Estimation of Bipartite Matchings for Record Linkage." *Journal of the American Statistical Association* 112: 600–12.
Sariyar, Murat, and Andreas Borg. 2016. Record Linkage in R. R package. Version 0.4-10. http://cran.r-project.org/package=RecordLinkage.
Sariyar, Murat, Andreas Borg, and Klaus Pommerening. 2012. "Missing Values in Deduplication of Electronic Patient Data." *Journal of the American Medical Informatics Association* 19: e76–82.
Scheuren, Fritz, and William E. Winkler. 1993. "Regression Analysis of Data Files that Are Computer Matched." *Survey Methodology* 19: 39–58.
Scheuren, Fritz, and William E. Winkler. 1997. "Regression Analysis of Data Files that Are Computer Matched II." *Survey Methodology.* 23: 157–65.
Steorts, Rebecca C. 2015. "Entity Resolution with Empirically Motivated Priors." *Bayesian Analysis.* 10: 849–75.
Steorts, Rebecca C., Samuel L. Ventura, Mauricio Sadinle, and Stephen E. Fienberg. 2014. "A Comparison of Blocking Methods for Record Linkage." In *Privacy in Statistical Databases*, ed. Josep Domingo-Ferrer. Springer, 253–68.
Tam Cho, Wendy, James Gimpel, and Iris Hui. 2013. "Voter Migration and the Geographic Sorting of the American Electorate." *Annals of the American Association of Geographers* 103: 856–70.
Tancredi, Andrea, and Brunero Liseo. 2011. "A Hierachical Bayesian Approach to Record Linkage and Population Size Problems." *Annals of Applied Statistics.* 5: 1553–85.
Thibaudeau, Yves. 1993. "The Discrimination Power of Dependency Structures in Record Linkage." *Survey Methodology* 19.
Winkler, William E. 1990. "String Comparator Metrics and Enhanced Decision Rules in the Fellegi–Sunter Model of Record Linkage." In *Proceedings of the Section on Survey Research Methods*. American Statistical Association. https://www.iser.essex.ac.uk/research/publications/501361.
Winkler, William E. 1993. "Improved Decision Rules in the Fellegi–Sunter Model of Record Linkage." In *Proceedings of Survey Research Methods Section*. American Statistical Association. http://ww2.amstat.org/sections/srms/Proceedings/papers/1993_042.pdf.
Winkler, William E. 2000. "Using the EM Algorithm for Weight Computation in the Felligi–Sunter Model of Record Linkage." Technical Report No. RR2000/05, Statistical Research Division, Methodology and Standards Directorate, U.S. Bureau of the Census.
Winkler, William E. 2005. "Approximate String Comparator Search Strategies for Very Large Administrative Lists." Research Report Series (Statistics) No. 2005-02, Statistical Research Division U.S. Census Bureau.
Winkler, William E. 2006a. "Automatic Estimation of Record Linkage False Match Rates." In *Proceedings of the Section on Survey Research Methods*. American Statistical Association.
Winkler, William E. 2006b. "Overview of Record Linkage and Current Research Directions." Technical Report, United States Bureau of the Census.
Winkler, William E., and Willian Yancey. 2006. "Record Linkage Error-Rate Estimation without Training Data." In *Proceedings of the Section on Survey Research Methods*. American Statistical Association.
Winkler, William E., Willian Yancey, and E. H. Porter. 2010. "Fast Record Linkage of Very Large Files in Support of the Decennial and Administrative Record Projects." In *Proceedings of the Secion on Survey Research Methods*.
Yancey, Willian. 2005. "Evaluating String Comparator Performance for Record Linkage." Research Report Series, Statistical Research Division U.S. Census Bureau.

Downloaded from https://www.cambridge.org/core. Harvard University, on 24 Apr 2019 at 23:40:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055418000783

EXHIBIT 10

UNITED STATES BANKRUPTCY COURT

DISTRICT OF MASSACHUSETTS

In re: TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Debtors.	Chapter 11 Cases 14-40987-MSH 14-40988-MSH 14-40989-MSH Jointly Administered
STEPHEN B. DARR, TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Plaintiff, v. BENJAMIN ARGUETA, A REPRESENTATIVE OF A CLASS OF DEFENDANT NET WINNERS Defendants.	Adversary Proceeding No. 16-4006
STEPHEN B. DARR AS TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Plaintiffs, v. MARCO PUZZARINI, A REPRESENTATIVE OF A CLASS OF DEFENDANT NET WINNERS Defendants.	Adversary Proceeding No. 16-4007

REPLY REPORT OF DR. CAMERON E. FREER

Respectfully submitted this 22nd day of May 2023



Cameron E. Freer, Ph.D.

TABLE OF CONTENTS

I. Introduction..... 1

II. Mr. Dennis lacks the requisite training and experience to meaningfully comment on the Freer Aggregation Methodology..... 1

A. Mr. Dennis lacks the requisite statistical training and experience..... 1

B. Mr. Dennis has never been engaged to perform a Fellegi-Sunter probabilistic record linkage analysis..... 6

C. Mr. Dennis did not analyze or offer an opinion on the majority of the Freer Aggregation Methodology or its analysis..... 8

i. Mr. Dennis barely reviewed my code..... 8

ii. Mr. Dennis formed no opinion on what sample size would be adequate to estimate input data quality..... 9

iii. Mr. Dennis performed no affirmative analysis to determine what he thought the appropriate approach was for any aspect of my methodology..... 9

iv. Mr. Dennis did not perform or offer any quantifiable error analysis or opinion..... 11

III. Mr. Dennis' lack of experience leads to many misunderstandings..... 13

A. Mr. Dennis misunderstands the role of name in the Freer Aggregation Methodology..... 13

B. Mr. Dennis misunderstands what constitutes sufficient data quality..... 22

C. Mr. Dennis misunderstands whether ground truth data is required to assess accuracy of the aggregation methodology..... 27

D. Mr. Dennis misrepresents the academic literature..... 28

i. Mr. Dennis misunderstands Christen regarding transitive closure..... 28

ii. Mr. Dennis misunderstands Christen regarding data quality..... 29

E. Mr. Dennis' statistical inexpertise leads to invalid criticisms of the Freer Report. 32

i. Mr. Dennis unreasonably attacks an example cluster illustrating high output quality..... 32

ii. Mr. Dennis misunderstands statistical significance and sample size..... 33

F. Mr. Dennis misunderstands transitive closure..... 36

G. Mr. Dennis is mistaken about the required output of the aggregation task..... 38

IV. Mr. Dennis' examples are cherry-picked and do not impeach the Freer Aggregation Methodology..... 41

V. The Freer aggregation methodology is agnostic to the choice of net equity formula..... 43

VI. Conclusion..... 44

EXHIBITS

Exhibit Number	Description
1	Materials Considered
2.1	All 12 fields from *representante* used in the Freer Aggregation Methodology, for 400 randomly-selected records chosen with fixed random seed
2.2	Values of the *rep_id* field from *representante* for these 400 records
2.3	Values of the *rep_nome* field from *representante* for these 400 records
2.4	Values of the *rep_email* field from *representante* for these 400 records
2.5	Values of the *rep_cel* field from *representante* for these 400 records
2.6	Values of the *rep_fone* field from *representante* for these 400 records
2.7	Values of the *rep_end* field from *representante* for these 400 records
2.8	Values of the *rep_numero* field from *representante* for these 400 records
2.9	Values of the *rep_cep* field from *representante* for these 400 records
2.10	Values of the *rep_login* field from *representante* for these 400 records
2.11	Values of the *rep_pwd_secondary* field from *representante* for these 400 records
2.12	Values of the *rep_cpf* field from *representante* for these 400 records
2.13	Values of the *rep_datanasc* field from *representante* for these 400 records
3	Values of the *rep_nome* field from the 300 randomly-selected User Accounts analyzed in the Freer Report

I. Introduction

1. I prepared an Expert Report (the "Freer Report") issued on April 29, 2022. The Freer Report contains my initial opinions, describes my qualifications, and includes my CV (as Exhibit 1 of the Freer Report).
2. I have prepared this Reply Report (this "Reply") in response to the Supplemental Rebuttal Expert Report of Joshua M. Dennis (the "Dennis Rebuttal") issued on April 14, 2023 and the transcript of the deposition taken of Mr. Dennis on May 9, 2023 ("Dennis Tr.").
3. I relied upon data and documents from a variety of sources during my work and in the course of generating this report. A complete listing of materials that I considered is presented in Exhibit 1 of this Reply, in conjunction with Exhibit 2 of the Freer Report. Note that I will refer to the aggregation methodology described in the Freer Report as the "Freer Aggregation Methodology", by which I mean the same thing as the "Freer Aggregation Process" or the "Borelian Aggregation Methodology".

II. Mr. Dennis lacks the requisite training and experience to meaningfully comment on the Freer Aggregation Methodology.

A. Mr. Dennis lacks the requisite statistical training and experience.

4. Mr. Dennis has limited to no experience with many of the statistical concepts and tools that I utilized in the Freer Aggregation Methodology. Mr. Dennis has a BS in Business from Skidmore College. Mr. Dennis' business coursework "involved multiple accounting courses, economics courses, finance courses, marketing courses. Things of that nature."[1]

[1] Dennis Tr., 33:17-34:3.

He minored in computer science, of which the only course dedicated to the manipulation of data that he remembers was a course teaching Microsoft Access.[2] Mr. Dennis does remember taking one semester-long statistics course.[3] Subsequent to graduating, Mr. Dennis took a continuing education course learning SQL that "spanned a couple weeks", but remembers no other formal continuing education related to data science.[4]

5. Mr. Dennis' actual probabilistic computing engagements appear to be limited to some form of fuzzy matching.[5] Even so, Mr. Dennis appears to have little comprehension of basic concepts in the field of entity resolution, or in the closely related and overlapping fields of record linkage and data matching.

6. Probabilistic record linkage is a method used to link data that refer to the same entities, such as individuals or businesses. It is called "probabilistic" because it involves the computation of the probability that each pair of records refers to the same entity.[6] It often incorporates rules and scoring systems that take into account a variety of data fields (such as name, email address, phone number, etc.) to generate the probability of a match.

7. Fuzzy matching, on the other hand, is a technique used to measure similarity between strings. This is quite different from estimating the probability that two complete records are a match.

8. Even when discussing fuzzy matching, Mr. Dennis demonstrates a lack of knowledge. He describes Jaccard similarity, Levenshtein distance, and TF-IDF as methods of fuzzy

[2] Dennis Tr., 34:16-35:17.
[3] Dennis Tr., 35:20-36:13.
[4] Dennis Tr., 36:14-38:12.
[5] Dennis Tr., 59:19-60:18, 62:7-23.
[6] When I say "entity" in this Reply, I am not referring to the concept of legal entity. Rather, I am referring to the concept of "entity" as it appears in the entity resolution, record linkage, and data matching literature: an *entity* is a distinct object (such as individual, organization, or location) that can be represented in multiple, potentially different ways across the data. The purpose of entity resolution is to identify when these different representations refer to the same object.

matching, but this is incorrect. TF-IDF (Term Frequency-Inverse Document Frequency) is not a fuzzy matching technique like Jaccard similarity or Levenshtein distance. TF-IDF is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. This mistake is especially strange given that Mr. Dennis claims to have used TF-IDF before:[7]

Q. You've used TF-IDF?
A. Correct.

Q. "TF-IDF" stands for what?
A. I'd have to go back and look. I don't recall.

Q. Is it a -- is it a fuzzy matching model?
A. It is, yes.

9. Mr. Dennis has little understanding of the notion of connected components,[8] an important concept in graph theory.[9] This notion is closely related to transitive closure,[10] a key aspect of the Freer Aggregation Methodology.

10. Mr. Dennis could not recall any specific engagements that involved transitive closure, and when he tried, Mr. Dennis again showed a lack of understanding. Transitive closure is not "putting things in one cluster and only one cluster" – that is called partitioning or creating a partition (which I discuss later on this Reply) – unlike what Mr. Dennis states:[11]

Q. Have any of your prior engagements involved transitive closure?

[7] Dennis Tr., 63:9-16.
[8] Dennis Tr., 90:15-91:5.
[9] Graph theory is a branch of mathematics germane to record linkage, and especially to the dynamic clustering approach of the Freer Aggregation Methodology.
[10] In simple terms, a connected component is a set of vertices that are all connected to each other by a path, but not to other vertices outside of the component. Transitive closure of the edges in a graph forms its connected components, thereby identifying all pairs of vertices that are reachable from each other, regardless of the specific path taken.
[11] Dennis Tr., 76:21-77:24; see also 122:14-123:22.

A. I'm trying to -- I apologize -- I'm trying to think. There have certainly been engagements that have required things to be bucketed and only input in one and only one bucket.

Q. Okay. When you're describing putting things in one bucket, did you use transitive closure in those instances, or was that just bucketizing individual bits of information?
A. My understanding is – with Dr. Freer is that transitive closure was putting something in one cluster and only one cluster.

Q. Right. But using a transitive process to do so; correct? If A equals B, B equals C, then A equals C? Well, broadly speaking.
A. I'd have to go back and think. It's possible.

Q. Okay. Well, can you think?
A. It's certainly possible. I can't think of anything sitting here today. I'll continue to think on it and let you know.

Q. Okay. Please do.

11. Mr. Dennis is not familiar with Cochran's Formula, used by statisticians to calculate an ideal sample size given a desired level of precision, desired confidence level, and the estimated proportion of the attribute present in the population.[12] Moreover, Mr. Dennis was unable to even generally explain the correlation between sample size and population size:[13]

Q. Well, generally speaking, does the required sample size grow in some linear fashion as relates to the overall data size -- dataset?
A. Not linearly, no.

Q. Well, how does it grow?
A. Well, I don't recall the math behind it.

Q. Is it logarithmic? Is it sublinear? Do you have any recollection?
A. I don't recall them.

12. Mr. Dennis does not have an understanding of what f-measure is (beyond what is described in my report), even though it is a basic statistical metric to measure accuracy in

[12] Dennis Tr., 80:9-11.
[13] Dennis Tr., 83:8-18.

binary classifications.[14] Even more concerning is that Mr. Dennis testified that he could not calculate how often one would expect to see false positives in an aggregation methodology with a precision of 0.980, a recall of 0.996, and an f-measure of 0.988 without referring back to my report.[15] (A precision of 0.980 straightforwardly implies a false discovery rate of 1 – 0.980 = 0.02, which means that 2% of the matches asserted by the methodology are false positives.) This may be because Mr. Dennis has never made such a calculation before:[16]

Q: Have you ever calculated false positives off of precision, recall, and F-measure?

A. No, I have not.

13. In fact, Mr. Dennis cannot recall any experience with calculating the statistical notions of precision and recall:[17]

Q. … I'm asking if you have any experience whatsoever calculating precision and recall as it relates to error rate?
A. Generally not.

14. Just as Mr. Dennis appears to have limited experience with commonly used data science statistical concepts, Mr. Dennis appears to have limited experience with commonly used data science tools. For example, Mr. Dennis claims only a "base knowledge" of Python (often the preferred programming language used in complex structured and unstructured data exploration).[18] Python is a very common tool of choice for data scientists because it includes a broad range of libraries used in data manipulation, including Pandas, which is specifically designed for large data set analysis.

[14] Dennis Tr., 90:11-14.
[15] Dennis Tr., 161:24-163:19.
[16] Dennis Tr., 163:22-164:1.
[17] Dennis Tr., 88:7-89:6.
[18] Dennis Tr., 45:11-14.

15. Mr. Dennis also claims to be only partially conversant with the Pandas Python package, having used Pandas libraries "a handful" of times.[19] In contrast, I have used Pandas every day that I was designing and implementing the Freer Aggregation Methodology, as well as for many years for a variety of projects involving large datasets. Pandas is an appropriate tool for working with a large dataset of nearly 11 million records.[20] I would expect any experienced data scientist to be fully conversant in using Python and its Pandas library for data analysis (or their equivalents in other languages).

16. Mr. Dennis did not use Splink, a well-known and peer-reviewed Python package for probabilistic record linkage, either before or *during* the TelexFree case.[21] This admission is surprising, as the Splink package is at the heart of the Freer Aggregation Methodology, and performs the key function of computing match probabilities between pairs of User Accounts.

B. Mr. Dennis has never been engaged to perform a Fellegi-Sunter probabilistic record linkage analysis.

17. Critically, even though the Fellegi-Sunter model underlies most of the standard approaches to probabilistic record linkage, the TelexFree engagement is the first time Mr. Dennis has been exposed to the Fellegi-Sunter model.[22] Mr. Dennis testified that he has never used the Fellegi-Sunter data aggregation method.[23] Nor has he studied the

[19] Dennis Tr., 45:24-46:12.
[20] In fact, Mr. Dennis' confusion about "numerous duplicate records" (footnote 25 of Dennis Rebuttal at 8) in the output of the Freer Aggregation Methodology is likely due to his use of tools that do not correctly parse the large CSV files that were included as Exhibits 6 and 7 of the Freer Report. Part of the codebase that I delivered to Mr. Dennis includes code that properly reads those CSV files, but it appears that Mr. Dennis did not make use of that code.
[21] Dennis Tr., 46:20-22.
[22] Dennis Tr., 57:19-58:3
[23] Dennis Tr., 62:24-63:3.

mathematics of how, in FS-EM, the relative weights of agreement across various fields are determined, an important concept for understanding and evaluating any probabilistic data aggregation methodology.[24]

18. What is even more problematic is that Mr. Dennis admits that he did not perform such an aggregation task *during* this project either, since he did not even execute the full sequence of code that I provided to him so that he could replicate my work. In fact, Mr. Dennis admits that of the code he examined (let alone executed), he primarily looked at only the cleaning and blocking sections.[25]

19. This highlights Mr. Dennis' inexperience with probabilistic record linkage because, as described in my initial report, the Fellegi-Sunter model, and specifically Fellegi-Sunter with Expectation Maximization (FS-EM), has become a standard method for performing probabilistic record linkage.

20. Mr. Dennis introduces a quote out of context[26] to cast aspersions on the Freer Aggregation Methodology, specifically to falsely imply that a gold standard is required for evaluation of FS-EM. It is not required, as I explain below in Section III.C of this Reply. Despite his insinuation, Mr. Dennis is not sure whether he has even *thought* about whether FS-EM requires a gold standard.[27]

21. Given Mr. Dennis' lack of affirmative analysis and inexperience with probabilistic record linkage in general, and with FS-EM in particular, he cannot offer a meaningful critique of my FS-EM aggregation methodology. Instead, as described in more detail below in Section IV of this Reply, he resorts to cherry-picking data that is not statistically

[24] Dennis Tr., 64:21-65:8.
[25] Dennis Tr., 26:3-8.
[26] ¶111 of Dennis Rebuttal at 36.
[27] Dennis Tr., 128:6-12

significant, so as to purportedly identify "potential" flaws – flaws that, if present, would be well within the stated error rates for the Freer Aggregation Methodology. If Mr. Dennis had a better understanding of probabilistic record linkage, he would have presented statistical evidence, rather than attempting to attack the Freer Aggregation Methodology with a small number of examples, presumably selected at his own discretion.

C. Mr. Dennis did not analyze or offer an opinion on the majority of the Freer Aggregation Methodology or its analysis.

22. In addition to lacking the appropriate training and experience to meaningfully critique the Freer Aggregation Methodology, Mr. Dennis did not even attempt to review or offer an opinion on my methodology in any meaningful way, other than to incorrectly claim that his cherry-picked examples of potential overclustering and underclustering invalidate my entire analysis, a topic that I address more fully below in Section IV of this Reply.

i. Mr. Dennis barely reviewed my code.

23. As an example of insufficient review, Mr. Dennis did not review my code for running the 10-step Freer Aggregation Methodology in full.[28] In fact, he primarily ran code only for two out of those ten steps, the blocking and cleaning steps.[29] And of blocking code, he reviewed only "a portion of it".[30] Since Mr. Dennis did not run through the code of the Freer Aggregation Methodology, he could not have possibly recreated the aggregation output.

[28] Dennis Tr., 20:18-22.
[29] Dennis Tr., 26:3-8; 31:18-32:12.
[30] Dennis Tr., 23:8-17; 74:8-11.

24. In light of Mr. Dennis' statements, I estimate the percentage of the overall code reviewed by Mr. Dennis to be less than 10% percent of the codebase. Of the portions he did review, Mr. Dennis admits that my code operates in the manner intended.[31]

ii. Mr. Dennis formed no opinion on what sample size would be adequate to estimate input data quality.

25. Mr. Dennis did not form a specific opinion on the degree to which the 300 name field values manually reviewed by me constituted an adequate sample size for verifying the quality of the name field.[32] In particular, Mr. Dennis did not determine how many records more than 300 he thinks I should have reviewed.[33]

26. It appears that Mr. Dennis did not realize that such a determination was a reasonable thing for a statistically literate expert to do: Mr. Dennis stated that he did not do an independent determination of the appropriate sample size for any purpose in this engagement because he was not asked to.[34] I discuss this issue of sample size in detail below in Section III.E.ii of this Reply.

iii. Mr. Dennis performed no affirmative analysis to determine what he thought the appropriate approach was for any aspect of my methodology.

27. Mr. Dennis testified that he was not asked to determine any affirmative analysis in this case.[35] Mr. Dennis testified having formed no opinion on the right or wrong approach is for each of the following aspects of my method:

[31] Dennis Tr., 25:18-26:2.
[32] Dennis Tr., 81:19-82:16.
[33] Dennis Tr., 82:17-20.
[34] Dennis Tr., 79:10-80:8.
[35] Dennis Tr., 97:17-98:5, 161:18-19, 198:9-10.

- He did not form an opinion one way or another about whether the analysis in the *rep_id_summary* table was accurate or reliable for purposes of distinguishing the TelexFree accounts from the Ympactus accounts and the hybrid accounts.[36]
- He did not form an opinion on what field cleaning process would have been appropriate for me to perform.[37]
- He did not determine how he thinks blocking should have been done.[38]
- He did not form an opinion or look into whether the weights in the FS-EM model were correctly determined.[39]
- He did not form an opinion on whether transitive closure was a "required step" in the aggregation process, because it was not something he was asked to do.[40]
- He did not perform any analysis to determine what percentage of the clusters in the aggregation output of the Freer Aggregation Methodology he thinks are error-free, because he was not asked to.[41]

28. Mr. Dennis takes the last example to the extreme by not even having an opinion on whether User Accounts which, in his own report, he insinuates are problematically clustered, are actually appropriately clustered or not. For example, he testified that he did not form any opinion on whether accounts in cluster 132785 (the "Maria Neves" cluster) were in fact over-aggregations,[42] despite referring to them in his Rebuttal as "potential

[36] Dennis Tr., 157:8-19.
[37] Dennis Tr., 72:22-73:4.
[38] Dennis Tr., 197:16-198:12.
[39] Dennis Tr., 66:23-67:8.
[40] Dennis Tr., 91:19-92:16.
[41] Dennis Tr., 134:23-135:10.
[42] Dennis Tr., 94:18-96:22.

issues".[43] He also did not do any analytical work to specifically arrive at a number of over-aggregations that in his opinion appear in this cluster,[44] despite discussing this cluster at length in his Rebuttal.[45]

29. Similarly, Mr. Dennis testified that he has no opinion on whether the "Shal Group" cluster 12650741 and "Dolarex" cluster 10992563, compared in his Rebuttal in paragraph 159 and Exhibit 9.4, should have been aggregated into one cluster or not.[46] This is despite specifically referring to this example as an "under-aggregation" in his Rebuttal.[47]

30. Mr. Dennis also did not form an opinion on whether the "DL1 INC" cluster 11849449, and "Linda Hackett" cluster 707775, compared in his Rebuttal in Exhibit 9.5, should be aggregated into one cluster or not,[48] despite again insinuating in his Rebuttal that it "appears to be under-aggregation".[49]

31. Because Mr. Dennis did not opine on what an appropriate approach would be for many aspects of the Freer Aggregation Methodology, it is difficult to see how he can meaningfully critique these aspects of the methodology.

iv. Mr. Dennis did not perform or offer any quantifiable error analysis or opinion.

32. Mr. Dennis' refusal to offer any opinion on what he thinks is appropriate permeates his entire error analysis. Specifically, Mr. Dennis did not perform or offer any quantifiable error analysis or opinion,[50] nor was he asked to do so.[51] As a result, he did not opine on

[43] ¶142 of Dennis Rebuttal at 47.
[44] Dennis Tr., 94:18-96:22
[45] ¶¶141-146 of Dennis Rebuttal at 47-48.
[46] Dennis Tr., 180:2-18, 185:13-22.
[47] ¶159 Dennis Rebuttal at 51.
[48] Dennis Tr., 190:5-191:11
[49] ¶161 of Dennis Rebuttal at 51.
[50] Dennis Tr., 98:6-15.
[51] Dennis Tr., 100:4-19.

or attempt to calculate what an appropriate error rate would be in this case.[52] For example, he testified that:

- He has not formed an opinion on whether a zero overclustering or underclustering error rate is required for a cluster to be acceptable.[53]
- He has not formed an opinion on what error rate would make a clustering methodology acceptable.[54]
- He has not endeavored to determine the number of missed links one would have to examine to validate or invalidate the Freer Aggregation Methodology.[55]
- He did not attempt to calculate the actual rate of overclustering in any cluster of the output of the Freer Aggregation Methodology.[56]
- He did not attempt to calculate the actual rate of underclustering in any cluster of the output of the Freer Aggregation Methodology.[57]

33. As I have illustrated, Mr. Dennis frequently fails to offer an opinion on key aspects of the Freer Aggregation Methodology. Even so, Mr. Dennis often incorrectly insinuates problems with the methodology, due to his misunderstanding. I explore his troublesome approach more fully in the rest of this Reply.

[52] Dennis Tr., 89:19-90:10.
[53] Dennis Tr., 159:24-160:5.
[54] Dennis Tr., 160:19-161:19.
[55] Dennis Tr., 167:7-19.
[56] Dennis Tr., 96:23-97:16, 114:20-115:8.
[57] Dennis Tr., 99:2-5, 114:20-115:8.

III. Mr. Dennis' lack of experience leads to many misunderstandings.

A. Mr. Dennis misunderstands the role of name in the Freer Aggregation Methodology.

34. The aggregation task in the TelexFree case is to form clusters of User Accounts based on which *person* they refer to. Mr Dennis seems not to understand this basic point. As a result, Mr. Dennis makes incorrect assertions about the Freer Aggregation Methodology,[58] and he incorrectly alleges instances of underclustering.[59] (Note that when I say "person" in this Reply, I mean to include both natural persons and non-natural persons, such as corporations.)

35. Sometimes two distinct persons will share an identifier such as address, phone number, or email address. (This can happen for two natural persons,[60] or a natural person and corporation,[61] or two different corporations.[62]) Even when two distinct persons share such identifiers, they remain separate persons, and the aggregation task is to place the corresponding User Accounts into different clusters. Mr. Dennis does not seem to understand this, and accordingly makes incorrect judgments about which clusters in the output of the Freer Aggregation Methodology constitute false negatives.

---

[58] ¶¶74-88, 108-124, 137, 157-164 of Dennis Rebuttal at 24-29, 35-42, 46, 50-52.
[59] ¶¶156-161 of Dennis Rebuttal at 50-51.
[60] For example, married couples very often share a street address, and sometimes share a phone number (especially landline) or even email address. The same can be true of family members, or unrelated people who share housing.
[61] A natural person and a corporation are not the same legal entity, even if the natural person owns the corporation, and even if the two entities share a mailing address, phone number or email address.
[62] Two different corporations are not the same legal entity, even if they are owned by the same persons, and even if they share a mailing address, phone number, or email address. For example, the U.S. Bankruptcy Court, District of Massachusetts has noted "In Massachusetts, corporations and their shareholders, or different corporations with common shareholders, are generally deemed distinct legal entities." (In re Plantation Realty Trust, 232 B.R. 279, 282 (Bankr. D. Mass. 1999).)

36. For example, Mr. Dennis asserts that Exhibit 9.4 of the Dennis Rebuttal is an instance of "under-aggregation".[63] It is not. The name fields in the respective clusters (some capitalization of "Shal Group LLC" and some variant of "Dolarex LLC") clearly refer to different persons.

37. Mr. Dennis makes numerous such mistakes. As another example, consider Exhibit 9.5 of the Dennis Rebuttal, which Mr. Dennis also insinuates is an instance of "under-aggregation".[64] Again the name fields in the respective clusters (some capitalization of "DL1 INC" and some variant of "Linda S. Hackett") clearly refer to different persons.

38. These are not mere oversights — Mr. Dennis incorrectly classifies these pairs of clusters as instances of "under-aggregation" because he misunderstands the aggregation task itself.

39. The record linkage literature is clear that the nature of an aggregation task, and specifically the choice of fields from the input data set, depends on what kind of object each cluster is meant to identify.[65] Mr. Dennis does not seem to understand this point.

40. In the TelexFree task, each cluster of User Accounts corresponds to a *person*, namely the Participant that is identified by the descriptions in the User Accounts of that cluster.[66] The

[63] ¶159 of Dennis Rebuttal at 51.
[64] ¶161 of Dennis Rebuttal at 51.
[65] For example, a widely-cited paper on entity resolution states, "How to match and combine records is also application specific. For instance, the functions used by that company to match customers are different from those used by others to match say products or DNA sequences." (O. Benjelloun, H. Garcia-Molina, D. Menestrina, Q. Su, S. E. Whang, and J. Widom. Swoosh: a generic approach to entity resolution. *The VLDB Journal,* 18, 255-276, 2009.)
Another paper on large-scale entity resolution states, "ER often requires extensive exploration and understanding of the data to formulate the matching algorithms. In both unstructured and semi-structured data scenarios, one needs to know which attributes are present and how frequently populated, which of these 'identify' an entity or are relevant for matching". (K. Qian, L. Popa, and P. Sen. Active learning for large-scale entity resolution. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1379-1388, 2017.)
[66] For more details, see Section III.G of this Reply below.

record linkage literature is clear that *personal identifiers*[67] are the fields that are relevant for identifying persons. Chief among these personal identifiers is name:[68]

> Names and other personal details play a crucial role in daily life because people are using them to identify individuals, ranging from family and friends, to work colleagues, and all the way to politicians and celebrities. For organisations both in the private and public sectors, names are often a primary source of identification of the individuals they are in contact with.
> Personal names are a major component of the information used in many data matching or deduplication processes to identify records that refer to the same individuals.

41. Indeed, the fact that names are used to identify persons is so commonplace that people often use names in this way without explicitly noticing that they are doing so. Despite protesting the use of the name field in the Freer Aggregation Methodology, Mr. Dennis uses names in exactly this way himself on many occasions, including when referring to Maria Neves[69] and Nael Lopes[70].

42. It is because of the nature of the TelexFree aggregation task and its dataset, in light of standard record linkage principles, that I justifiably chose the 11 personal identifier

[67] "The identifying data collected about individuals generally include their names and addresses, dates of birth, social security or drivers license numbers, telephone numbers, and email addresses." (P. Christen, *Data Matching*, Data-Centric Systems and Applications, Springer-Verlag, 2012 at 42.)

[68] Christen, 2012 at 42.

[69] Exhibit 3 of the Freer Report includes all of the User Accounts of the cluster of the Freer Aggregation Methodology having Cluster ID 132785. It is labeled simply as "Cluster 132785". Nevertheless, Mr. Dennis refers to the same set of records in Exhibit 16 of the Dennis Rebuttal as "Maria Neves User Account Detail". This is entirely natural to do. Where did he get the idea that this cluster corresponds to Maria Neves? Almost certainly from the field *rep_nome*, which, in this cluster, usually contains some variation of the name "Maria Neves".

[70] Mr. Dennis refers to "Nael Lopes" as a "Participant" on the basis of one record in Cluster 119427 (¶124 of Dennis Rebuttal at 41.). Looking at the table of values he provides (¶123 of Dennis Rebuttal at 41), whose last line contains a single record with *rep_nome* "nael lopes", it is clear that he is inferring who the Participant is via the contents of the name field. (In fact, the *rep_email* value for this User Account is "saminha_nunes@yahoo.com", but note that Mr. Dennis does not refer to the Participant as "Saminha Nunes".) Again, this illustrates how it is entirely reasonable and standard to use names to identify persons.

fields[71] used in the Freer Aggregation Methodology, with name in particular having a special role.

43. In contrast, consider the following example: suppose one was aggregating records of real estate property deeds, where each record denotes a sale, and the desired clustering contains one cluster of deeds per property. In that case, the most important fields would be those that identify the location of the property, such as address, GIS coordinates, or Legal Description. Note that in such records, there might be one or more fields referring to a person's name (e.g., the property owner), but such fields would *not* have a special role in this context, because the goal is for each cluster to correspond to a property (by agreeing on *where* it is), not to a person (by agreeing on *who* it is).

44. Indeed, as Chief Scientist of Remine (2017-2018) I needed to (i) perform data matching among large databases of consumer records, with a goal of one cluster of consumer records per US adult, and also (ii) perform data matching among large databases of deeds for residential properties, with a goal of one cluster of deeds per property. These two tasks had very different requirements, even though some field names were roughly in common between the two kinds of data (e.g., both had some notion of name, and some notion of address). It would have been inappropriate to use the techniques from the consumer matching problem on the deed dataset, and vice-versa.

45. Analogously, in the TelexFree aggregation task, although there are several fields containing address components, it would be inappropriate to cluster together User Accounts merely because they describe the same address, precisely because the goal is to

[71] The particular fields that I chose for inclusion in the Freer Aggregation Methodology consist of the name field (*rep_nome*) along with 10 other personal identifier fields. These were selected on the basis of their relevance to the task of personal identification, as well as on the data quality of their contents, as I describe in the Freer Report at 23-31.

have one cluster per person, not one cluster per location. Similarly, it would be inappropriate to cluster merely on the basis of email addresses, phone numbers, or other potential identifiers that can be shared between distinct persons.

46. The fact that distinct persons sometimes share personal identifiers other than name (such as address, phone number, etc.) is well-understood in the data matching literature.[72] The standard solution is to *disambiguate* by name, i.e., to use the name field to distinguish among records that share an address or other identifiers but not a name.[73] Mr. Dennis does not seem aware of this fact.

47. I accomplish this disambiguation in the Freer Aggregation Methodology in the course of the blocking step.[74] This is a standard and reasonable way to accomplish the disambiguation[75] — which it is crucial that the methodology perform via some mechanism.

48. Even as the name field plays a special role in disambiguation, all 11 fields I selected play an important role in the Freer Aggregation Methodology, and the FS-EM technique analyzes the contents of *all* of these fields in determining the relative weights of these

[72] Winkler has developed data matching techniques for decades at the U.S. Bureau of the Census, pioneering the use of Fellegi–Sunter with Expectation Maximization (FS-EM). In a widely-cited survey of matching and record linkage, Winkler refers several times to "nonmatches such as husband-wife or brother-sister pairs that agree on address information". (W. E. Winkler, Matching and Record Linkage. In *Business Survey Methods* (ed. Cox, Binder, Chinnappa, Christiansen, Colledge, and Kot). Wiley Series in Probability and Statistics, Chapter 20, pp. 355-384, 1995.)

[73] Winkler notes that "results that are closer to the truth" insofar as the method "divides the set of pairs that agree on address into those agreeing on name and demographic information and those that disagree." (Winkler, 1995.)

[74] A mathematically equivalent result could be obtained by appropriately setting pairwise scores following Step 6, but this would be more computationally intensive, as I explain in the Freer Report at 50-54.

[75] For example, Winkler describes a circumstance where performing blocking based on name is warranted in order to distinguish individuals, explaining that "Name information generally provides more distinguishing power than receipts, sales, or address information." (Winkler, 1995.)

(*Distinguishing power* is a notion in probabilistic record linkage, which Winkler defines as follows: "A procedure, or matching variable, has more distinguishing power if it is better able to delineate matches and nonmatches than another." (Winkler, 1995.))

fields. As such, the Freer Aggregation Methodology is not based primarily on any single field.

49. Therefore, name is not a "starting point",[76] nor are User Accounts "aggregated primarily"[77] by the name field, as Mr. Dennis falsely alleges by insinuating that the Freer Aggregation Methodology takes the same approach as the very different Martin Report — even going so far as to gratuitously quote from the June 2021 Court Memorandum for over a page about the Martin Report, not the Freer Report[78]. Rather, name is used to disambiguate, as is appropriate for an aggregation methodology that identifies persons, and no single field is the primary basis of the aggregation.

50. Mr. Dennis cites 3 paragraphs (Paragraphs 5, 89, and 90)[79] of the Freer Report in his grandiose but incorrect claim that a particular "understanding" of mine is of "paramount importance" and "foundational to the entirety of the Freer Aggregation Process", that I failed to provide any meaningful support demonstrating why such an understanding is grounded in fact, and that this failure renders my "Aggregation Process … unreliable for determining Net Equity." These claims are all false.

51. Mr. Dennis, in his confusion about name, repeatedly misinterprets the "understanding" in Paragraph 5 of the Freer Report.[80] It is neither "of paramount importance"[81] nor "foundational to the entirety of the Freer Aggregation Process",[82] as he falsely alleges, nor is the "understanding" itself false, as he claims.

[76] ¶76 of Dennis Rebuttal at 25.
[77] ¶70, 79, 97 of Dennis Rebuttal at 25, 26, 31.
[78] ¶79 of Dennis Rebuttal at 26-27.
[79] ¶70, 83 of Dennis Rebuttal at 21, 27-28.
[80] ¶¶70, 79-80, 131 of Dennis Rebuttal at 21-23, 26-27, 44.
[81] ¶74 of Dennis Rebuttal at 24.
[82] ¶¶74-75 of Dennis Rebuttal at 24.

52. Furthermore, Mr. Dennis incorrectly insinuates that Paragraph 90 of the Freer Report (describing the special role of name) depends on the "understanding" in Paragraph 5. It does not. The relevant issue in Paragraph 90 is the "major source of complexity with the TelexFree dataset" as is described in Paragraph 89 and fully documented in footnote 27. It merely uses Paragraph 5 as motivation (as it is where this complexity was earlier described).

53. To wit, Mr. Dennis criticizes the following Paragraph 5 of the Freer Report:[83]

> Because of the pyramid nature of TelexFree's scheme, Participants often recruited friends and relatives into the scheme. My understanding of the TelexFree scheme is that, in the course of User Account creation, it was common practice for recruits to use their own name while using the contact information (email address, phone number, street address, and other fields) of their recruiters, or for family members to share the same contact information, or for a business and an individual to share the same contact information.

54. Mr. Dennis states that "The Freer Aggregation Process is predicated on"[84] this "understanding". This is false. In fact, as I explain in Paragraph 89 of the Freer Report[85] (which Mr. Dennis also quotes), the pyramid nature of the scheme (referred to in Paragraph 5 of the Freer Report) is part of what makes the data aggregation task complex. As I note there, in some datasets, when aggregating by persons, there is not much need to disambiguate based on name because persons who are unconnected with each other are unlikely to share identifiers in the dataset. With TelexFree, this is not the case – due to its pyramid nature, one *does* need to disambiguate based on name. This fact is true independent of the "understanding" in Paragraph 5 of the Freer Report.

[83] Freer Report at 2.
[84] ¶70 of Dennis Rebuttal at 21.
[85] Freer Report at 24.

55. The "understanding" in Paragraph 5 of the Freer Report did not inform my choice of methodology, other than to suggest that it was important to consider disambiguating (which must, in turn, happen via the name field for the reasons above). Even if the "understanding" were not in fact true or justified, disambiguating based on name is still standard and appropriate in sufficiently complex datasets.

56. Mr. Dennis incorrectly states, "However, this understanding is entirely unsupported and counter to real-world case facts."[86] Even though this "understanding" did not inform my choice of methodology, it does happen to be true. The pyramid nature of the scheme is well-documented.[87] The fact that Participants often recruited friends and relatives into the scheme is also well documented.[88] All three "common practices" described in the "understanding" are also true and documented:

    a. As far back as 2015, the Trustee documented that "some Participants entered the scheme using the Email Address of another Participant"[89].

    b. Family members do commonly share contact information such as phone number and address.[90]

    c. Businesses and individuals do sometimes share the same contact information.[91]

57. Paragraph 89 of the Freer Report[92] states "As described in Paragraph 5, a major source of complexity with the TelexFree dataset is the prevalence of User Accounts that contain

[86] ¶70 of Dennis Rebuttal at 21.
[87] "Each of the Debtors in these jointly administered cases operated a Ponzi and pyramid scheme. This ruling is the law of the case in each of the jointly administered cases." (Amended Order on Motion by Chapter 11 Trustee for Entry of Order Finding That Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, December 21, 2015 at 2.)
[88] P. J. Owen, "TelexFree, other scams target local immigrants", *Telegram & Gazette*, Worcester, MA. June 30, 2014.
[89] ¶46 Affidavit of Stephen B. Darr in Support of Motion by Chapter 11 Trustee for Entry of Order Finding That Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, October 7, 2015 at 11-12.
[90] See, for example, Exhibit 9.11 of the Initial Dennis Rebuttal of July 31, 2020.
[91] See, for example, Exhibit 9.5 of the Dennis Rebuttal.
[92] Freer Report at 24.

different names, but the same email address, phone number, address, etc." This is immediately followed by a footnote:[93]

> For multiple examples of distinct Participants (with different names) having the same contact information as each other in several fields, see the following Exhibits of the July 31, 2020 Rebuttal Expert Report of Joshua W. Dennis: Exhibits 9.1 and 9.5 (individual and business), Exhibits 9.2 and 9.3 (same email or address), Exhibit 9.4 (different businesses with the same address and email), Exhibits 9.6 and 9.11 (same last name and phone, but different first names), Exhibit 10 (same email, address, and last name, but verified by affidavits to be spouses – see July 29, 2020 affidavits of Frantz Balan and Lusette Balan).

58. Although this "major source of complexity" was first described in Paragraph 5, the conclusion that Paragraph 89 draws is true independently,[94] and does not rely on Paragraph 5 for its conclusion.

59. In summary, I did not rely on this "understanding" in designing or implementing the methodology. Rather, my design and implementation of the methodology are based upon the contents of the data, established data science principles, and the nature of the task. In particular, Mr. Dennis' assertion that this "understanding," was "of paramount importance"[95] is false, as is his claim that "[i]t is foundational to the entirety of the Freer Aggregation Process".[96] Mr. Dennis' conclusion is also false: "Thus, if Dr. Freer's

[93] Freer Report at 24, footnote 27.

[94] The "understanding" in Paragraph 5 of the Freer report describes 3 circumstances (which did occur, as I analyzed above in this Reply). Each of these 3 circumstances leads to "User Accounts that contain different names, but the same email address, phone number, address, etc.", as I explain in Paragraph 89 of the Freer Report, quoted above. In the quoted footnote, I provided detailed evidence for the existence of such User Accounts in the TelexFree dataset (in fact, all taken from examples that Mr. Dennis earlier provided in his previous rebuttal report dated July 31, 2020). I do believe that this constitutes "a major source of complexity with the TelexFree dataset", and for this reason, one needs to consider methods for disambiguation. Furthermore, the existence of such User Accounts described in that footnote and leading to this complexity, is true independent of the "understanding" in Paragraph 5 of the Freer Report.

[95] ¶74 of Dennis Rebuttal at 24.

[96] ¶74 of Dennis Rebuttal at 24.

assumption regarding the Name Field is found to be inaccurate, it would render the entirety of his analysis unreliable."[97]

B. Mr. Dennis misunderstands what constitutes sufficient data quality.

60. Mr. Dennis misunderstands what constitutes sufficient data quality. He quotes from Christen to insinuate that the TelexFree dataset constitutes "garbage-in",[98] which is far from the truth, and which demonstrates that Mr. Dennis misunderstands what the record linkage literature is talking about with respect to data quality.

61. The TelexFree dataset is of more than sufficient quality to warrant performing an aggregation methodology, as I analyzed in detail in the Freer Report.[99] The dataset is overwhelmingly not "garbage-in" on the whole.[100]

62. Mr. Dennis then uses his mistaken understanding of data quality and his misapprehension of the Christen quote to further insinuate that the aggregation output is "garbage-out".[101] This, too, is nonsense. The output is of high quality, as I demonstrated in the Freer Report.[102]

63. Mr. Dennis also asserts that the minimal validation performed by the TelexFree website at the time of data input implies that the data is "garbage-in".[103] Yet examining the actual data (as I have done) reveals that it is not "garbage-in". Mr Dennis' statements suggest that he has not done *any* statistical evaluation of the data quality.

---

[97] ¶75 of Dennis Rebuttal at 24.
[98] ¶112 of Dennis Rebuttal at 37.
[99] Section III.B of Freer Report at 23-37.
[100] The Freer Report occasionally mentions "garbage" such as "..", "11111", or ",", in the context of certain field values. These represent a very small portion of the total User Accounts, as I have quantified (¶128 and elsewhere in the Freer Report).
[101] ¶112 of Dennis Rebuttal at 37.
[102] Freer Report 73-78.
[103] ¶112 of Dennis Rebuttal at 36-37.

64. Mr. Dennis appears especially concerned,[104] and unreasonably so, about the idea that names were entered into an Internet website in a minimally validated way. This is neither a problem nor unusual. In general, data entered into Internet websites for consumer products is only minimally validated. Full validation of user input to consumer websites is not common, in part because it causes commercial friction (reducing user engagement or purchases). To the extent that validation does happen on data, it is most often applied to structured data fields such as email address or date of birth, which have a standardized format. Street addresses can sometimes be validated using GIS databases, but that can be more difficult for databases of global consumers.

65. Names, on the other hand, are most often validated through single-sign-on providers, such as "Google Sign-In" or "Facebook Login". Such services were substantially less common in 2012-2014, and in any case, users find validation of names in this way to be "invasive".[105] Therefore it is neither surprising nor worrisome that the TelexFree dataset was only minimally validated.

66. As a result, it is perfectly reasonable for an aggregation methodology to make use of minimally validated data in the TelexFree case. Of course, regardless of the level of validation, the data quality still has to be sufficient, as I have determined it to be.

67. I justified the data as being of high quality in the Freer Report already, but as a further illustration, I will now walk through a statistically valid subsample to show exactly how much Mr. Dennis misunderstands the quality of the TelexFree dataset when he speaks of "garbage-in". For the 12 fields I have selected from the *representante* data table for the

[104] ¶¶95-96 of Dennis Rebuttal at 31.

[105] L. Bauer, C. Bravo-Lillo, E. Fragkaki, and W. Melicher. A comparison of users' perceptions of and willingness to use Google, Facebook, and Google+ single-sign-on functionality. In *Proceedings of the 2013 ACM Workshop on Digital Identity Management*, pp. 25-36. 2013.

Freer Aggregation Methodology, I illustrate their quality and describe their general properties.

68. Exhibits 2.1 through 2.13 show randomly-sampled values of the 12 fields included from *representante.* Exhibit 2.1 shows all 12 fields for 400 randomly-selected records (chosen with fixed random seed, as described below in Section III.E of this Reply), which provides a sense of how the values in the various columns generally cohere with one another. For example, portions of the name in *rep_nome* often appear as part of the email address in *rep_email*, and the two phone numbers (*rep_cel* and *rep_fone*) usually have a similar format to each other (e.g., are of similar length and share initial digits). Also observe that the three address components (*rep_end, rep_numero,* and *rep_cep*) typically appear believable in combination (e.g., the number in *rep_numero* is sometimes part of *rep_end*).

69. Exhibit 2.2 shows the 400 values occurring in Exhibit 2.1 for the first of these fields, *rep_id*, sorted numerically. They are randomly distributed within the range of *rep_id* values that occur in *representante*.

70. Exhibit 2.3 shows the 400 values that occur in Exhibit 2.1 for the field *rep_nome*, sorted alphabetically. They are generally of high quality, and are quite similar in character to those values shown in Exhibit 3 (see Section III.E.ii of this Reply below).

71. Exhibit 2.4 shows the 400 values that occur in Exhibit 2.1 for the field *rep_email*, sorted alphabetically using the characters in reverse order, so as to draw out the patterns in domain names. (Empty values are represented with a gray box.) Note that 85.5% have a domain name among the top four values: “gmail.com” (48.25%), “hotmail.com” (26.5%), “yahoo.com” (6.75%), “163.com” (4%). In fact, 98.5% of the 400 values are in the

standard format for email addresses as defined in an Internet specification concerning email address, RFC 2822,[106] using the same email validation test that I ran in the Freer Report.[107]

72. Exhibit 2.5 shows the 400 values that occur in Exhibit 2.1 for the field *rep_cel*, sorted alphabetically. Observe that it very often appears to be plausibly a phone number (e.g., contains numerical digits 0 through 9, and is of an appropriate length).

73. Exhibit 2.6 shows the 400 values that occur in Exhibit 2.1 for the field *rep_fone*, sorted alphabetically. Observe that it, too, very often appears to be plausibly a phone number.

74. Exhibit 2.7 shows the 400 values that occur in Exhibit 2.1 for the field *rep_end*, sorted alphabetically. Observe that it very often appears to be plausibly a street name (sometimes including a house number as well).

75. Exhibit 2.8 shows the 400 values that occur in Exhibit 2.1 for the field *rep_numero*, sorted alphabetically. (Empty values are represented with a gray box.) Observe that it often appears to be plausibly a street number; in some cases, the number is already included in that User Account's *rep_end* value and does not appear again here (see Exhibit 2.1).

76. Exhibit 2.9 shows the 400 values that occur in Exhibit 2.1 for the field *rep_cep*, sorted alphabetically. Observe that it, too, very often appears to be plausibly a postal code (such as ZIP Code within the US).

77. Exhibit 2.10 shows the 400 values that occur in Exhibit 2.1 for the field *rep_login*, sorted alphabetically. Its values appear consistent with being a user-selected login name.

[106]RFCs are specifications for Internet services, and RFCs 2821 and 2822 concern email addresses. For RFC 2822 specifically, see "RFC 2822", retrieved from https://datatracker.ietf.org/doc/html/rfc2822

[107] Freer Report at 26-27.

78. Exhibit 2.11 shows the 400 values that occur in Exhibit 2.1 for the field *rep_pwd_secondary*, sorted alphabetically. It is null for 75% of these User Accounts, it takes the value "f1fb1ea9a408b1f321062e86ec4bae3e" twice, and it is distinct for the remaining 24.5% of values, consistent with being the md5sum hash of a user-supplied password (when non-null).
79. Exhibit 2.12 shows the 400 values that occur in Exhibit 2.1 for the field *rep_cpf*, sorted alphabetically. It is always non-null, and often appears to plausibly be a national ID value (analogous to Social Security Number in the US). However, as I analyzed in the Freer Report,[108] it turns out that the contents of *rep_cpf* often vary widely within a collection of User Accounts that largely agree on other field values, though this is not apparent from these 400 values.
80. Exhibit 2.13 shows the 400 sorted values that occur in Exhibit 2.1 for the field *rep_datanasc*, sorted chronologically. It is null for 77% of these User Accounts, and otherwise takes distinct values. All but one of these values is in the range between "1947-10-26" and "1996-10-04" (which all seem highly plausible as birthdates), and one is "2013-01-28" (which seems implausible given the timing of the TelexFree scheme).
81. Note that (with the exception of *rep_cel* and *rep_fone* contents looking very similar to each other, and fairly similar to the contents of *rep_cpf*), one could easily infer how to label the contents of each of these fields among the 12. In other words, even without the field names at the top of each page, it is clear that Exhibit 2.3 contains names, and that Exhibit 2.4 contains email addresses, etc.
82. This illustrates that the input data is consistent with its labeled field name, and is of sufficient quality for use in the Freer Aggregation Methodology.

[108] Freer Report at 30.

C. Mr. Dennis misunderstands whether ground truth data is required to assess accuracy of the aggregation methodology.

83. Mr. Dennis is confused about the potential need for ground truth data in the Freer Aggregation Methodology.[109] Such data is not required for either running the methodology or for evaluating its output.

84. Some machine learning algorithms fall under the category of "supervised learning", and require a dataset that is labeled with "ground truth" for both training the model and for evaluating the accuracy of the trained model. Other machine learning algorithms perform "unsupervised learning", and do not rely on labeled data for either training or assessment. Unsupervised learning algorithms identify patterns and structure within the input data itself during training.[110]

85. Unsupervised learning techniques are crucial in record linkage, as Linacre (author of the Splink software package) explains: "Manual labelling of comparisons is error prone and costly so isn't a great option."[111]

86. The standard method in the field of probabilistic record linkage, the Fellegi-Sunter model with Expectation Maximization (FS-EM), is an unsupervised learning technique that accomplishes high-quality record linkage from a dataset that is unlabeled, such as the TelexFree dataset.[112]

87. The technique behind the Splink Python package, which implements FS-EM for large datasets, has itself been extensively tested using gold standard data, and has found to be

[109] ¶111 of Dennis Rebuttal at 36.
[110] C. M. Bishop, *Pattern Recognition and Machine Learning,* Springer, 2006.
[111] R. Linacre, *The Intuition Behind the Use of Expectation Maximisation to Train Record Linkage Models,* October 14, 2022. Retrieved from https://www.robinlinacre.com/em_intuition/
[112] Referring to "[h]ow unsupervised learning is used to estimate model parameters in Splink", Linacre notes that "[t]he Expectation Maximisation (EM) algorithm provides a solution" to "the problem of training a record linkage model" without labeled data. (Linacre, 2022.)

quite accurate and with well-calibrated error rates.[113] As a result, Splink is a reliable probabilistic record linkage package for use in the unsupervised learning setting of TelexFree data.

D. Mr. Dennis misrepresents the academic literature.

88. It is unclear whether Mr. Dennis is being intentionally misleading, or simply misunderstands the content (due to his inexpertise), but Mr. Dennis systematically misrepresents the academic literature. Accordingly, one should be cautious in reading his evaluations of academic literature.

i. Mr. Dennis misunderstands Christen regarding transitive closure.

89. The way Mr. Dennis quotes Christen about transitive closure demonstrates a lack of understanding of Christen's point, and shows that Mr. Dennis misunderstands the reason that the Freer Aggregation Methodology uses dynamic clustering.

90. In Paragraph 126 of the Dennis Rebuttal, Mr. Dennis says "On this point, Section 6.8 of the Christen Text addresses issues managing transitive closure of clusters stating:"[114] and then quotes Christen:[115]

> The transitivity of matches can also lead to **problems** in that 'chains' of records, where individual pairs are classified as matches, are formed. The records at the two ends of the chain can, however, be quite different from each other, and they would not be considered to correspond to a match. [bold emphasis added]

[113] T. Enamorado, B. Fifield, and K. Imai. Using a probabilistic model to assist merging of large-scale administrative records. *American Political Science Review*, 113(2), 353-371, 2019.
[114] ¶126 of Dennis Rebuttal at 42.
[115] Section 6.8 of Christen, 2012.

91. Mr. Dennis incorrectly sees the circumstance of chaining as an "issue"[116] with the Freer Aggregation Methodology, rather than as an unavoidable and standard circumstance that *any* aggregation methodology must address. Rather, Christen's text and Mr. Dennis' quote of my deposition[117] are saying that in developing an aggregation methodology, one must carefully address this circumstance, *not* that the aggregation methodology is inherently problematic because of this inevitably-arising circumstance.

92. Indeed, in a more recent text (of which Mr. Dennis is seemingly unaware), Christen himself proposes a way of addressing this challenge:[118]

> One challenge with both traditional probabilistic data linkage and modern supervised pairwise classification techniques is that they can lead to inconsistent results due to nontransitivity (Christen, 2012). If we have classified record pair (R1, R2) as a match and (R1, R3) as a match, then the pair (R2, R3) should also be a match. However, when the pairs are looked at separately, the third pair might have been classified as a non-match – or even not compared at all if blocking was used. **This issue can be overcome by considering data linkage classification as a clustering problem** (Hassanzadeh, Chiang, Lee, & Miller, 2009) where the aim is to group all records that refer to the same entity into one cluster. [bold emphasis added]

93. I do precisely as Christen suggests, and overcome this "issue" by the use of dynamic clustering (which views the data linkage task as a clustering problem) in Step 8 of the Freer Aggregation Methodology.[119]

ii. Mr. Dennis misunderstands Christen regarding data quality.

94. Mr. Dennis shows a lack of understanding of data accuracy, and demonstrates such by misleadingly quoting from Christen.

[116] ¶125 of Dennis Rebuttal at 42.
[117] ¶127 of Dennis Rebuttal at 42.
[118] P. Christen. Data linkage: The big picture. *Harvard Data Science Review*, 1(2), 2019.
[119] Freer Report at 63-72.

95. Christen writes:[120]

> Arguably the most important data quality dimensions for data matching and deduplication are *accuracy* and *consistency*, **because a large portion of efforts in the indexing, comparisons and classification steps (that will be covered in Chaps. 4–6) deal with inaccurate and inconsistent data. If data would be of perfect quality, then data matching could be accomplished through straightforward database join operations and no sophisticated indexing techniques or approximate comparison functions would be needed. As long as data are of imperfect quality, however, techniques are needed that can deal with inaccurate and inconsistent data while still achieving high matching quality.** [bold emphasis added]

96. Mr. Dennis quotes the first clause of this quote in Paragraph 94 of the Dennis Rebuttal.[121] He conveniently omits the rest of the sentence and the rest of the paragraph, and completely misses the point. Christen explains that *because* data is not perfectly accurate, researchers in data matching have developed methods for dealing with imperfectly accurate data. As Christen explains, the data matching task is easy and sophisticated techniques are unnecessary when the data is perfectly accurate. The subject of data matching and its associated academic literature exist precisely *because* data is seldom perfectly accurate.[122]

97. These "techniques … that can deal with inaccurate and inconsistent data while still achieving high matching quality" are in fact the subject of Christen's book. The rest of the book following this introduction describes precisely the techniques that I use in the Freer Aggregation Methodology to achieve "high matching quality" in the face of "imperfect quality" input data. These techniques described by Christen begin with field selection and data cleaning, followed by blocking, field and record comparison using

[120] Christen, 2012 at 40.
[121] Dennis Rebuttal at 30.
[122] Note that Mr. Dennis does not even say what would constitute such reliability and accuracy. He specifically testified that he does not have an opinion of what is a sufficient quantum of proof to verify that data in this case was reliable and accurate. (Dennis Tr., 173:21-175:11.)

string nearness measures for fuzzy matching, Fellegi-Sunter with Expectation Maximization to form clusters based on these comparisons, and assessment of error rates. The Freer Aggregation Methodology performs precisely those steps that Christen recommends in the continuation of the quote that Mr. Dennis omitted, so as to obtain "high matching quality".

98. Mr. Dennis might not understand this, but the presence of errors and variations in field values such as names and addresses is to be expected, and record linkage can often still be successfully performed. For example, Christen et al. write, "Furthermore, since the identification of the same entity is most often based on imperfect data, the linkage has to consider variations and errors in quasi-identifiers[123] such as names and addresses."[124] The same authors go on to say that "When data are manually entered (typed), then typographical errors are not uncommon, especially in name and address values"[125]. The task of record linkage is precisely to merge records in the face of input data errors: "Record linkage tries to link databases containing error-prone indirect quasi-identifying identifiers… This generally requires error-tolerant comparison functions, also known as approximate matching functions".[126] In fact, this is often possible for messy data; the book describes a number of approaches that "can achieve high linkage quality even for databases that contain records with errors and variations in quasi-identifying attribute values."[127]

99. By following the steps that Christen describes, it is possible to obtain high quality data matching, even when the input dataset is imperfect, as many real-world datasets are. The

[123] Christen refers to descriptions in records as "quasi-identifiers".
[124] P. Christen, T. Ranbaduge, and R. Schnell. *Linking sensitive data: Methods and techniques for practical privacy-preserving information sharing.* Springer, 2020 at 25.
[125] Christen, Ranbaduge, and Schnell., 2020 at 60.
[126] Christen, Ranbaduge, and Schnell, 2020 at 9.
[127] Christen, Ranbaduge, and Schnell, 2020 at 260.

Freer Aggregation Methodology follows these steps, and accordingly obtains a high quality aggregation output.

E. Mr. Dennis' statistical inexpertise leads to invalid criticisms of the Freer Report.

100. Several of Mr. Dennis' critiques of the Freer Report are flawed due to his poor understanding of statistical analysis and sample size, as I now illustrate in two instances.

i. Mr. Dennis unreasonably attacks an example cluster illustrating high output quality.

101. Referring to Exhibit 3 of the Freer Report, Mr. Dennis writes, "Dr. Freer provides no discussion regarding if or how this single Cluster, presumably selected at his own discretion, is representative of the entire dataset."[128] To claim that it was chosen at my "discretion" is unreasonable; as I stated, I selected it because it is the "cluster that has the greatest Total Net Equity",[129] as Mr. Dennis conveniently omits.

102. Mr. Dennis misunderstands the statistical relevance of such an example. I stated that it is an "example cluster", chosen "[t]o illustrate several aspects of the aggregation methodology[130], rather than "representative of the entire dataset" as Mr. Dennis claims.

103. Furthermore, this cluster does in fact illustrate the high quality of the output. For example, consider its false discovery rate (FDR). In Exhibit 16 of the Dennis Rebuttal, Mr. Dennis has manually grouped its 471 User Accounts into three sections. The first section has two User Accounts (those with *rep_id* 11501403 and 15376207); the *rep_nome* values are "maria lares" and "rita lares" respectively, and are plausibly false

[128] ¶100 of Dennis Rebuttal at 32.
[129] ¶214 of the Freer Report at 75.
[130] ¶214 of the Freer Report at 75.

positives for this cluster. The second section has three User Accounts (those with *rep_id* 196685, 1902924, and 16337571); the *rep_nome* values are "maria neves","Maria Neves", and "maria neves" respectively, but the *rep_email* values differs from those in the rest of the cluster, and these records are potential false positives for the cluster. Mr. Dennis' grouping of these User Accounts indicates as much. Even assuming that all five of the User Accounts in the first two sections were false positives, the FDR for the records in this cluster would be approximately 1.1%, consistent with the 1.98% overall FDR that I estimate in the Freer Report.[131]

ii. Mr. Dennis misunderstands statistical significance and sample size.

104. Mr. Dennis also does not appear to understand the concept of statistical significance in relation to sample size. He incorrectly insinuates that a sample size of 300 out of millions of user accounts was too small to draw a valid conclusion about the quality of the name field.[132] In particular, Mr. Dennis' focus on the percentage of records used in the estimate (0.0028%) rather than the number (300) is misleading and reveals his lack of understanding of statistical principles.[133]

105. In the Freer Report, I describe my methodology for estimating the quality of the name field:[134]

> Indeed, the name field (*rep_nome*) in the TelexFree dataset is of very high quality (despite variations in case, punctuation, spaces, and other typical issues that data cleaning and fuzzy matching can account for). First, the name field is only empty in 0.002% of User Accounts and is never null. Second, nearly all of the name values contain strings consisting entirely of alphabetical characters or the symbols commonly found in names

[131] ¶213 of Freer Report at 74.
[132] "Dr. Freer performed a manual review of just 300 Name Field values from this remaining population, or 0.0028% of remaining User Accounts with Name Field values that are both five characters or longer and contain only the characters commonly found in names." (¶104 of Dennis Rebuttal at 34.)
[133] For example, Mr. Dennis is unfamiliar with Cochran's Formula, a standard statistical formula for determining how to choose sample size. (Dennis Tr., 80:9-83:18)
[134] Freer Report at 25-26.

(period, comma, hyphen, space, or apostrophe) – as opposed to numbers or other symbols – which is an indication that the value is a legitimate name. Specifically, 97% of the name field values are both five characters or longer and contain only the characters described above. Furthermore, a manual review reveals that in a random sample of 300 name field values meeting this criterion, 98% look like a name. Here are the first 10 values taken from that random sample: …

106. Exhibit 3 shows the 300 values of *rep_nome* on the basis of which I estimated that "98% look like a name". Note that the first 10 values are the same as in the Freer Report. I set a *random seed* in a specific way that makes the methodology reproducible, and demonstrably *not* cherry-picked.[135]

107. Note that while these values shown are all within the subset of 97% for which the "name field values are both five characters or longer and contain only the characters described above", I manually looked at this subset only so as to quickly obtain a valid lower bound. In fact, the total portion of *rep_nome* values that "look like a name" is somewhat higher than my lower bound, because some values in the unreviewed 3% also contribute to overall name field quality.[136]

108. Mr. Dennis incorrectly asserts that "Dr. Freer estimates that in 95% of the User Accounts, the value of the *rep_nome* field appears to be a name."[137] In fact, I state, "I estimate that in **at least** 95% of the User Accounts, the value of the *rep_nome* field appears to be a name."[138] [bold emphasis added] This is because, as noted above, I did not manually review 3% of the records. As a result, Mr. Dennis is also incorrect when he alleges that "by Dr. Freer's own admission, there are **over** 500,000 User Accounts that do

[135] I randomly sampled 100 values with each of the random seeds 101, 102, and 103 using the *sample* method of pandas after first subselecting the 10,606,222 records of the 10,987,617 User Accounts that were both five characters or longer and contain only alphabetical characters, period, comma, hyphen, space, or apostrophe. Setting a random seed to 101, then iterating through values 102, 103, etc., is a standard technique to get representative samples in a way that is reproducible and demonstrably not cherry-picked.

[136] For example, this 3% includes many valid Chinese names of 4 or fewer characters, or business names containing a number.

[137] ¶105 of Dennis Rebuttal at 34.

[138] ¶94 of Freer Report at 26.

not appear to contain a name in the Name Field."[139] [bold emphasis added] Mr. Dennis is making a basic mathematical error; the correct conclusion to be drawn from my statements is an *upper* bound, not a lower bound, on the number of "User Accounts that do not appear to contain a name in the Name Field".

109. At the time of writing the Freer Report, I noted that 7 values of *rep_nome* (from the subsample of size 300) appear likely to be something other than actual names, while the remaining 293 are plausibly names (albeit with occasional typos, abbreviations, and unusual capitalization). These 7 values are: "sd lkjhgf", "xxxxxxxxxxxx xxxxxxxxxxx", "aaaaa aaaaa", "LARRY KING LIVE", "VOIP VOIP", "svgb wang", "lightning carrier". This was the basis of my estimate that "98% look like a name".

110. These 300 field values provide a meaningful estimate of how often field values of *rep_nome* appear to be a name, contrary to Mr. Dennis' insinuation that 300 records is insufficiently small to draw a conclusion. The sample size that is needed to draw a statistically significant conclusion typically grows much more slowly than linearly in the population size. For example, using the Sample size calculator[140] of Qualtrics, a large firm that enables millions of surveys annually, the Ideal Sample Size on a population of size 11,000 to obtain a 5% Margin of Error with 90% Confidence Level is 265, while on a population size of 11,000,000 (1000 times larger), the Ideal Sample Size is still just 271 (less than 300). For this reason, it is misleading for Mr. Dennis to refer to the percentage of records ("0.0028%") used in the estimate, rather than the number ("300"), and statistically illiterate for Mr. Dennis to suggest that 300 is far too small of a sample size on which to base any conclusion.

[139] ¶105 of Dennis Rebuttal at 34.
[140] Qualtrics, "Sample size calculator", April 12, 2023. Retrieved from https://www.qualtrics.com/blog/calculating-sample-size/

111. With only a modestly larger sample size, the confidence could be increased even further, and the margin of error made even smaller. For example, a sample size of 1067 yields 3% Margin of Error with 95% Confidence Level on a population size of 11,000,000. Using the same methodology as for the 300 samples and the same initial random seed, I have reproducibly obtained 1067 samples of *rep_nome* meeting the above criteria, and found 21 of these records to be likely something other than names. As a result, I estimate that out of "the name field values [that] are both five characters or longer and contain only the characters described above", approximately 98.03% "look like a name", consistent with my earlier 98% estimate, and now with even higher confidence and lower margin of error. This is hardly surprising, as 300 was already sufficient for a statistically significant conclusion, and further illustrates how Mr. Dennis' criticism of that sample size of 300 was unfounded.

F. Mr. Dennis misunderstands transitive closure.

112. Mr. Dennis misunderstands the concept of transitive closure.[141] As a result, several of Mr. Dennis' attacks in the Dennis Rebuttal[142] are unfounded.

113. In an aggregation task, a set of records is partitioned into clusters, with one cluster per entity that is identified by the descriptions contained in these records.[143] (In the case of TelexFree, each record is a User Account, and a cluster of User Accounts identifies its Participant.) The definition of such a *partition* of the set of records is that it is mutually

[141] Dennis Tr., 77:4-23.
[142] ¶¶125, 126, 128, and 131 of Dennis Rebuttal at 42-44.
[143] "By definition, an ER [entity resolution] process partitions a set of references into disjoint subsets (sometimes called clusters), where all the records in that subset are judged by the process to refer to the same entity." (J. Talburt, *Entity Resolution and Information Quality*. Elsevier, 2011 at 84-85.)

exclusive and collectively exhaustive ("MECE"): it is a collection of non-overlapping, non-empty clusters of records whose union is equal to the original set of records.[144]

114. Partitions satisfy a key property: they are transitively closed, in the sense that if records A and B are in the same cluster as each other, and records B and C are in the same cluster as each other, then A and C must be in the same cluster as each other. This basic mathematical result is well-known in the entity resolution literature, and it follows from the fact that every record appears in one and only one cluster of a partition.[145]

115. In particular, the output of *any* aggregation task will be transitively closed, regardless of the technique that is used. Mr. Dennis does not seem to understand this, and he unfairly disparages the Freer Aggregation Methodology on account of his failure to understand:[146]

> Q. Okay. And do you have any opinion as to whether transitive closure was something that was a required step in this process that Dr. Freer engaged in? … in other words, if you don't do transitive closure, is there another way to skin the cat? …
> A. I think it depends.
>
> Q. Depends on what?
> A. Depends on the approach taken.

116. The fact that every partition is transitively closed means that some false positives and false negatives will be unavoidable in any aggregation task. To simplify Mr. Dennis' example,[147] consider three records A, B, C, where A and B largely agree in their descriptions, B and C largely agree in their descriptions, and A and C substantially disagree in their descriptions. One might ask for a clustering in which A and B occur in

[144] "A partition of a set S is simply a collection of nonempty subsets of S … that do not overlap … yet is large enough to cover all of S" (Talburt, 2011 at 82.)
[145] Theorem 3.4 of Talburt, 2011 states that every partition is transitively closed in the sense that the set of pairs of records that occur in the same cluster as each other is a relation that satisfies the transitive closure property (see Definition 3.1). (Talburt, 2011 at 81-83.)
[146] Dennis Tr., 91:19-92:6.
[147] ¶125 of Dennis Rebuttal at 42.

the same cluster, B and C occur in the same cluster, and A and C do not appear in the same cluster. However, this is impossible to achieve by *any* technique. Every partition of the records will result in either (i) A and B being in different clusters (potentially a false negative), or (ii) B and C being in different clusters (potentially a false negative), or (iii) A and C being in the same cluster (potentially a false positive). This fundamental tension is inherent to aggregation itself, and not to any particular chosen technique. Mr. Dennis misunderstands this basic point, and unreasonably blames the Freer Aggregation Methodology for this "issue".[148]

117. Indeed, this tension explains why any reasonable aggregation methodology will tend to have some false positives and some false negatives, a related point on which Mr. Dennis appears confused.[149] Mr. Dennis seems to think that finding a small number of false positives or a small number of false negatives exhibits a failure of a methodology, rather than understanding that this is inevitable in any aggregation methodology.

G. Mr. Dennis is mistaken about the required output of the aggregation task.

118. Mr. Dennis is wrong about the required output of the aggregation task.[150] As a result, Mr. Dennis incorrectly claims that I have "Failed to Deliver the Full Scope" of my "Engagement".[151] The required output in the TelexFree case is a set of clusters of User Accounts such that each cluster identifies a Participant. This is the output of the Freer Aggregation Methodology.

119. More concretely, the required output is a set of clusters of User Accounts, where each cluster belongs to a single Participant and every Participant has exactly one cluster. Mr.

[148] ¶125 of Dennis Rebuttal at 42.
[149] ¶125 of Dennis Rebuttal at 42 and Dennis Tr., 124:19-126:12.
[150] ¶¶70, 163-165 of Dennis Rebuttal 22-23 and 52-53.
[151] Dennis Rebuttal at 52.

Dennis seems to misconstrue this process in his Rebuttal,[152] believing that the Freer Aggregation Methodology is a "purely mathematical exercise"[153] that does not achieve the intended task. However, the Freer Aggregation Methodology does indeed produce the required output, as described in the Freer Report and as I now outline:

a. A Participant, as per the Court's definition, is a person.[154]

b. A person is identified through various *descriptions*,[155] such as their name (including any aliases), addresses, phone numbers, email addresses, and other personal identifiers.[156] For example, a Participant might be identified as follows, from the descriptions in its cluster: "The Participant is a person called Maria Neves or Maria T Neves, living at one or more of these 3 addresses, and with one or more of these 4 phone numbers…"

c. Each User Account has descriptions that identify a Participant, including name, address, phone numbers, and email address.[157]

d. The aggregation methodology creates *clusters* of User Accounts that identify Participants using the descriptions contained in these User Accounts. Specifically, starting with the set of User Accounts, it generates a set of clusters such that (i)

[152] ¶¶70, 163-165 of Dennis Rebuttal at 22-23 and 52-53.
[153] ¶¶70, 163 of Dennis Rebuttal at 22 and 52.
[154] Memorandum in Support of Proposed Supplemental Order Respecting Motion by Chapter 11 Trustee for Entry of Order Finding that Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, January 19, 2016 at 3.
[155] "One of the most critical tasks for improving data quality and increasing the reliability of data analytics is Entity Resolution (ER), which aims to identify different descriptions that refer to the same real-world entity." (V. Christophides, V. Efthymiou, T. Palpanas, G. Papadakis, and K. Stefanidis. An overview of end-to-end entity resolution for big data. *ACM Computing Surveys* (CSUR) 53, no. 6: 1-42. 2020.)
[156] "The identifying data collected about individuals generally include their names and addresses, dates of birth, social security or drivers license numbers, telephone numbers, and email addresses." (Christen, 2012 at 42.)
[157] "ER aims to identify different descriptions that refer to the same real-world entity appearing either within or across data sources, when unique entity identifiers are not available. Typically, ER aims to match structured descriptions (i.e., records) stored in the same (a.k.a., deduplication), or two different (a.k.a., record linkage) relational tables." In this terminology, the task for the TelexFree case is one of deduplication. (Christophides et al., 2020.)

each User Accounts is in one and only one cluster and (ii) each cluster contains those and only those User Accounts that, through the set of their descriptions, identify a Participant.[158]

120. Note that having multiple descriptions is helpful towards contacting the Participant, as it provides a broader range of ways to reach them.

121. Also, note that there is no predetermined list of Participants, nor is their total number known in advance.[159] The number or list of Participants are not required as input to the aggregation methodology.[160] Instead, these details emerge from the aggregation methodology itself. Since there is no predetermined list of Participants to which User Accounts or clusters can be assigned, the only way that an aggregation methodology "assigns" User Accounts to a Participant is via the way that descriptions contained in the User Accounts of a given cluster identify a Participant.

122. The way I have described the task of the aggregation methodology is standard in the entity resolution ("ER") literature:[161]

> [A]n external view of identity is one in which some number of attribute values for an identity have been collected, but it is not certain if it is a complete collection of values or even if all the values are correct. … In an external view, the collection of attribute values that have been linked together by the system comprises its view of the identity of the entity. In other words, an ER system based on an external view builds its knowledge about entity identities piece by piece. The external view of identity more closely

[158] In the standard approach to entity resolution with a single dataset as input, "matching entities are typically grouped within disjoint clusters such that any two entities in a cluster should match with each other and no entity should match with entities of other clusters." (A. Saeedi, M. Nentwig, E. Peukert, and E. Rahm. Scalable matching and clustering of entities with FAMER. *Complex Systems Informatics and Modeling Quarterly* 16: 61-83, 2018.)

[159] This circumstance is common in record linkage: "clustering algorithms must be *unconstrained* algorithms, that is, the algorithms do not require as input the number of clusters or other domain specific parameters. For this framework, they … should be able to produce a large (and unknown) number of clusters." (O. Hassanzadeh, F. Chiang, H. C. Lee, and R. J. Miller. Framework for evaluating clustering algorithms in duplicate detection. *Proceedings of VLDB '09*: 1282-1293, 2009.)

[160] "[T]he core characteristic of clustering algorithms is that they produce a set of disjoint entity clusters without requiring as input the number of clusters or any labeled dataset for training". (Christophides et al., 2020.)

[161] Talburt, 2011 at 31-32.

resembles the experience of a business or government agency using ER tools in an effort to link their records into a single view of a customer or agency client.

123. In summary, I have created a set of clusters of User Accounts, each of which identifies a Participant. This is the required output for an aggregation methodology in the TelexFree case.

IV. Mr. Dennis' examples are cherry-picked and do not impeach the Freer Aggregation Methodology.

124. In the course of his Rebuttal, Mr. Dennis provides a small number of examples of false positives and false negatives that he incorrectly thinks impeach the methodology.[162] They are not statistically significant or representative, but are merely anecdotal. Mr. Dennis appears to have specifically searched for false positive and false negative examples, an approach colloquially known as "cherry picking":[163]

> Q. And the tables that were generated by your office, did you include and attach all of the tables that you generated in the course of your investigation or just selective ones?
> A. Selected ones.
>
> Q. All right.

125. It is unsurprising that he found such examples, because they will exist in any large aggregation output. In fact, my stated error rates imply that there are many more such examples, and nevertheless the aggregation methodology is reliable and the results are of high quality.

126. In contrast, Mr. Dennis does not understand how error rates work, and confuses anecdotal evidence for statistical evidence:[164]

[162] ¶¶115-119, 128-129, 133-134, 138-139, 149-162 of Dennis Rebuttal at 37-39, 42-46, 48-52.
[163] Dennis Tr., 18:12-17.
[164] Dennis Tr., 98:16-99:1.

Q. I'm really just asking you did you perform any error rate calculation overall for which you considered to be overclusters in Dr. Freer's clustering?
A. I certainly provided plenty of examples of instances that appear to be overclustering or instances where his clusterings don't comport with the stated goals.

127. It is difficult to even imagine an expert in record linkage who has no experience calculating precision and recall, which are fundamental measurements of Type I and Type II errors. Yet Mr. Dennis admits that he has no such experience:[165]

Q. And while that's happening, did you do any work -- well, are you familiar with the terms "precision" and "recall" in statistical work -- in statistical analysis?
A. To some degree.

Q. When you say "to some degree," what do you mean?
A. I certainly read Dr. Freer's report and how he described those terms.

Q. Other than reading Dr. Freer's report and how he described those terms, do you have independent experience calculating precision and recall?
A. That's not something I've been asked to do.

Q. I mean generally speaking.
A. There are specific definitions, and they're, again, not something I typically focus on or I've been asked to focus on.

Q. I'm not asking whether you typically focus on. I'm asking if you have any experience whatsoever calculating precision and recall as it relates to error rate?
A. Generally not.

128. Mr. Dennis appears unaware that a nonzero error rate is both appropriate and inevitable. As a result of this inexperience, Mr. Dennis menacingly states that even a single false positive or single false negative potentially dooms an aggregation methodology:[166]

Q. Okay. Let me give you a hypothetical. In a given aggregation methodology, in your opinion, would a single missed link, either overaggregation or underaggregation, would that invalidate the methodology?
A. Potentially.

[165] Dennis Tr., 88:7-89:14.
[166] Dennis Tr., 164:2-8.

129. This may explain why Mr. Dennis does not seem to think that error rates are important. Mr. Dennis does not comment on or evaluate my error rate calculations:[167]

> Q. Well, let's take it from the abstract to the concrete. In this case, did you calculate error rate, what would be an appropriate error rate for the output of Dr. Freer's cluster methodology?
> A. I didn't opine as to what an appropriate error rate would be in this case, no.
>
> Q. Whether you opined or not, did you attempt to calculate what would be an appropriate error rate?
> A. No, I did not.

130. Mr. Dennis does not do his own calculations of overall error rates:[168]

> Q. Okay. And as a result of the quantification exercise, did you attempt to calculate any overaggregation or underaggregation error rates? …
> A. Not in this case.

131. Throughout the Dennis Rebuttal, Mr. Dennis doesn't even express his own examples or points in terms of error rates.[169]

132. In summary, Mr. Dennis' anecdotal evidence should be disregarded from a statistical perspective, and certainly does not render the methodology unreliable, as he falsely claims.[170]

V. The Freer aggregation methodology is agnostic to the choice of net equity formula.

133. The Court has previously approved the formula for Net Equity per User Account.[171] I have no reason to believe that Mr. Martin incorrectly implemented that formula, nor does Mr. Dennis make such a claim.

[167] Dennis Tr., 89:19-90:7.
[168] Dennis Tr., 115:2-7.
[169] ¶¶115-119, 128-129, 133-134, 138-139, 149-162 of Dennis Rebuttal at 37-39, 42-46, 48-52.
[170] ¶70 of Dennis Rebuttal at 21-22.
[171] Supplemental Order Respecting Motion by Chapter 11 Trustee for Entry of Order Finding That Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, January 26, 2016, at 2-3.

134. Steps 1-9 of the Freer Aggregation Methodology do not use Net Equity as an input. These steps are the portion of the Freer Aggregation Methodology that determine clusters of User Accounts.[172] In Step 10, I tabulate the Total Net Equity for each cluster by summing up the Net Equity values for each User Account in that cluster. This is the only step that makes use of Net Equity in the Freer Aggregation Methodology, and the clusters of User Accounts are determined prior to this step.

135. Mr. Dennis confirms these facts:[173]

> Q. Okay. So with respect to steps one through nine, none of those steps relied upon Huron's net equity calculation; correct? …
> A. There are -- I don't believe so. I don't believe so. I believe the net equity calculation was brought on or attached at the end.

136. Even if, for whatever reason, the Court chooses to approve a different Net Equity formula per User Account than the one it previously approved, such an action will have no impact on the reliability of the Freer Aggregation Methodology. The only change required would be to re-tabulate Total Net Equity for each cluster as a sum of the new Net Equity values for each User Account.

137. In light of these facts, Mr. Dennis' arguments about Net Equity[174] are moot to the reliability of the clusters produced by the Freer Aggregation Methodology.

VI. Conclusion

138. Mr. Dennis is mistaken when he concludes that the Freer Aggregation Methodology is unreliable. Through the Freer Report, this Reply, and the implementation of the Freer

[172] Mr. Dennis acknowledges that "The ninth step in the Freer Aggregation Process was to determine the final clusters" (¶61 of Dennis Rebuttal at 18).
[173] Dennis Tr., 156:21-157:4.
[174] ¶¶35 and 36 of the Dennis Rebuttal at 10-11.

Aggregation Methodology, I have shown that the Freer Aggregation Methodology is both appropriate to the task and reliable.

139. The Freer Aggregation Methodology is based on scientifically valid reasoning that was properly applied to the facts at issue. The science is encapsulated in several fields of mathematics and computer science, including entity resolution, data matching, and probabilistic record linkage as well as graph theory, statistics, and large-scale data analysis. My academic background and work experience, as described in the Freer Report, show me to be more than qualified in these fields.

140. The aggregation technique used in the Freer Aggregation Methodology, namely probabilistic record linkage using Fellegi-Sunter with Expectation Maximization (FS-EM), can be and has been tested extensively in the last two decades. FS-EM is a standard technique for probabilistic record linkage, has been peer-reviewed and published, and there are standards controlling its use, as described by Winkler,[175] [176] Christen,[177] Enamorado et al.,[178] and others. For more details, see Section II.A.5.i of the Freer Report.[179] The use of FS-EM is appropriate in the TelexFree case.

141. Splink, the Python library for probabilistic record linkage used in my analysis, is an implementation of FS-EM that is designed to work on very large datasets, such as the 10,987,617 TelexFree User Accounts at hand. Splink is maintained and published by the UK Ministry of Justice and peer-reviewed by the UK Office for National Statistics.[180] The science behind the Splink package is also peer-reviewed and published, and the

[175] W. E. Winkler, Using the EM Algorithm for Weight Computation in the Fellegi-Sunter Model of Record Linkage, American Statistical Association, *Proceedings of the Section on Survey Research Methods,* 667-671,1988.
[176] Winkler, 1995.
[177] Christen, 2012.
[178] Enamorado, Fifield, and Imai, 2019.
[179] Freer Report at 13.
[180] "Splink: Data linkage at scale", GitHub. Retrieved from https://github.com/moj-analytical-services/splink

technique has an established error rate.[181] The continued existence and active public maintenance, by a number of contributors, of the Splink package speaks to the standards controlling its operation and to the wide acceptance in the scientific community of the underlying science. Splink is widely used, with over 5 million downloads as of 2023.[182] For more details, see Section IV.C of the Freer Report.[183]

142. The error rate for an aggregation performed using Splink can be estimated from the data set via formulae described in academic literature[184], in order to determine how frequent false positives and false negatives are. Using these formulae, I have estimated the error rates for the Freer Aggregation Methodology (false discovery rate of 1.98% and false negative rate of 0.377%).[185] These quantities imply a precision of 0.980, a recall of 0.996, and an f-measure (which synthesizes both precision and recall, hence accounting for both false positives and false negatives) of 0.988. This f-measure quantity is considered "very good" by a survey of data linkage techniques.[186]

143. As is expected of any aggregation methodology, there are some false positives and some false negatives in the output of the Freer Aggregation Methodology, but the prevalence of these false positives and false negatives appears to be consistent with the estimated error rates.

144. The input dataset is sufficiently reliable to be used in the Freer Aggregation Methodology. In particular, the name field (*rep_nome*) is of high quality, and my use of the name field in blocking and in pairwise matching is both appropriate and justified.

---

[181] Enamorado, Fifield, and Imai, 2019.
[182] R. Linacre, *Splink and the Open Source Dividend*, March 9, 2023. Retrieved from https://www.robinlinacre.com/open_source_dividend/
[183] Freer Report at 40.
[184] T. Enamorado, B. Fifield, and K. Imai, 2019.
[185] Freer Report at 74.
[186] A. Ferrante and J. Boyd. A transparent and transportable methodology for evaluating Data Linkage software. *Journal of Biomedical Informatics* 45, no. 1: 165-172, 2012.

145. The output clusters are of high quality, and the descriptions in each cluster identify its Participant.

146. The Freer Aggregation Methodology only makes use of Net Equity after these clusters are formed. It uses the Court-approved Net Equity formula to tabulate Total Net Equity for each cluster according to the Court-approved tabulation method, thereby determining the Net Winners and Net Losers.

EXHIBIT 1

Exhibit 1

MATERIALS CONSIDERED
Description
Supplemental Rebuttal Expert Report of Joshua M. Dennis, April 14, 2023
Deposition of Joshua W. Dennis, May 9, 2023
Affidavit of Stephen B. Darr in Support of Motion by Chapter 11 Trustee for Entry of Order Finding That Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, October 7, 2015
Amended Order on Motion by Chapter 11 Trustee for Entry of Order Finding That Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, December 21, 2015
Memorandum in Support of Proposed Supplemental Order Respecting Motion by Chapter 11 Trustee for Entry of Order Finding that Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, January 19, 2016
Supplemental Order Respecting Motion by Chapter 11 Trustee for Entry of Order Finding That Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, January 26, 2016
In re: Plantation Realty Trust 232 B.R. 279,282 (Bankr. D. Mass. 1999)
L. Bauer, C. Bravo-Lillo, E. Fragkaki, and W. Melicher. A comparison of users' perceptions of and willingness to use Google, Facebook, and Google+ single-sign-on functionality. In *Proceedings of the 2013 ACM Workshop on Digital Identity Management*, pp. 25-36. 2013
O. Benjelloun, H. Garcia-Molina, D. Menestrina, Q. Su, S. E. Whang, and J. Widom. Swoosh: a generic approach to entity resolution. *The VLDB Journal,* 18, 255-276, 2009.
C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
P. Christen. Data linkage: The big picture. *Harvard Data Science Review*, 1(2), 2019.
P. Christen, T. Ranbaduge, and R. Schnell. *Linking sensitive data: Methods and techniques for practical privacy-preserving information sharing.* Springer, 2020
P. Christen, *Data Matching,* Data-Centric Systems and Applications, Springer-Verlag, 2012.
V. Christophides, V. Efthymiou, T. Palpanas, G. Papadakis, and K. Stefanidis. An overview of end-to-end entity resolution for big data. *ACM Computing Surveys* (CSUR) 53, no. 6: 1-42. 2020.
T. Enamorado, B. Fifield, and K. Imai. Using a probabilistic model to assist merging of large-scale administrative records. *American Political Science Review*, 113(2), 353-371, 2019.
A. Ferrante and J. Boyd. A transparent and transportable methodology for evaluating Data Linkage software. *Journal of Biomedical Informatics* 45, no. 1: 165-172, 2012.

O. Hassanzadeh, F. Chiang, H. C. Lee, and R. J. Miller. Framework for evaluating clustering algorithms in duplicate detection. *Proceedings of VLDB '09*: 1282-1293, 2009.

R. Linacre, *The Intuition Behind the Use of Expectation Maximisation to Train Record Linkage Models,* October 14, 2022. Retrieved from https://www.robinlinacre.com/em_intuition/

R. Linacre, *Splink and the Open Source Dividend*, March 9, 2023. Retrieved from https://www.robinlinacre.com/open_source_dividend/

P. J. Owen, "TelexFree, other scams target local immigrants", *Telegram & Gazette*, Worcester, MA. June 30, 2014.

K. Qian, L. Popa, and P. Sen. Active learning for large-scale entity resolution. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1379-1388, 2017.

A. Saeedi, M. Nentwig, E. Peukert, and E. Rahm. Scalable matching and clustering of entities with FAMER. *Complex Systems Informatics and Modeling Quarterly* 16: 61-83, 2018.

J. Talburt, *Entity Resolution and Information Quality*. Elsevier, 2011.

W. E. Winkler, Using the EM Algorithm for Weight Computation in the Fellegi-Sunter Model of Record Linkage, American Statistical Association, *Proceedings of the Section on Survey Research Methods,* 667-671,1988

W. E. Winkler, Matching and Record Linkage. In *Business Survey Methods* (ed. Cox, Binder, Chinnappa, Christiansen, College, and Kot). Wiley Series in Probability and Statistics, Chapter 20, pp. 355-384, 1995.

"RFC 2822", retrieved from https://datatracker.ietf.org/doc/html/rfc2822

Qualtrics, "Sample size calculator", April 12, 2023. Retrieved from https://www.qualtrics.com/blog/calculating-sample-size/

EXHIBIT 2

Exhibit 2.1

All 12 fields from *representante* used in the Freer Aggregation Methodology, for 400 randomly-selected records chosen with fixed random seed.

rep_id	rep_nome	rep_email	rep_cel	rep_fone	rep_end	rep_numero	rep_cep	rep_login	rep_pwd_secondary	rep_cpf	rep_datanasc
19409231	ping chu	lh197109222022@163.com	15046380954	15046380954	shiyan	1	157400	jaq168e		[redacted]	None
22067481	riev samann	rievsamann@gmail.com	070215889	070215889	pouk chas	291	855	samann0001			None
29203692	Thais Medeiros Silva		689958478	689958478	Rua Jo	1485	6890000	fo53420618c702a			None
12840111	Aida luz Millan mendez	finanzas@fem-ingenieria.com.co	3175005899	3810703	Cll 12b 32A 86	32a86	760041	3183365469			None
19201259	SAYON RATANA	sophannarithhorn@gmail.com	12349973	12349973	#35E, St.378	345	855	successbiz42a7			None
32653674	petra luna	bishopjma73@yahoo.com	9565298110	9565298110	12345 main		78572	rexal29			None
32766744	xue bai	bxue2737@gmail.com	15279989002	85434321	chunxi road,chengdu	100#	610000	bx27373			None
14987623	Ali H Omar	aomar8049@gmail.com	6146320875	6146320875	12312 parkwood dr #110	12312	55337	shaareey1	575ea7987311a647196c8669b1c57b70		1970-10-05
13693377	Braulio Enrique Gavarrete Panchame	zileonela@gmail.com	50499957388	50499957388	Col. Santa Monica segunda calle bloque 2 casa 37	37	504	gionela11	45e1cc9e358bb3f36f7cbe74e5fc0eec		1966-11-10
11244297	RAUL MARCAS VALDEZ	misionsietevmt@gmail.com	944635541	944635541	MZ B LT10 AA HH NUEVA JUVENTUD	12	511	dorcasmar	611de0b10a10b9fae5f3a9bdabeb04fb		1982-04-14
31054766	Jose Sousa	duarte69sousa@gmail.com	0351962969869	0351962969869	Travessa da Paragem. Moradia D	12	9125-194	17jdnst1			None
12006479	beliza alcca quispe	beliza_love_30@hotmail.com	231488	231488	irb. ttio la florida	l-7	84	bebelonita32			None
28240802	marcelo bruno matheo calumbi	marcel21964@hotmail.com	557182094025	557182094025	avenida curralinho	03	41705155	marcelotelexba2	fb5f2f27be2de104ac2b192f3e874dda		1980-09-04
33501542	guoqi han	tfgyl679@163.com	13604507219	13604507219	bayan		151800	hgq404			None
15648265	cristian escobar	gustavoescobar6@gmail.com	6172067515	6172067515	l	l	l	ianalejandros2			None
31190562	sebastian jimenez	sebasjim_lds@hotmail.com	4169939547	905 8214181	1285 Garcia street		55828-868	sebasjim18			None
22759639	freddy vivar	ricardo_vivar_camavi@hotmail.com	593998792632	593998792632	cooperativa el triunfo	656	070650	riky0044			None
17784207	Renald Paul	renaldpaul@hotmail.com	5148935434	5148935434	6985, boul. des Laurentides, Laval,QC	6985	H7H3A1	successcall101			None
15009737	Ramon De Leon	ramondeleonc@gmail.com	18095445555	18095445555	manz 21	59	22000	ridlc7f			None
11602575	JAVIER MAXIMO LEIVA CIRIACO	javiermaximolc@gmail.com	968665387	968665387	Psje. Talara Pueblo Joven el Porvenir		15169	javierml7voip1			None
14851591	Wendson Santos Fonseca	edimaxnetwork@gmail.com	17991853555	1733431478	RUA JOÃO MATEUS DE MORAES	231	14702114	wendsonfonseca1	f1fb1ea9a408b1f321062e86ec4bae3e		1985-05-18
11002661	Lara Giane da Silva	laragianes@hotmail.com	1281015840	1288394867	R: Mariana	292	12233220	cl12laragiane	d87febe46d5519ef1874c36eacd46f2e		None
21495295	Francisco Almazan	pacohuichol@hotmail.com	6641612812	6646806455	Meseta	1342	22240	wichol4tres			None
24866050	RADDYAWATI CHAN	chanshanazara@gmail.com	60195944575	60341019493	DESA VIEW TOWER	2254	53100	richbeauty3c			None
28930870	Victoriano Mazariego	mauricioalarconmartinez@gmail.com	72007700	72007700	Tercera av Norte y Novena Calle Poniente Varrio El		00000	2i9xwxqejjbf910			None
14471807	william amaral machado	williamamaral.stgo@gmail.com	5599444546	5597232254	Rua Augusto Haller	1000	98804710	williamamaraluy	8e1ee624dc8676881fdc12cb002d8708		1991-03-29
25908762	cesar machiavello	nbruzzone31@gmail.com	099761661	47333767	larrañaga	215	50000	milces62			None
32009236	Renato Marins	renatopenedo@gmail.com	9782103090	9782103090	Po Box 3232		01930	8q5tnw4on8yon83			None
11613041	fabrizio	fabriziobonardi79@gmail.com	00393454638036	00393454638036	via rosa luxemburg 13	00393	21040	fabrizio79	749c6b091e2d50b185c593c1c679634e		1979-10-01
30520312	gutierre santana	gutcris2004@hotmail.com	9783330861	9783330861	5 elm dr	5	01749	gucall8			None
6909814	ALBERTO GUILLEN TELLO	aguillentello@gmail.com	6622569250	6622803192	CAMPECHE 270	270	83190	guillentello3			None
10895323	Filipa Andreia Rocha	fapgrocha@gmail.com	00351965550681	00351965550681	Caminho Velho da Palmeira	72	9300-304	stacecilia	b802b8d91fd69363661f43ae17fca18f		1985-12-22
18050007	Estefania Tapia	moniquelombeida@gmail.com	59322437118	59322437118	francisco chira y j paredes	115	170127	roxysteff15	4919433b44e074dceda655cf5788a76e		1990-08-12
19286159	cristhiam chamorro	skio.12@hotmail.com	3208411664	3208411664	cll 9 20-25	25	00000	pan20			None
34050578	Nor Rahiniza Abdul Ghani	rahiniza.tlxfree@gmail.com	01112455783	01112455783	C-21-1-12, Pangsapuri Sri Abadi	01112	11900	rahiniza10			None
14580551	Alexis Cabral Morel	alexandrareiscrrn@hotmail.com	0973852882	0973852882	km30	0000	0000	alexandrareis1	149815eb972b3c370dee3b89d645ae14		1993-02-09
12741791	vitor gonçalves	fontainha11@gmail.com	00351926102236	00351926102236	estrada da camacha	217	9060-010	omeufuturocl			None
14665365	Sebastiana Ururi Pumahuanca	adolforober@gmail.com	946621498	946621498	jr las palmeras		051	sebasp3			None
28414936	lkj jhg	e.kd025@gmail.com	8099109686	8099109686	kjh	mjb	kjb	80991096867			None
14092857	Henrique Urias Santos	uriashenrique@gmail.com	558100612647	558199612647	Dona Marina dechamps	62	54340095	urias2015	734c085fb7ebb2b95b7e0e82d452d837		1961-07-15
23229189	cristian mamani coaquira	kennedymamanic@gmail.com	952917617	952823647	agrupacion ciudad nueva Mz 19 Lt 06	a 3	23000	2503sa7			None
26588988	Andrés Saavedra	andrex1963@gmail.com	661829310	661829310	Plaza Arturo Piera	7	46018	gpsiter1126			None
22825309	gaspar condori cruz	wcmgaspar@gmail.com	954729758	954729758	av. antisuyo	c-11	084	joaquinvoip22			None
10481737	Marcio de Oliveira Gomes	marciogomesoline@uol.com.br	556699542801	556699542801	Main ST	320	02740	mgomesgroupcl			None
31485092	meiying gao	gaomeiying1969@gmail.com	15960133009	15960133009	zhangzhou		363000	gmq042	04973d1800f7c29d2b5daccdb01f31a7		1969-05-15
22779907	longshen chen	3589447775@qq.com	8613589447775	8613589447775	huangheroad	5	515000	njmy001009			None
15269983	adriana menegati	jozineiusa2012@gmail.com	2672352798	2672352798	6237 crafton st	6237	19149	neydrica7	2ff287f725a4e9d9d7c04b7a819c8f29		1974-11-02
22442207	da liu	lx51876@163.com	155770888245	155770888245	gui tang		537499	lx51125			None
22172165	maria villegas	maricita_aqp@hotmail.com	0051959369858	0051316021	mno.bustamante	j-2	051	23majolivi01			None
24603072	Andy Asian	myaperu@hotmail.com	981461007	981461007	Av. Camino Real Nº 476		001	myaperu3voip8			None
17420867	lourdes zambrano marcillo	lourdesosita@outlook.com	593984999360	s/n	condor ñan	205	0000	1982pao			None
28806740	soeun savry	soeunsavry7@gmail.com	85516957695	85516957695	phnom penh		855	lucky4phone5			None
24990458	jefferson herrera	mnghghg@hotmail.com	6466664	464646656	gjgugygyg	15465	jghgjhj	alexpaqu3			None
25827594	carlos santos	cerqueirafree@gmail.com	966380983	966380983	rua airo	21	4730	cartania710			None
15948487	Enrique Torello	enriquet444@hotmail.com	+34627046884	Ybañez	Trías de Bes	73/77	08440	kikebcn	2f4e613b3796bb305942e81621f1f088		1965-02-09
15622367	denisson carvalho fontes	denissonnet@gmail.com	557999665426	557999665426	capitol ave	1603	82001	dcf	03435e70b2d275b16191f12f3b41df2f		1986-04-24
23128273	gilberto guayamis galoc	guayamisg@gmail.com	942845624	942845624	jr huallaga 434		001	guaytas4			None
13274413	Newton cruz	newton.cruz7@gmail.com	8134439434	8134439434	rua do triunfo		52120-090	newtonusa772	42aa6053bc606f9645d3832265e9f548		None

29027608	César Magno Abreu Castro	autoferraztelexfree@gmail.com	00351912121505	00351912121505	urbanização vista alegre		9125-902	autof38		[redacted]	None
30602924	William Silva	wcsjunior@yahoo.com	5108606309	5108606309	104 villa drive		94806	newmoneywill7	800e365301349f2b62fc5bcbf6ac2b39		1982-06-19
19566151	Gentil Fernandes da Silva	isaiasleite2010@hotmail.com	559492817557	559492817557	AV. Duque de Caxias	226	68520000	gfs09			None
13310573	Christian Baca Silva	nancysilhua@gmail.com	995795449	3863123	conjunto residencial campoy block l 515	515	LIMA 36	christian8943	1106ef9ecd422a4e62912c68ac8d2df0		1989-02-02
8990114	Julia Keko Aki	akijuh8@gmail.com	09064631070	09064631070	Kunimoto2254-2	2254-	4370012	akijuh13c4			None
27968580	Ho Lai Fong	poloamon@gmail.com	85226714333	85265996128	Flat E,11/F King Win Fty. Bldg.,	00852	00000	mags0003	ade7c1dc609408b28708a9d241f17889		1969-08-29
30347198	Salomé Chagas	salomechagas16@gmail.com	351261937300	351261937300	Rua do Juncal	16	2520-042	xagas7109			None
15424181	Francia Feliz	francia_1303@hotmail.com	8294326954	8095964381	Av. españa edif 23	23	809	franciacliente			None
16362967	Eduardo Ferreira Junior	ebpmultnive1@gmail.com	5584-88672797	5584-20106105	Av. dos Caiapós	2025	66835	compostela20	802d8c9b92946f3e6b368312cd1383e1		1975-05-30
17655023	abc abc	a0925147655@gmail.com	1111111111	1111111111	22222222222222222222222222		70000	hkt0001			None
12839407	ADAUTO GOMES PINHEIRO NETO	adautopinheiro@hotmail.it	351913958066	351913958066	RUA DA ESPERANÇA	2	2500155	walfigueiredo			None
17150481	Erik Damian Munoz Monserrate	mercedesrocio@hotmail.es	593994734538	59323741018	Carlota Jaramillo	13	EC230101	taguaexitoso02	3d863b367aa379f71c7afc0c9cdca41d		1996-01-26
18285155	Carlos Rodrigo Jacome Ponce	carlos.jacome.p@gmail.com	0984619495	3310256	Antonio Sanchez s/n y Francisco Ecija	S/N	170137	cjacome1	235b567112edea1e02589e9a734126fb		1979-04-25
15244391	CARLOS E. GOMES	cecgomes20@comcast.net	9548512332	9548512332	2912 NW 67TH COURT		33309	cecgomes016000			None
15804223	Betzabe Laureano Lazaro	jehovajhire001@gmail.com	51968084164	511968084164	Av. Universitaria Km	5	0062	betty567voip1			None
22593609	Jose Aparecido de Andrade	joseaparecidodeandrade4@gmail.com	556596038070	556596038070	Rua Marcelo Terceros Banzer	2	304	javitorias8cl6			None
15058255	hassan abdi karim	shaafici1@gmail.com	441212706722	441212706722	26 stirling Road	26	b169bg	shafici4	8ce87b8ec346ff4c80635f667d1592ae		1974-08-24
26719794	Natalina Aliprandi	natalialiprandi@hotmail.com	393206565660	393206565660	via Degli Stretti	07	31100	raffa1209			None
24596252	Nélio Ferreira	nelioferreiranegocios@gmail.com	914264638	914264638	Rua Barreira Vermelha nº2A		2510-049	bvp059			None
1837442	danimari llc	soporte@myonecallxpress.com	7043607312	7043607312	1918 coliseum drive	1918	23666	c17043607312			None
18715363	Paula Prak	paulaprak@gmail.com	3016930631	3016930631	1818 Granby Way		21702	Luckyspirits01	b918cc34caecf0777981414c38908144		1981-02-28
24762910	yolanda landaverde	osorioleslie23@gmail.com	6179554744	6179554744	133 irvin st		02149	abregos7s			None
26175142	oupalavanh bouaphet	oupalavanh@gmail.com	0781093781	0781093781	2º REP CAMP RAFFALLI CCL		20260	bouaphet	9551f54626799ab0f93c314eaadaa1c0		1983-02-04
12486309	keqing qin	13666303168@163.com	18660377988	18660377988	weihai		264200	qkq598			None
31188226	Vinicius Wetter voip	viniciuswetter@hotmail.com	46874654845	54546545645	joseverissimo	16	35053340	lucypoc4398036			None
18542715	carlos perez	carlos_perez79@hotmail.com	573214570978	5718824435	cra 4 este 30 21		0000	8006607713			None
18623951	julio cesar	julio.cesar.prado.jcp@gmail.com	86344407	86344407	rua senador pinheiro ramos	786	54753110	juliocc5			None
29554222	sandy aguirre	sandyta1823@yahoo.com	6174130243	6174130243	9 walden		02151	3santato1			None
25618846	ds fd	zxy13581932995@126.com	13891278831	12548453486	shanxi		719000	fdbs			None
31051852	Segundo Gregorio Celi Guerrero	jessyestefany_1997@hotmail.com	593990413417	593990367147	via a la loma	xxxx	070650	comandante0016			None
19582027	guohua chen	cgh130121@126.com	13081029083	13081029083	beijing		100000	cgh114a	381ee61c5844233b757c8dbff6ba975d		1968-05-06
19042495	hgcvuygihop vjhbiuhojp	sinlimitesavanza@gmail.com	987564	786543	fdtyfug	69785	78965	ag928			None
22425649	wreew sfasd	rert@163.com	ewrewtre	trruyt	truyuyt	gkjhj	tyuiuy	bafrm222c2			None
20635873	Armando V. Sousa	armandotelex@gmail.com	5088603665	5088603665	66 Harrington Farms		01545	avsc9			None
24824544	Eduard Pfau	agatazxc289@gmail.com	+77015261730	+77015261730	Stepnoy4	12	100000	drandpool5	78c4318cd553838269207ece540b906f		1976-07-19
19176359	Abel Ernesto Chacon Chacon	drfranciscog1962@gmail.com	593998812240	59322699199	Av. Tnte Hugo Ortiz y Moro Moro	S/N	170105	aech09			None
20976333	humberto vladimiro suarez rios	hsuarezrios@yahoo.com	974219466	974219466	cuatro torres	407	084	milingo3			None
17440663	Pavel Danilin	pashadan86@gmail.com	+79170638522	+79170638522	40-letia Pobedy,8	339	432072	pasha358	a08e91f1a15579e233ce4854ed66efe2		1986-07-13
23586116	Evaldo cunha	evaldoatlanta@gmail.com	6788552995	6788552995	2313 cajun dr		30066	feraatlanta46c		..	None
15515561	van prom	promchiro@aol.com	2096145465	2096145465	1031 Mchenry ave		95350	168peterc1			None
28514922	Mario David Montiel Lopez	lopemar@hotmail.com	983428883	61571428	calle acosta ñu	37	7000	loperro52			None
30224954	shijie da	13563087333@qq.com	8613563087333	8613563087333	jiqirod kuansahn	18	250023	lbf1623013			None
23996308	Mapril Pena	maprilpena@gmail.com	961265353	961265353	estrada nacional 360	71	2500-609	vopiu103			None
33605882	Jose Teixeira	hessotelexfree@gmail.com	447876558572	447876558572	Bond Streeet	63	10	ynwcu4tfwl1rr5i			None
25713568	cesar augusto rodrigues aguero	carodrigues123@gmail.com	34722317313	34981634675	a barreira	17	15172	1topcc1			None
10849613	Percy Choquehuanca Mendoza	percy.1972@hotmail.com	951409268	326127	Jr. Roberto Barriga	772	051	percychm	680f53fc72a58404c6e6256c48bf753a		1972-01-15
16774341	Egidio Fulberti	squadratelex@gmail.com	3343158136	3343158136	Via Strada Ciosa	41	18100	fulberti47	3e106f801582b714ab41d5fa18e00b87		1947-10-26
33812486	Helena Mantese	bambola66@gmail.com	4156374504	4156374504	540 D ST Apt 9		94901	14hm7v3			None
18527109	Eduardo Alves	eduardo_alves4@hotmail.com	42323871932	2712389132	Rua Guarani	200	29052303	gruporedevix	a3eda7c758497f437c166ab6f568983b		1987-12-06
32755544	diego mavalo	diegomavalo@hotmail.com	459898477896	459846229385	Juana de Oriol	656	40000	mavalo10			None
23437753	Ala Khamiankova	a.belikov.staff@gmail.com	+375257843934	+375257843934	Lavrinovicha	1-22	247351	freenight2cl1			None
30467382	angel suazo	angelexousia@hotmail.com	33395926	31920172	RES EL DORAL	CASA	504	mybless81voip6			None
28381230	Isabel Freitas	isabel.freitass@sapo.pt	961063303	961063303	Caminho dos Saltos	1-B	9050-219	lebazi3p6			None
27794690	Guilherme Costa	telexfree.moneynet@hotmail.com	19783023434	19783023434	07 home st	07	01749	solac	c54f739fe53a1f36ae7c15b21e6bda85		1992-02-22
30912060	mingzhong wang	yangzhiwen86@163.com	13083626044	13083626044	henanluoyang		471000	kpqq555			None
11173547	monaco giovanni	pino.varratta@libero.it	3454911746	3454911746	via lequile	178/c	73100	pinovar	71d6c631e36a3432a5899bb09c51d52a		1962-01-06
23507474	gui long yang	colin920605@gmail.com	15811916535	15811916535	gaungdong		516000	ygl666a8			None
32798142	zhi zou	13487591499@qq.com	8613487591499	8613487591499	beisan road	3	610081	zouzhi004			None
12029215	janser rodriguez	janserrodriguez@gmail.com	8097107067	8097107067	carretera mella 102	102	23000	janserrodriguez			None
16370859	vanesa zhagui	luis.suculanda@yahoo.com	6313352714	6313352714	jarison	2	11937	vanesas05			None
31632038	Ana Luís	sofia_luis5@hotmail.com	965135327	965135327	Loural	0	9240-216	asgluis35			None
14462651	Genner Bolivar Puello	gennerbolivar@gmail.com	593994536919	593994536919	Avenida 17 de julio # 8-65	8-65	00000	preciosom15			None
15890279	carlina meza	hucajoca@hotmail.com	3125192104	3125192104	calle6 No 4AE 20	4AE 2	54001	telexcameca			None
27588896	Manuel Alves	mmalves.telexfree@gmail.com	00351912301180	00351224670384	Rua do Barreiro	43	4485777	maa1voip2			None
22549637	Angelica dasilva	brazilian.angel21@hotmail.com	7324060098	7324060098	westside av.	653	08861	a331cli			None
13024661	eris mejia mejia	jrjr0320@hotmail.com	19788769629	19788769629	14 bruce st	14	01841	eris1234			None
18874971	shou dong wang	ccl15367418087@gmail.com	13311883817	13311883817	shenlanlu	23	518000	wsd1681	838ef0e61e11ca0469110b10d157ca81		1973-02-08
27858064	UGT ghj un GF tgjf	jossandra1993@hotmail.com	VB CF y DG u d	GN g UF DCC gg	fdhfzfgigjuggf	VB g	23000	1234vi1			None

30736220	fabio bomfim	fabioguimaraesbomfim@hotmail.com	968708635	968708635	rua 1 de maio	11	2625-622	fgb3voip7		[redacted]	None
12998949	Jorge Velasco	alfredo_vlsc@yahoo.com	5593334957	5593334957	586 W Stacie Avenue		93257	elpalomo	2ad5495bb5022419d894270f829b5125		None
29198424	Antonio Beitum Beitum	claudiociuffa@hotmail.com	554399352030	554399378003	Wood Dale Rd	468 N	60191	niko1955eua8			None
26020118	yi luo	royal0816@163.com	13761279502	13761279502	shang hai	2	200940	ly5009			None
30169436	mengqing ma	taochen75@gmail.com	8058869666	8058869666	2459 jacaranda dr	2459	93036	honey09a			None
25409672	jesus alfonso aguirre hernandez	japroservi@hotmail.com	9931679450	9931679450	av. revolucion	633	86150	jaah84			None
22553765	abraham laura pacco	franklin_444_5@hotmail.com	973505754	973505754	Jr. jose antonio encinas	S/N	051	abrahaml001			None
27253936	Luis Luna	alizabelpg@hotmail.com	8498629761	8498629761	Calle 10		10100	8497012235f			None
16885985	Cosme Rodrigues	cosmemh@icloud.com	15089332888	15089332888	13 Burkeley St		02914	cosmemh30	4c91dd5327d6d3867566013261ac73a6		1976-10-12
13390271	Rene Silva de Oliveira	1988renesilva@gmail.com	8487482872	8487482872	Rua Dom Antônio de Almeida Lustosa	132	59073110	renenatal5	3ebcf2ea126331c42b12da5abb3da6b0		1988-02-17
17760751	Otilio Prado	suporteplanobbc@gmail.com	553199844030	553199844030	.	.	.	bhinvest114			None
12309161	ana paula	nina.bbom@yahoo.com.br	5519982686928	5519982686928	calle del maestro c cero cora	s/n	7800	larapy5			None
22878987	salustia perez	salusperez@hotmail.com	2274822368	2274822368	calle miguel negrete	sn	74180	os8			None
25579732	Ricardo Kazuhiro Ikematsu	r.ikematsu@gmail.com	08054318685	08054318685	yorikido	1376-	3700535	rkike7voip5			None
28043426	jesus condori gutierrez	maxsacogu@gmail.com	986866244	986866244	micaela bastidas	111	051	maxsavoip45			None
32366660	GEOVANNY PELAEZ	geo.-9@hotmail.com	34676362107	34676362107	COSTA DE SOL	3	28033	ffonos8			None
22325361	Ageu Divino Borges	vencedorusa@gmail.com	6599883386	6599883386	Rua 405 Qda 124	34	70088611	borges07voip8			None
26462494	rekiya salama	sofiefashion@gmail.com	08032092699	08032092699	Suit 3, Mammy, Lungi Barracks	3	234574	hajiya			None
20302025	ernestina tavarez	brayanvas00@hotmail.com	8293923776	8097368776	Gurabo Calle 20	68	51000	entavarez9			None
28082806	claudia chaves	duarte.jose93@gmail.com	15084105696	15084730554	14 shadowbrook lane	47	017571144	lh0rb1ymfondifx			None
12230337	Kadi Ka	wbeyanca@gmail.com	45667	45677	56799	00100	00254	kash1			None
33835388	adalia Buendia	michael_1803@hotmail.com	00593988161297	00593988161297	coop. santa Eufracia Mz. D , Solar 1		090610	adaco31			None
17133111	Robert Rodricks	robertrodricks@gmail.com	3013177872	3013177872	3356,Old Line Avenue		20724	robert7878a	00c66aaf5f2c3f49946f15c1ad2ea0d3		1954-06-19
19446541	Benedicto Carreño Castro	benecartage@hotmail.com	0573173757607	0575133574	Carrera 47 # 55-30	47	0000	benedi99			None
10924917	blessing adat umo	eganndem@yahoo.com	08032691565	08032691565	wilbahi estate	56	23409	greenleadafrik6			None
28622356	Ionel Iacob	elenatirnovanu@gmail.com	3334357995	3334357995	via pinelli	6	10100	gyanni	263bce650e68ab4e23f28263760b9fa5		1976-12-01
23939388	bruno silva	brunoglleal@globo.com	5516981711961	551636302223	antonio josé moreira	125	14055320	burusac1			None
16650765	Mohammad Muhsen	miagee76@gmail.com	0119627971970 51	0119627971970 51	44 alhekma st	.	.	saifs2	e489562a6c8de9769278c6cee58c47e3		1978-08-24
8233698	graziana	grrrace@yahoo.com	3331398595	3331398595	via bartali	1	58100	diegof114			None
13310459	Ruben diaz	rdb94@hotmail.com	7025040232	7025040232	pobox 570372	123	89157	cubalibre38			None
28683680	Angeles Alvarez	lauralopezalvarez@gmail.com	691684848	691684848	C/ Sant Jaume	8	08980	alvarez53c2			None
23055097	lucinda maria de paula neta	lu_depaula1@hotmail.com	78191874	78191874	calle san nicolas	22	15	lubolfone7			None
31382076	eduardo alvarez morales	azakatsuki1515@gmail.com	51999080254	51999080254	Jr Jose de San Martin 150		01	cel2voip08			None
31809602	wellington de jesus pereira	alexandrausa2013@gmail.com	7543675913	7543675913	3550 blve dr#103		33064	07wjp303			None
12326283	Maria Lorena Castillo Aveiga	doctornico71@hotmail.com	59322504150	593999831134	Carvajal 195 y la Gasca	bloq	170143	lorecast	5e8667a439c68f5145dd2fcbecf02209		1980-07-22
29145636	juan carlos buitrago	juank1875@gmail.com	3208431196	3208431196	cll 20 6 53	0	057	eser8			None
24326730	cece cece	idultelexfree@gmail.com	081355203254	081355203254	tarakan	11	4121	cece007			None
17617649	marino belen	tintasbm@yahoo.com	8096133626	8096133626	Manzana 25 No.5 el eden	5	11201	marinobelen	ca6949e70eabe8adbc8632697f6b988f		1980-12-18
28510658	gilma amaya	jcedillor@gmail.com	99072736	99072736	sps	1	504	gamaya827			None
15997647	Alexis Rekeraho	alexisbwana@gmail.com	250788590092	250788590092	Kigali		250	iranzi6	c0bb9bbaf56198988ba3a662ff7450d3		1975-01-01
14615005	enia valentin	epvalentin@gmail.com	17173302810	1713302810	185 foal court	185	17602	valeusa100	3f1ab41efddee120ba34f197705d1a01		1975-07-10
24754476	Grace Semugenze	gnkalubo@gmail.com	0772436082	0772436082	P.O. Box 21940 Kampala	256	25641	graceofchrist3e			None
14186869	wendy pichardo	wlpmnv@yahoo.com	6177784040	6177784040	odonnell lane	6	02026	wlp09	56eb2fed43098c5f44e1a37ea007251b		1976-06-23
25681294	licinio1 amaral1	licinioamaral@gmail.com	9991310303	2732440203	Rua x1	1231	02411310	teamarraza13103			None
32799100	jing guo	gg921jj@163.com	18511060629	18511060629	shijiazhuang		050000	rry022a2			None
23746160	ZENON QUISPE CHOQUEVILCA	zenon.telexfree66@gmail.com	979232336	979232336	CUSCO - CANCHIS - SAN PEDRO	S/N	084	06zenon			None
17205639	Jose Hernandez	joguiro@yahoo.com	(829)639 4590	(809)472-0404	Av. Sarasota 20	20	11207	joguiro	5a01f0597ac4bdf35c24846734ee9a76		1967-09-09
13283601	Dl1 Inc	davidbeeba@gmail.com	8573122571	7816200168	97 bellevue avenue	.	02176	ldh43	1849c91d7b856fa80fb45b4130d6fd3b		2013-01-28
21375225	jose antonio callapiña condori	jositan12@gmail.com	989649964	984264269	jositan12@gmail.com	572	084	vace1			None
21603193	rundong zhao	12345@qq.com	1310000	1310000	tiedong	1231	114000	egs010			None
30478712	jose rodrigues	duartesousa79@gmail.com	351967043776	351967043776	Estrada do Pedregal	1	9350-033	2sousafn101			None
17366803	erica mazo	Ericamazo1021@gmail.com	7813001513	..	878 broadway		02150	valexa09	5f14615696649541a025d3d0f8e0447f		1973-10-23
32159622	aylaiheang07 aylaiheang07	yemren@live.com	8559294123	8559294123	battambang	5	855	heang567418			None
3422538	aubri salvador	telexfree@telexfree.com	+51 1 2642019	+51 1 2642019	Av Juan Pezet 1765 - San Isidro	17765	3788	aubrisalvador_tlf	92f632fea36a4ddaebe79564dbe6029f		1978-01-01
21995387	guibin huang	hgb19770329@163.com	13209496565	13209496565	laizhou		261426	s2e			None
18301429	antonio rudrigues	school_69_6@hotmail.com	351967876352	351967876352	america	1	21345	clientejsa03			None
24648950	fernando carreño	fmorenodaza@yahoo.es	573014942381	577949516	cll 152c 133-09	152	0000	pachamar15			None
15994557	thabit salim	thabit.salim@yahoo.com	0701396396	0701396396	kawanda	02	256	c21paluka4			None
23991518	Jason Azevedo de Farias	jason.farias@hotmail.com	556892298749	556892298749	Rua 13 de maio	06	69900205	jason1f2			None
29340744	henry kiddu	henrykiddu6911@gmail.com	7813304744	7813304744	26 beacon st		01803	c2csudk7ixrs935			None
15187881	marco fernandes	markffaria@hotmail.com	3194066658	3134278516	rua clarissas	680	31730725	voipmarcoplus	ec038f4aef0cdd8eb081faf735e1ab89		None
16190613	mercedes garcia	m.luis.madera@hotmail.com	8092585655	8095322990	calle azua	9	10147	anglada9			None
13187479	Raysa Marleny Leonardo Olaverria	rayzamarleny@hotmail.com	8299046324	8095566891	calle bermudez #128	60	22000	raysa0806	d5e29feb654a6681f1e29c3ff36a3cfa		1993-03-08
28704006	Paulo De Freitas	paulorjfreitas@gmail.com	+351967468465	+351967468465	Rua Nova Quinta Deão Edif. Cuibem 1 2º E		9050-071	gtt308pro2			None
13708471	carmen elena alban orozco	carmenelenaalban@gmail.com	3176165383	8225702	calle3 No. 17 a 16	3	00000	carmenelena1	59e020e1b52110 6aec6cdaf41f5b9ee4		1989-08-08
30328944	Elias Bonfin Amaral	dkfigueiredo@gmail.com	19-34526600	19-34526600	Avenida Laranjeiras	1322	13480000	bonfin205voip			None
29329960	bai sun	hjsgf@163.com	18765343728	18734526783	jilin		136000	sbl0208			None
23779744	cliff monique	godie.banjo@gmail.com	414754245741	0782043285	entebbe	256	256	bann8			None

20236581	daniela meneses	lydamurcia@hotmail.com	3145825126	3145825126	calle 14 #11_44		19003	3145825125		[redacted]	None
12253845	Luis Fernandes	Luisfernandes-87@hotmail.com	964851693	964851693	Caminho dos Tis	24	9325-172	luisfernandes5	5ff2f237f9e624088563a70efe2528ca		1987-03-25
20613851	Gisselle Ricca	gisselle rica@hotmail.com	091211606	091211606	Juan Paullier	2446	11800	12redgigi8			None
11010505	EDIMAX NETWORK CORPORATION	lidertelexfree@gmail.com	9546874023	9546874023	1310 PARTRIDGE CLOSE	1310	33064	edimax56	f1fb1ea9a408b1f321062e86ec4bae3e		1962-08-26
30114492	ANDERSON MURILLO	CHARK100@GMAIL.COM	573108169326	5712893264	cr72 22d - 54	72	11001000	KIRIN02	eaca99ab01438061f06b232dde801fc5		1983-08-19
21676513	cesar lozano	aricar22@hotmail.com	3108634173	3061012	calle 4b # 53 f 19		00000	cesaraugusto298			None
17353657	Juan Evangelista Villalvir Martinez	jvillalvir@hotmail.com	+50499333582	+50499333582	Col. Sandoval sorto	10	504	jvillalvir01	62e201f3d00d21ca737fde0e4ed9b545		1987-07-28
15360707	Elen white Magalhaes lopes	gilneyw@hotmail.com	07388347867	07388347867	Rua Helio Campo Dallorto	25	45995000	voipwhitee			None
24717556	hafizs mustaffa	hafiz92@yahoo.com	0148299303	0148299303	lot 55 D kg teluk baru	55	26040	hafizs10			None
26152056	marianna porcelli	porcellimarianna@gmail.com	00393204811016	00393204811016	strada del salice vecchio km 4	48	71122	marianna36			None
33561636	Paulo Beloni	paulobeloni_foz@hotmail.com	554599646613	554535243184	rua macaé	394	85868150	azebeloni3			None
30812884	jose rodriguez	grupogalicia2014@gmail.com	0034629412629	00982545193	castrolazaro-portomarin	2	15011	mzc0001			None
23105921	M Itv	marullo_05@hotmail.com	Exceden	U tvt	NBA	Ih y	23000	gegre7			None
13042245	hongyan li	zrr8695@gmail.com	18050898223	18050898223	henan china		400001	lhy5882	6649c7f2602777662f075f0d14d84b35		1976-05-09
30422180	ivan fernandes	navivan54@hotmail.com	5615063846	5615063846	647 brittany n		33446	lnd2jgapftbcm5s			None
15816073	Edith Machaca cruz	edith_shande@hotmail.com	0059179454547	0059179454547	Potosi	118	00591	cnoe1			None
17690393	ANA MALHEIRO	drummond.a.telex@gmail.com	351925534038	351925534038	CAMINHO DAS COURELAS	40	9020080	2aikidoc			None
15611301	Edgar Rodriguez	e.rodriguez.joa@hotmail.com	8095654857	8095654857	Repardo 13 #17		00000	edgardaniel02	68d47877655c857fedb09e7b7afe7af0		1987-09-22
10859665	Rafael Frias	friasr@gmail.com	(809) 964-0979	(809) 964-0979	Calle 1ra, Res Don Jose II	32	10100	charmed17			None
27574160	ma vena	bluevisal@ymail.com	8559898	8559898	#12,st2	12	855	maake8			None
31030906	OSCAR MORA	reglamide@gmail.com	3165890172	3165890172	CALLE 1 # 12-58	1	00000	color115			None
31122188	Wen zhi	telexapp8@gmail.com	17184382292	17184382292	671 47 street,3a		11220	n10c9			None
12101361	MARÍA DEL PILAR CASTELLANOS PUENTES	mapi2803@gmail.com	3204950505	8028806	Calle 72B No. 70F-45	72	00000	mapy02			None
20162527	sergey daerty	rhcnbmzg@snapunit.com	523524246	523524246	rue degrennb	51 b	001111	ermos4			None
30765762	hortencia torres j	hortencia1263@gmail.com	7705497709	7705497709	454 parkwood way	454	30236	horto037			None
19367271	Evgeniya Lebedeva	evaeva01012014@gmail.com	79128701598	79128701598	Avtozavodskaya 68	53	426068	evalika22			None
15631715	Carmen Duran	carmenduran9@yahoo.com	8299161605	8299161605	carretera veron punta cana	14	23000	angelia01	043db66e8d0fc4512a0bc9f840f1554d		1975-11-09
12373697	Bayron Emilio Garcia Amaya	moshe107@hotmail.com	50495060089	50495060089	1 calle, 3 ave. Bo. Medina No.22	22	504	hechos8	58c8221746f2c81a450c45f48adec438		1971-08-05
28473488	hugo mamani choquehuayta	edy.dance@gmail.com	984999180	984999180	comunidad sencca chectuyoc		084	hug005			None
32518998	Jose Lucero	ljose7194@gmail.com	0959230639	593-2955016	la primavera	xxx	070650	kacr1			None
15359791	alba y Castro	alba.castro95@yahoo.com	6172704232	6172704232	258 vine st	2L	02149	albacastro14	a947d4cd95e5d0c30f0420b21bca3b27		1978-07-12
17748067	Francis Kinyua	anmur2012@gmail.com	5086151681	5086151681	115 Brookline St	1	01603	nkirote110			None
29016170	roseli felix	roseeffelix@gmail.com	5511992851262	551120745126	Rua das boas noites	779	08225180	rosefel4a			None
17342739	Carla Regina Rodrigues da Silva Sueth	viniciussuethtelexfree@gmail.com	17992179898	17992179898	Josefina Reis Assunção	31	15047147	carlacsueth7			None
15018283	Mark Vang	yumvang@yahoo.com	6124082185	6124082185	1383 Summer Ave	1383	55109	richmarkvangb			None
16513577	1 1	songwenjie0012@gmail.com	11111111111	11111111111	xian	20	710000	swj			None
11940641	Dustin Leavitt	duster.1@cox.net	4806888158	4806888158	3764 East Bart Street		85295	lduster5	daffe63d7b852b29209b0ecb668c8525		1957-05-09
33696458	melissa alvarez velasquez	irvins_gsv@hotmail.com	943958087	943958087	c		01	rosette404			None
16881825	andres ocampo posada	andresocaposada@hotmail.com	573114608493	5714832952	Cra 136A 145 30 Casa 36	136	0000	loren0901	a431d70133ef6cf688bc4f6093922b48		1973-12-30
32325898	Sandro Scamporlino	sandroscamporlino@gmail.com	3495652865	3495652865	Viale Tito Labieno 70		00174	samaem24	81e3b96f6932b28a4074797685dbd8eb		1967-09-06
17641357	xiaohong duan	duanxiaohong1919@126.com	13937805296	13937805296	henan		475000	dxh06	d0970714757783e6cf17b26fb8e2298f		1967-01-16
32410022	manoel cleto cordeiro	cletocor@bol.com.br	558796834374	558796834374	rua luis epaminondas	324	56640000	movini7v1			None
12599305	wertyuiop	jonaula@gmail.com	dfghjkghjk	23456789	wertyuio	qweri	234567	wilsonr691			None
31740540	Guu Xang	kouvuepc@gmail.com	9	9	9		95828	pabkuv31			None
13535571	Antonio Fidalgo	lfidalgo_28@yahoo.com	617 386 9689	617 386 9689	220 Union Street	220	02368	afidalgo71	2f8339d149a1ea2e112e47d9e29e74d4		None
18662359	Marie Fanny Desanges	claireimage2000@yahho.com	6176987267	6176987267	Livermore	102	02126	fannydes007			None
26081656	fabiana Leite Ribeiro	clauass12@gmail.com	84 96071106	84 96071106	r: Rio Solimões	7699	59068360	frex26174c3			None
11139601	paulina andrea arias perez	paulina.arias.perez@gmail.com	965456922	965456922	las palmas 3381 depto 4 lan-A Hualpen	3381	4260257	pauliariasp	58ca638d0bc81650ed425b23fa0a0567		1985-11-07
30957314	jose silva	marcokeiros@hotmail.com	1	1	tra	23	4505-505	carlosalegro01			None
17056207	Clara Esquivel	clarita1850@gmail.com	573134208892	5714936899	carrera 25	46 a	00000	mayola199			None
20587585	cassiano marko	cassianomarko@yahoo.com.br	4632322806	4632322806	Av. Generoso Marques, 1562		85550-000	inte5			None
18338867	milda weck	ilse1073@hotmail.com	3214210768	8627357	cra 42 No. 18 a 08	301	41000100	casacambulos3			None
22625973	ramona apaza apaza	zacarias.raq@gmail.com	956545460	956545460	jr. miraflores	s/n	051	erikakaty6			None
13480503	jose escolero	elfenix2280@hotmail.com	8602092045	8602092045	21 silver lane c1		06118	escolero2			None
13673021	Palmira Diniz	palmira.diniz@gmail.com	+3518970862	+351968970862	Rua do Damasqueiro	3ºI	9325-024	sanlis2	509c5eca2b6523ce3719ddb1c2ab883d		1983-10-23
5342132	fabio wanzeler	almeganetwork@hotmail.com	087088	08u087	23425435	333	33073	10voiccer			None
20923117	Ramon Tice	tice.telexfree@gmail.com	8293759656	8092336874	Av. francia	18	51000	8434			None
15763669	Rodrigo DeJesus	rodrigo.realtor@gmail.com	7742791900	7742791900	437 Captain Eames Circle		01721	dgbr0501	e83e154707cd1a7f351a5ac05026ad8e		1980-05-01
27179502	jislaineborges borges	jislaineborges@live.com	3399273139	3399273139	taunton ave	88	02914	jishb63c			None
11124597	hongliang gong	zhbllq1@163.com	13704518112	13704518112	nangang		150001	ghl15			None
25300804	Adriana Fuzikawa	dripaloma73@gmail.com	08095398656	08055091804	yura cho 1620 10		3730036	dripa37			None
24409536	Maria Soledad Chamy Roman	schamy@axbridgegold.com	51959345479	51959345479	Calle Las Gardenias 232 dpto. 401 Valle Hermoso	401	01	schamyp145			None
26182512	pedro kilo	kilokilo@gmail.com	456457	5646	kiev	22	43556	kolik3			None
15273029	shoujie han	hanshoujie7987@163.com	13945567987	13945567987	lanxi		151500	hsj1948	ec5bac90e1f795e4f0ef8d140e24e34a		1948-03-06
18932255	miriam cusi sucasaire	cabezon.tibilon@gmail.com	976107701	976107701	jr. miraflores	905	051	jesusmamani6			None
13693843	b b	barbara_radloff@hotmail.com	1111111111	1111111111	b		48471	bbbb11			None
30573942	000000 0o0	bancofinan@gmail.com	00	00	bancofinan@gmail.com	ll	00	royh8			None

28738876	sgs sdgsdv	iachimvitalik3@gmail.com	34342324	234141	adfaf	45	142352	anka9999		[redacted]	None
23966676	jonathan yuri alves santiago	yurisantiagotelexfree@gmail.com	553175983820	553136240711	rua doze	172	33822770	voipsantiago7			None
22855173	Vinicius Lima	wetferg@tubaraomail.com	2814274456	2814274456	4392 Candlelight Drive	2370	77520	turbod2lwxe0			None
28905288	Gladys Alexander	sylvom@yahoo.com	08037037160	08037037160	Zone 11 gariki Abuja	11	0009	omonus43			None
33056690	8 po	abc123@hotmail.com	0934561789	0934561789	1	112	334	malidy4c8			None
21767977	vanda maria almeida	vanda.almeida@outlook.com	559899935406	559899935406	rua da felicidade	86	6500000	valmeidavoip07			None
19839823	ester evangelista de matos	ester.ev-2006@hotmail.com	6984593205	6984593205	rua:das turmalinas	1598	76875828	esterfone09			None
12264801	Francisco Javier Vergara Redondo	javiertelexfree@hotmail.com	34646090274	34646090274	Alfonso XIII	18.	21002	bjaviervg6			None
12083907	Lisset Marcela Giraldo Lujan	lissethmarcela6@hotmail.com	573006918601	5745969874	carrera 52 B 75 50 apto 210	52	5001000	0511lisslujan87			None
33247320	rui martins	joaocamacho2011@hotmail.com	00351968343440	00351968343440	rua mateus vicente 11	n 11	1500 445	rlmartins5c			None
19248329	Gladys Sala	gladysala29@hotmail.com	8299943445	8299943445	peaton 6 #50 Hato Mayor	50	51032	jordangil05			None
26322170	vangie05 vangie05	lvmrktg@gmail.com	555	555	123 Crysler		64130	vangie05			None
19294639	CHEAG THEAV	mafidaro166@gmail.com	0965051886	0965051886	phnom penh	41	855	theav10			None
15421341	sabato scala	fiammetta.bianchi64@gmail.com	3479824991	3479824991	via murolo	17	80039	scasab	382facfdafa0c490ec81be9b93b9dd1b		1964-06-24
32180460	Birgit Maria Kiekeben	mariakharting2014@gmail.com	00351916428707	00351212488727	rua Diogo Cao	26	2750-159	smarttlx4			None
21997289	feng wang	54684646@163.com	13605339874	13605339874	gao qing qu	10	255000	sxh5670			None
21068441	LAI YEE FOOK	rizatelex@gmail.com	60133662	60133662	Flat A, 3/F., To Li Garden	00852	00852	20140320272			None
13062527	Maria Sky	alejandro3jo@gmail.com	5557779999	5557779999	22union st Lynn mass	22	01902	na1			None
26617496	alex coelho	galocampiao@gmail.com	6177410002	6177410002	30 franklin st	1	02148	7iylx8gx17pt59c			None
29184610	Jhon Tibeth Rengifo Zavaleta	florimer.7@gmail.com	968546719	968546719	PRLG.JOSE GALVEZ 1360 EL PROGRESO		043	celoso2			None
10532909	venilson leandro da silva	ueslainevenilson@hotmail.com	9046541972	9046541972	403 branch wood ln		32256	vinnytel10			None
21894517	diana giron	dipagiba@hotmail.com	3213717436	2486749	mz 4 casa 15	15	00	minudiana10			None
17235575	Francisca Adima Naciemnto Andres	alfeganacimento@gmail.com	71112520	71112520	Calle Eugenio Von Boeck	S/N	10000	lilanacimento02	677ff5f13ed72ab9580c7a61af6f1645		1996-10-04
30470928	Jose Rebelo	telexrebelo@gmail.com	917601158	917601158	Rua francisco franco	3 E	2500	rebelo116			None
15354283	maria ortiz	anapastrana81@gmail.com	44347443335	44347443335	1305 kelston pl apt107	1305	28212	alicia2030	2d95666e2649fcfc6e3af75e09f5adb9		None
30004816	Bertha Contreras	reyesdeleoncastillo@gmail.com	50250486529	50250486529	Barrio El Porvenir		17011	titacontreras7	10f87c89f3fb79f8e86435548376a4a5		1994-08-16
16982689	FLAVIO GUSTAVO SANCHEZ CHAVEZ	gustavo@sarasantojitos.net	6862094330	6862530968	AV NARCIZO MENDOZA	1739	21257	flagusc02			None
26146704	luisella gobino	comunicazioni3plus@gmail.com	3356481793	3356481793	via cavourrina	14	12060	sagara3vp03			None
26838936	Fernanda Lopes de Mello		689958478	689958478	Rua Jo	1485	6890000	fo533ccda358bca			None
21424083	cleo mar rengel jimenez	margarita.rengel@hotmail.es	593999462340	5932490644	la sirena Oe4-51	51	170146	cleoren61			None
18023483	osvaldo antonio valdez vasquez	osvaldovaldez29@hotmail.com	8094493735	8094493735	c/primera #9 res. mendoza 2	9	11000	osvaldovaldez02	ee8b9658fd8e7133494ec5be4dbe897f		1975-04-04
26032112	jingwen yu	tfysyy679@163.com	13504559512	13504559512	beilin		152000	11dede08			None
26399666	Wilfer garcia	wgarcia@sidepsa.com.co	3147910652	3147910652	cl 41c n 62a 17 int 401	41	5001000	hernando310			None
10696403	edna o. pereira	ednapereirao13@gmail.com	9786015573	9787106268	1207 middlesex st		01851	edna0010	d93591bdf7860e1e4ee2fca799911215		1961-02-23
25170986	Saori Shiiki Nakao	saory.shiiki@gmail.com	09051188255	09051188255	sasagawa 4 chome 101 - 12	12	5100944	saory2013jp6e			None
18582907	Lalin HENG	lalin.heng@yahoo.com	0061405810099	0061405810099	37A Ashburton		6102	neath1234			None
28578540	MARYI DAVILA	yulixime05@gmail.com	0573105777173	05714564323	CALLE 22A N° 16A - 05	22	11001	mariyy08			None
19469725	wenhua zhao	zhaownhua888@163.cm	13945874178	13945874178	tl		153000	zwh788803			None
14113409	LOGANATHAN NAGALINGAM	logavisa@gmail.com	60123904755	60123904755	No.57, Lorong Bentara 16A,		41200	jindiya1tb4			None
15024209	Bruno Sousa	bmrs7.inetese@gmail.com	00351961840174	00351961840174	rua do concelho	18	9125-012	th3b3st	e02676412b538fe93f4a90805848bdf8		1991-04-13
32448086	Clara Hernandez	45448663	3144674854	2222561	CLL 54 16-18	1	50001	claraher5			None
19149861	ldhvkjbn fjvlj	isabelpascoa89@gmail.com	986547231	958554555	montalegre	3	5470-000	pascoa3006			None
22171995	Martha Patino	ayalamartha31@gmail.com	573004191340	5723965844	carrera 83B #15-122	15122	00000	alfi1	ac802e53178ac650a139574c534f872c		1969-08-24
24626918	Sou Thira	tolahor168@gmail.com	016807871	012756571	E1- 89, Street 55, Sangkat Boeung Reang, Khan Daun Penh, Phnom Penh City, Cambodia	89	855	hhkhmer5			None
26135776	Valdemir dos Santos	felipeandreu_ccb@hotmail.com	18996242626	1836445990	Rua Luiz Catharin	379	16201-235	kikasplinha2			None
25343132	Tony Falconi	tony_falconi@hotmail.com	016981041290	016981041290	Rua Rio Pardo	23	14270000	tonlucca205			None
23905874	mario puente de la vega aymachoque	jair0_12@hotmail.com	131231	983	av.las parhuas		03201	vegapo26			None
15288461	Dwi Prasetiyo	yunusemail@yahoo.com	+6287854497098	+6287854497098	Desa Sumberwudi Karanggeneng RT.001/001	001	62254	pras99	1200addfe20e301049ec29bec4d40331		1977-01-12
21838565	sdkjnk skjfru	murilodasilva60@gmail.com	98793280	9407502	ksjdfseu	11	02215	40mds			None
32416710	Aliaksandr Belkin	sasha1986m@gmail.com	375297415137	375297415137	Buinickogo	53-4	222304	sasha1986m09			None
23404545	carla alves	carlapalves82@gmail.com	964308633	291960000	rua da olaria ed alta vista bl2 2 i		9125-043	peugeot781c5			None
34024830	Mirela Soares Fran		689958478	689958478	Rua Jo	1485	6890000	fo534b36e7a106b			None
32107782	Maria de Lourds O. Silva Moreira	rogercaracas@gmail.com	7791719844	7791719844	Av: Brasil	692	45000000	hog5m5awshdoplp			None
25009474	Helen Gularte	helengularte@hotmail.com	59899951676	59822083814	Joaquin Requena	2374	11800	hegu07voip			None
28628768	DANIEL SANCHEZ SANCHEZ	danilotero1@gmail.com	0034657286949	0034657286949	CALLE VIRGEN DE GUADALUPE,8	8	03400	daniellotero16f			None
17135677	Daniel Martinez	ileanamartinez781@yahoo.com	2398229236	2398229236	1102 angelo ave.		33971	danielcuba52	d7f3c343f886ee24f320aa8a737747d1		1971-11-15
31209640	Gabriel Gonzalez	trebol64@aol.com	13473589713	13473589713	5618 48th ave		11377	kehilah40			None
13967743	sonia teles	zieffj@yahoo.com	6177109238	6177109238	825 aventura drive		37122	sty026	6e0d0cdf9e48538255bc003a8587e6d0		1969-02-03
31398256	rongbao qu		18678869979	18678869979	shangdongshengjinanshizhongqu	350	250000	qrb88010			None
22218041	alline novaes	allinenovaes@hotmail.com	556182072016	556182072016	Sha cj 06 ch 15 cs 19	19	72000000	allinenovaes32	e10adc3949ba59abbe56e057f20f883e		1980-06-30
31750172	qingli sun	xrj268268@126.com	13801633366	13801633366	yananxilu1764hao606		200050	am9008			None
31270836	Elio Antonio Pires	elio.pires@yahoo.com.br	0554209748209	0554234465427	rua Pref. Antonio Witchemichen	311	84400000	liderpires	0dec527c556b0354a28dafffc266063b		1975-04-19
12038525	endis miranda	endis011@gmail.com	2405155154	2405155154	11218 evans trail	101	20705	en87503			None
26616012	Alma Cedillo	todosganan55@yahoo.com	6825529465	6825520465	6367 Hanger Park Dr	6367	76119	cedillo119			None
21812225	pamela carreon chambi	mpamelitaper5@gmail.com	941249125	941249125	urb, guardia civil		051	perpame2			None
12843537	Tcvision Real Estate	tcvision2009@gmail.com	203 275 7753	203 275 7753	133 Garfield Ave	.	06606	tcvision38	935f099f5e344faddbae2ad5d08b681a		None
26355750	Bruno Peralva	brunosalema@hotmail.com	03533711623	03533711623	Rua Ismael Nery	12301	São Paulo	anaclaudia2c1			None

30482382	violet egbadon	violetegbadon@yahoo.com	08023374003	08023374003	3 odudu eleyiwo oniru estate	3	234	obbehi11		[redacted]	None
12974955	Choua Thao	chouathoj69@gmail.com	2092334176	2092334176	Dresden Way	6036	95212	chouathao1	4ecc04528d462a27230d8334e79b24b6		1969-03-02
15784159	haggeo del rosario	haggeo_rosario@hotmail.com	8299410223	8299410223	calle teniente amado	20	22000	alambre2	7dadbe5a37682c50d0600552bc389b33		1967-06-14
31950162	jose carlos dos santos	jcksantosju@outlook.com	6696103380	6699812366	rua porto velho	1050e	78575000	jcsantososa9	6b9db1f28f1e2ed917a49b3f6dcbd42b		1972-02-05
20992199	Suani Gutierrez	suany_3@hotmail.com	2414201637	2414182970	topacio	5b	90453	swane06			None
20132909	SANTOS BALVINO PEDRAZA HUAMAN	lueno7@hotmail.com	51983754785	51983754785	URB LA NUEVA GRANJA	S/N	03001	santosbalvino09			None
31554866	rosita valarezo	carlosvalarezo93@gmail.com	0993034631	052638078	calle 5 de junio	s/n	130103	rosiva1005			None
13098427	jiayu zheng	zzy6679a1@163.com	18088771801	18088771801	nangang		150001	zzy6679a4			None
22309379	Natanael Velasco	olimpogarciah@gmail.com	089987858	032792321	Alberto	001	160150	equipo185			None
13348423	Noe Hernandez Montero	nohemo50@hotmail.com	51968694270	5174497230	Av Alcantara Navarro	s/n	073	19noemonterovp			None
11060967	Zefiel Silva	zeguedes_075@hotmail.com	554599991187	554599124621	rua mato groso ap.22	995	85816020	zeguedesmacvel	d43070b938bdae4eda7f980f67a44e30		1975-11-21
20549351	Agnaldo Candido	agvip23@gmail.com	12981488666	01239332197	Rua Java	233	12235500	sa1140			None
18397293	rosalba ahumada	roahuma2@gmail.com	3127481519	8231814	calle 12 # 18 b 42		00000	rosalba1234			None
30351078	Ricardo Martins	alexcmartins13@gmail.com	962501314	962501314	estrada Padre A gostinho João Cardoso	39	9230-112	2kalexcmartins7			None
31352478	vvv vvv	lop.asen@hotmail.com	3970909	123	asd	123	00929	icce9			None
33905840	Ann Dingo	option20@gmail.com	8014659283	8014659283	1165 E Zach St		83706	muckluck4			None
10543101	aminah binti manaf	barakah62@gmail.com	0167734679	0167734679	jln balau 4	33	81750	barakah05a3			None
29513156	fernanda pereira	fernandapereiramadeira@gmail.com	0351963159315	0351963159315	avenida calouste gulbencan	9	9000-011	v7pife001			None
25736686	Miguel Pta	miguelp299@gmail.com	964953007	964953007	Rua Dom Dinis	o	9125-128	96495300727			None
26858334	elisabet marleny hidalgo rojas	msrojas770@gmail.com	51999056655	5114232467	Jr Nazca 656 dpto 303 torre 1	303	0001	chefaf07ip			None
16938215	cristiane lima	delimafabioo@hotmail.com	07908432781	07908432781	153	153	e16 4eg	londonx65	9fe6f4d1ef8fab724cc1779b732a042d		1978-03-09
20121573	segundo arsecio herrera cabrera	jennyherrera725@yahoo.es	0997476702	2726531	La Concordia	001	230101	chelcho001			None
23243981	ken Vang	chayc.cha@gmail.com	9165393054	9165393054	7564 chateau ct		95828	kenv32			None
23582532	FANNY LUCIA FEICAN MATUTE	fannitafe@hotmail.com	593993951668	593993951668	LIRIÑAÑ	LOTE	170146	jesusmiluz34			None
28361926	rosa fuentes	rosafuentes074@gmail.com	9254217242	9254217242	.	.	02149	grosacf09			None
14882899	Handryus de Medeiros Moresco	handryusmoresco@gmail.com	4599157777	4530389816	R Dr Flauzino Mendes	353	85819750	7aguiasroze7	e6f2e66a147445dfb2f26048e4781e29		1980-08-20
19963569	voz1046 heller	uplinebruno@gmail.com	555591004560	555591004560	rua cel pillar	336	98005190	voz1046			None
22553517	william ccollana mejia	wiliamax1@hotmail.com	973539904	973539904	manahuañunca	l-2	84	sonso32			None
22700501	qingxia liu	lqx666a@163.com	13512607447	13512607447	bawujiu		156300	lqx760b7			None
18725587	carlos sandoval	vicente_091195@hotmail.com	59366557892	59332874567	tarqui	76-93	00000	jolvi1cl6			None
26995892	osmarina pereira	agroalba@mhnet.com.br	4288032008	4288032008	casa	2063	83900000	osmarinaw2a			None
23452199	Esperanza Suárez	espesuarez-1234@hotmail.com	5593993279679	5932824177	Llano Grande:calle Carapungo y Psje: Princesa Pacha	12	170155	oresty06			None
18529075	ffuggb fhhgvff	jp_pa@hotmail.com	235678	56776	calle64A #50-24	2356	0000	serna04			None
18158445	Lamech Muluya Lamech Muluya	muluyakibudde@gmail.com	+256712538522	+256712538522	Block 223plot 2165 nabwojo	07125	256	kalabaya9			None
16136951	Daniel Fernando Almeida Silva	danisilva8181@gmail.com	00351918814581	00351918814581	Rua Sao Martinho	215	3700 384	danielasilva03	2dcdd05ece6efa19f17274bbc1364330		1981-10-24
29971722	ameriquito911c ameriquito911c	jrodriguez6507@hotmail.com	8293819106	8093801727	center st #1	1	02136	ameriquito911c			None
33910392	ruby song	rubysong003@yahoo.com	7182169951	7182169951	144-51roosevelt ave apt 8D flushing ny 11354		11354	rubysong04s3			None
17348805	Nasim Akhtar	hsiwani@hotmail.com	8328944357	8328944357	15310 Beecham dr	15310	77065	sara986			None
33661606	jose cardenas	enrriquedante1@hotmail.com	156156156156	5456456456	av. dsadwqdqw	32	001	76enchufex			None
22981403	JONATAS REIS	pekmanbbom@hotmail.com	556934714151	556934714151	av independencia	990	7691600	peka2			None
28643742	luis kasimoto	luiskasimoto@gmail.com	08047222021	0276569157	yorikido1649-1miura koopo 105	miura	3700535	luk26105f			None
21946811	Cristhian Huamancari Cubillas	cabazolo105@gmail.com	051963751971	051963751971	Av. Marco Polo1461	31	01	cubillas4vp3			None
22134655	marli eugenia dos santos marli eugenia	joserobertoramalho@outlook.com	18-996551669	18-36423154	fairfield road	20	182	marlibgi1x0			None
30433494	rosangela cino	rosaangelacino@libero.it	3472663376	3472663376	via kennedy	14	92024	fiocco3			None
28107942	roberto pinto	robpin1958@yahoo.com.br	5521994086384	552126701321	rua nelson lameira pinho	260	26440590	robpin20168			None
31910156	ELIZABETH MORALES MUÑOZ	helo977@hotmail.com	573112082396	573143297019	AV 3 #15-87 EDIF ORIPAYA		540006	elizamor10			None
12240909	Edison Oswaldo Jurado Aleman	wcm777soporte.ec@gmail.com	593995721476	593995721476	Quito	777	00000	visionariodeluz	a14523e002a82deb3b41ab9ab2a7f01d		1971-03-13
16678043	Berlissa reyes	reyesjunior28@gmail.com	18199052424	18299052424	san francisco de macoris		31000	ramses8824			None
11142915	Gisele de Vasconcelos	voipsenax@yahoo.co.uk	442056986547	442056986547	22 Malin Court		HP3 9AQ	5voipsenax341			None
17526707	jerseilian jesse	jerseilian@bol.com.br	99314089	32277072	rua alberto torres	77	69918272	jerseilian999			None
15820355	leonardo meira	lozico5@gmail.com	.	5088722268	20 bonito dr.	.	01701	mlml123	ed07470d794db05903a5de1e1cc64043		1963-05-05
28740870	jkhcdb yewtye	a977312519@gmail.com	43545656	78327848	khjdsjh	1	93776	mllk4			None
14542109	Faustino Torres	telexfreesuccess@gmail.com	4079286403	4079286403	5514		32821	interc38			None
23510178	rosa yanqui yanqui	rositayanqui18@gmail.com	951137080	951137080	jr.pierola	2558	051	yulys107			None
23598116	Rosaura Elena Miranda De Villanueva	invercion4@gmail.com	987600049	4941191	Ex Calle Union 107		001	buja26			None
103975	GHACHEM Cedric Sadri	cedric.gh@gmail.com	33603458392	33603458392	rue	11	94440	cedricgh	ca81257d0a041dd986e9972ec6f4e225		None
30998902	vicente sanchez	eljefemaximo@hotmail.com	678262473	678262473	teide	..	...	delavegamaximo			None
23661504	65546465 654	6545@564	213	65454654546	654	546	21000	cvt04		[redacted]46	None
19746185	angel leoncio vasquez rojas	angelvr1971@hotmail.com	593992162854	59372421505	cdla jaime roldos	sn	030450	todopoderoso810			None
12255389	anthony giraldo echeverry	anthonygiraech10@gmail.com	3116087504	3137212032	calle 13 Nº 8 - 30		472	anthonygiraldo	90ce1abf15835bcdf2326b31f8913bda		1986-07-03
21745071	Kim Mov	hongseng57@gmail.com	098868696	098868696	PP	PP	855	mov18006			None
31246572	Ly Cheng	luolyeng@yahoo.com	012818713	012818713	No.41, st.143 Phnom Penh	01281	855	pscheng7d			None
15519015	Lit Jang Fang Xie	fongxie11@hotmail.com	7873996719	7873996719	Urb. Ciudad Jardin	149	00987	rob8	dceeb85ee6ef9808a39800fe69f64114		1989-03-07
28831180	andrea paula	paulorobert156@bol.com.br	8293067550	8293067550	praca bom jesus	194	57910-000	voipvalmm35			None
16832997	huiqun wu	whq2013666@126.com	15213125413	15213125413	chongqing		400000	whq0006	f9fbe35b8769c57b7b324c8ec0fc59cd		1969-09-23
23338299	Rosalinda Trinidad Venancio	rosalindatelexfree@gmail.com	051999595113	051999595113	centro poblado huarichaca		01	rosalindatv187			None

21045701	a a	seeveelee@gmail.com	1	1	1	1	11204	usnewying007c4		[redacted]	None
32600292	rosa margot miranda pomasunco	faranibar3@hotmail.com	0512611293	0512611293	av. sucre 1125 dpto. 44	1125	01	trome023			None
26960496	lana luiz	emshego@hotmail.com	876543908	87890936	breton	99	20765	pesas29			None
14991993	dolores gonzalez gonzalez	samy.cortes@yahoo.com	573207214497	573007056185	calle 13 No 24-70	13	5001000	dologg03	934b535800b1cba8f96a5d72f72f1611		1965-05-18
16006241	CAIDA ALICIA GOMEZ PAREDES	joseloparati@hotmail.com	51991287674	51991287674	av. 6 de diciembre 20	20	063	kaydita18voip			None
4808828	MASSIMILIANO MINORCHIO	mminorchio@libero.it	+393479126620	+393479126620	VIA PERTINI	21	40059	massimilianom	c1f2c71ded85e3a3eb6fe1c5b690417e		1976-12-17

Exhibit 2.2

The *rep_id* field values from *representante* for these 400 records, sorted numerically.

rep_id
103975
1837442
3422538
4808828
5342132
6909814
8233698
8990114
10481737
10532909
10543101
10696403
10849613
10859665
10895323
10924917
11002661
11010505
11060967
11124597
11139601
11142915
11173547
11244297
11602575
11613041
11940641
12006479
12029215
12038525
12083907
12101361
12230337
12240909
12253845
12255389
12264801
12309161
12326283
12373697

12486309
12599305
12741791
12839407
12840111
12843537
12974955
12998949
13024661
13042245
13062527
13098427
13187479
13274413
13283601
13310459
13310573
13348423
13390271
13480503
13535571
13673021
13693377
13693843
13708471
13967743
14092857
14113409
14186869
14462651
14471807
14542109
14580551
14615005
14665365
14851591
14882899
14987623
14991993
15009737
15018283
15024209
15058255

15187881
15244391
15269983
15273029
15288461
15354283
15359791
15360707
15421341
15424181
15515561
15519015
15611301
15622367
15631715
15648265
15763669
15784159
15804223
15816073
15820355
15890279
15948487
15994557
15997647
16006241
16136951
16190613
16362967
16370859
16513577
16650765
16678043
16774341
16832997
16881825
16885985
16938215
16982689
17056207
17133111
17135677
17150481

17205639
17235575
17342739
17348805
17353657
17366803
17420867
17440663
17526707
17617649
17641357
17655023
17690393
17748067
17760751
17784207
18023483
18050007
18158445
18285155
18301429
18338867
18397293
18527109
18529075
18542715
18582907
18623951
18662359
18715363
18725587
18874971
18932255
19042495
19149861
19176359
19201259
19248329
19286159
19294639
19367271
19409231
19446541

19469725
19566151
19582027
19746185
19839823
19963569
20121573
20132909
20162527
20236581
20302025
20549351
20587585
20613851
20635873
20923117
20976333
20992199
21045701
21068441
21375225
21424083
21495295
21603193
21676513
21745071
21767977
21812225
21838565
21894517
21946811
21995387
21997289
22067481
22134655
22171995
22172165
22218041
22309379
22325361
22425649
22442207
22549637

22553517
22553765
22593609
22625973
22700501
22759639
22779907
22825309
22855173
22878987
22981403
23055097
23105921
23128273
23229189
23243981
23338299
23404545
23437753
23452199
23507474
23510178
23582532
23586116
23598116
23661504
23746160
23779744
23905874
23939388
23966676
23991518
23996308
24326730
24409536
24596252
24603072
24626918
24648950
24717556
24754476
24762910
24824544

24866050
24990458
25009474
25170986
25300804
25343132
25409672
25579732
25618846
25681294
25713568
25736686
25827594
25908762
26020118
26032112
26081656
26135776
26146704
26152056
26175142
26182512
26322170
26355750
26399666
26462494
26588988
26616012
26617496
26719794
26838936
26858334
26960496
26995892
27179502
27253936
27574160
27588896
27794690
27858064
27968580
28043426
28082806

28107942
28240802
28361926
28381230
28414936
28473488
28510658
28514922
28578540
28622356
28628768
28643742
28683680
28704006
28738876
28740870
28806740
28831180
28905288
28930870
29016170
29027608
29145636
29184610
29198424
29203692
29329960
29340744
29513156
29554222
29971722
30004816
30114492
30169436
30224954
30328944
30347198
30351078
30422180
30433494
30467382
30470928
30478712

30482382
30520312
30573942
30602924
30736220
30765762
30812884
30912060
30957314
30998902
31030906
31051852
31054766
31122188
31188226
31190562
31209640
31246572
31270836
31352478
31382076
31398256
31485092
31554866
31632038
31740540
31750172
31809602
31910156
31950162
32009236
32107782
32159622
32180460
32325898
32366660
32410022
32416710
32448086
32518998
32600292
32653674
32755544

32766744
32798142
32799100
33056690
33247320
33501542
33561636
33605882
33661606
33696458
33812486
33835388
33905840
33910392
34024830
34050578

Exhibit 2.3 The *rep_nome* field values from *representante* for these 400 records, sorted alphabetically.

rep_nome
000000 0o0
1 1
65546465 654
8 po
ADAUTO GOMES PINHEIRO NETO
ALBERTO GUILLEN TELLO
ANA MALHEIRO
ANDERSON MURILLO
Abel Ernesto Chacon Chacon
Adriana Fuzikawa
Ageu Divino Borges
Agnaldo Candido
Aida luz Millan mendez
Ala Khamiankova
Alexis Cabral Morel
Alexis Rekeraho
Ali H Omar
Aliaksandr Belkin
Alma Cedillo
Ana Luis
Andrés Saavedra
Andy Asian
Angeles Alvarez
Angelica dasilva
Ann Dingo
Antonio Beitum Beitum
Antonio Fidalgo
Armando V. Sousa
Bayron Emilio Garcia Amaya
Benedicto Carreño Castro
Berlissa reyes
Bertha Contreras
Betzabe Laureano Lazaro
Birgit Maria Kiekeben
Braulio Enrique Gavarrete Panchame
Bruno Peralva
Bruno Sousa
CAIDA ALICIA GOMEZ PAREDES
CARLOS E. GOMES
CHEAG THEAV

Carla Regina Rodrigues da Silva Sueth
Carlos Rodrigo Jacome Ponce
Carmen Duran
Choua Thao
Christian Baca Silva
Clara Esquivel
Clara Hernandez
Cosme Rodrigues
Cristhian Huamancari Cubillas
César Magno Abreu Castro
DANIEL SANCHEZ SANCHEZ
Daniel Martinez
Daniel Fernando Almeida Silva
Dl1 Inc
Dustin Leavitt
Dwi Prasetiyo
EDIMAX NETWORK CORPORATION
ELIZABETH MORALES MUÑOZ
Edgar Rodriguez
Edison Oswaldo Jurado Aleman
Edith Machaca cruz
Eduard Pfau
Eduardo Alves
Eduardo Ferreira Junior
Egidio Fulberti
Elen white Magalhaes lopes
Elias Bonfin Amaral
Elio Antonio Pires
Enrique Torello
Erik Damian Munoz Monserrate
Esperanza Suárez
Estefania Tapia
Evaldo cunha
Evgeniya Lebedeva
FANNY LUCIA FEICAN MATUTE
FLAVIO GUSTAVO SANCHEZ CHAVEZ
Faustino Torres
Fernanda Lopes de Mello
Filipa Andreia Rocha
Francia Feliz
Francis Kinyua
Francisca Adima Naciemnto Andres
Francisco Almazan

Francisco Javier Vergara Redondo
GEOVANNY PELAEZ
GHACHEM Cedric Sadri
Gabriel Gonzalez
Genner Bolivar Puello
Gentil Fernandes da Silva
Gisele de Vasconcelos
Gisselle Ricca
Gladys Alexander
Gladys Sala
Grace Semugenze
Guilherme Costa
Guu Xang
Handryus de Medeiros Moresco
Helen Gularte
Helena Mantese
Henrique Urias Santos
Ho Lai Fong
Ionel Iacob
Isabel Freitas
JAVIER MAXIMO LEIVA CIRIACO
JONATAS REIS
Jason Azevedo de Farias
Jhon Tibeth Rengifo Zavaleta
Jorge Velasco
Jose Aparecido de Andrade
Jose Hernandez
Jose Lucero
Jose Rebelo
Jose Sousa
Jose Teixeira
Juan Evangelista Villalvir Martinez
Julia Keko Aki
Kadi Ka
Kim Mov
LAI YEE FOOK
LOGANATHAN NAGALINGAM
Lalin HENG
Lamech Muluya Lamech Muluya
Lara Giane da Silva
Lisset Marcela Giraldo Lujan
Lit Jang Fang Xie
Luis Fernandes

Luis Luna
Ly Cheng
M Itv
MARYI DAVILA
MARÍA DEL PILAR CASTELLANOS PUENTES
MASSIMILIANO MINORCHIO
Manuel Alves
Mapril Pena
Marcio de Oliveira Gomes
Maria Lorena Castillo Aveiga
Maria Sky
Maria Soledad Chamy Roman
Maria de Lourds O. Silva Moreira
Marie Fanny Desanges
Mario David Montiel Lopez
Mark Vang
Martha Patino
Miguel Pta
Mirela Soares Fran
Mohammad Muhsen
Nasim Akhtar
Natalina Aliprandi
Natanael Velasco
Newton cruz
Noe Hernandez Montero
Nor Rahiniza Abdul Ghani
Nélio Ferreira
OSCAR MORA
Otilio Prado
Palmira Diniz
Paula Prak
Paulo Beloni
Paulo De Freitas
Pavel Danilin
Percy Choquehuanca Mendoza
RADDYAWATI CHAN
RAUL MARCAS VALDEZ
Rafael Frias
Ramon De Leon
Ramon Tice
Raysa Marleny Leonardo Olaverria
Renald Paul
Renato Marins

Rene Silva de Oliveira
Ricardo Kazuhiro Ikematsu
Ricardo Martins
Robert Rodricks
Rodrigo DeJesus
Rosalinda Trinidad Venancio
Rosaura Elena Miranda De Villanueva
Ruben diaz
SANTOS BALVINO PEDRAZA HUAMAN
SAYON RATANA
Salomé Chagas
Sandro Scamporlino
Saori Shiiki Nakao
Sebastiana Ururi Pumahuanca
Segundo Gregorio Celi Guerrero
Sou Thira
Suani Gutierrez
Tcvision Real Estate
Thais Medeiros Silva
Tony Falconi
UGT ghj un GF tgjf
Valdemir dos Santos
Victoriano Mazariego
Vinicius Lima
Vinicius Wetter voip
Wen zhi
Wendson Santos Fonseca
Wilfer garcia
William Silva
ZENON QUISPE CHOQUEVILCA
Zefiel Silva
a a
abc abc
abraham laura pacco
adalia Buendia
adriana menegati
alba y Castro
alex coelho
alline novaes
ameriquito911c ameriquito911c
aminah binti manaf
ana paula
andrea paula

andres ocampo posada
angel leoncio vasquez rojas
angel suazo
anthony giraldo echeverry
antonio rudrigues
aubri salvador
aylaiheang07 aylaiheang07
b b
bai sun
beliza alcca quispe
blessing adat umo
bruno silva
carla alves
carlina meza
carlos perez
carlos sandoval
carlos santos
carmen elena alban orozco
cassiano marko
cece cece
cesar augusto rodrigues aguero
cesar lozano
cesar machiavello
claudia chaves
cleo mar rengel jimenez
cliff monique
cristhiam chamorro
cristian escobar
cristian mamani coaquira
cristiane lima
da liu
daniela meneses
danimari llc
denisson carvalho fontes
diana giron
diego mavalo
dolores gonzalez gonzalez
ds fd
edna o. pereira
eduardo alvarez morales
elisabet marleny hidalgo rojas
endis miranda
enia valentin

erica mazo
eris mejia mejia
ernestina tavarez
ester evangelista de matos
fabiana Leite Ribeiro
fabio bomfim
fabio wanzeler
fabrizio
feng wang
fernanda pereira
fernando carreño
ffuggb fhhgvff
freddy vivar
gaspar condori cruz
gilberto guayamis galoc
gilma amaya
graziana
gui long yang
guibin huang
guohua chen
guoqi han
gutierre santana
hafizs mustaffa
haggeo del rosario
hassan abdi karim
henry kiddu
hgcvuygihop vjhbiuhojp
hongliang gong
hongyan li
hortencia torres j
hugo mamani choquehuayta
huiqun wu
humberto vladimiro suarez rios
ivan fernandes
janser rodriguez
jefferson herrera
jerseilian jesse
jesus alfonso aguirre hernandez
jesus condori gutierrez
jiayu zheng
jing guo
jingwen yu
jislaineborges borges

jkhcdb yewtye
jonathan yuri alves santiago
jose antonio callapiña condori
jose cardenas
jose carlos dos santos
jose escolero
jose rodrigues
jose rodriguez
jose silva
juan carlos buitrago
julio cesar
ken Vang
keqing qin
lana luiz
ldhvkjbn fjvlj
leonardo meira
licinio1 amaral1
lkj jhg
longshen chen
lourdes zambrano marcillo
lucinda maria de paula neta
luis kasimoto
luisella gobino
ma vena
manoel cleto cordeiro
marcelo bruno matheo calumbi
marco fernandes
maria ortiz
maria villegas
marianna porcelli
marino belen
mario puente de la vega aymachoque
marli eugenia dos santos marli eugenia dos santos
meiying gao
melissa alvarez velasquez
mengqing ma
mercedes garcia
milda weck
mingzhong wang
miriam cusi sucasaire
monaco giovanni
osmarina pereira
osvaldo antonio valdez vasquez

oupalavanh bouaphet
pamela carreon chambi
paulina andrea arias perez
pedro kilo
petra luna
ping chu
qingli sun
qingxia liu
ramona apaza apaza
rekiya salama
riev samann
roberto pinto
rongbao qu
rosa fuentes
rosa margot miranda pomasunco
rosa yanqui yanqui
rosalba ahumada
rosangela cino
roseli felix
rosita valarezo
ruby song
rui martins
rundong zhao
sabato scala
salustia perez
sandy aguirre
sdkjnk skjfru
sebastian jimenez
segundo arsecio herrera cabrera
sergey daerty
sgs sdgsdv
shijie da
shou dong wang
shoujie han
soeun savry
sonia teles
thabit salim
van prom
vanda maria almeida
vanesa zhagui
vangie05 vangie05
venilson leandro da silva
vicente sanchez

violet egbadon
vitor gonçalves
voz1046 heller
vvv vvv
wellington de jesus pereira
wendy pichardo
wenhua zhao
wertyuiop
william amaral machado
william ccollana mejia
wreew sfasd
xiaohong duan
xue bai
yi luo
yolanda landaverde
zhi zou

Exhibit 2.4

The *rep_email* field values from *representante* for these 400 records, sorted alphabetically using the characters in reverse order.

rep_email
45448663
6545@564
CHARK100@GMAIL.COM
voipsenax@yahoo.co.uk
zhaownhua888@163.cm
zzy6679a1@163.com
zhbllq1@163.com
lh197109222022@163.com
royal0816@163.com
54684646@163.com
lx51876@163.com
yangzhiwen86@163.com
hanshoujie7987@163.com
13666303168@163.com
hgb19770329@163.com
tfgyl679@163.com
tfysyy679@163.com
lqx666a@163.com
hjsgf@163.com
gg921jj@163.com
rert@163.com
cgh130121@126.com
zxy13581932995@126.com
whq2013666@126.com
xrj268268@126.com
duanxiaohong1919@126.com
schamy@axbridgegold.com
cosmemh@icloud.com
telexfree@telexfree.com
yemren@live.com
jislaineborges@live.com
vanda.almeida@outlook.com
lourdesosita@outlook.com
joserobertoramalho@outlook.com
jcksantosju@outlook.com
anthonygiraech10@gmail.com

option20@gmail.com
clarita1850@gmail.com
murilodasilva60@gmail.com
msrojas770@gmail.com
jehovajhire001@gmail.com
endis011@gmail.com
henrykiddu6911@gmail.com
fontainha11@gmail.com
Ericamazo1021@gmail.com
ayalamartha31@gmail.com
nbruzzone31@gmail.com
danisilva8181@gmail.com
anapastrana81@gmail.com
shaafici1@gmail.com
danilotero1@gmail.com
songwenjie0012@gmail.com
jozineiusa2012@gmail.com
anmur2012@gmail.com
jositan12@gmail.com
clauass12@gmail.com
drfranciscog1962@gmail.com
barakah62@gmail.com
carlapalves82@gmail.com
roahuma2@gmail.com
mapi2803@gmail.com
alexandrausa2013@gmail.com
ednapereirao13@gmail.com
alexcmartins13@gmail.com
carodrigues123@gmail.com
osorioleslie23@gmail.com
agvip23@gmail.com
hortencia1263@gmail.com
andrex1963@gmail.com
dripaloma73@gmail.com
duarte.jose93@gmail.com
carlosvalarezo93@gmail.com
iachimvitalik3@gmail.com
evaeva01012014@gmail.com
grupogalicia2014@gmail.com
mariakharting2014@gmail.com
fiammetta.bianchi64@gmail.com
rosafuentes074@gmail.com
ljose7194@gmail.com

joseaparecidodeandrade4@gmail.com
invercion4@gmail.com
cabazolo105@gmail.com
colin920605@gmail.com
yulixime05@gmail.com
azakatsuki1515@gmail.com
e.kd025@gmail.com
a0925147655@gmail.com
juank1875@gmail.com
taochen75@gmail.com
zrr8695@gmail.com
lozico5@gmail.com
mpamelitaper5@gmail.com
salomechagas16@gmail.com
mafidaro166@gmail.com
bambola66@gmail.com
zenon.telexfree66@gmail.com
miagee76@gmail.com
pashadan86@gmail.com
gustavoescobar6@gmail.com
florimer.7@gmail.com
bxue2737@gmail.com
hongseng57@gmail.com
ccl15367418087@gmail.com
soeunsavry7@gmail.com
newton.cruz7@gmail.com
rositayanqui18@gmail.com
reyesjunior28@gmail.com
tolahor168@gmail.com
akijuh8@gmail.com
telexapp8@gmail.com
tcvision2009@gmail.com
a977312519@gmail.com
aomar8049@gmail.com
gaomeiying1969@gmail.com
chouathoj69@gmail.com
duartesousa79@gmail.com
fabriziobonardi79@gmail.com
agatazxc289@gmail.com
isabelpascoa89@gmail.com
miguelp299@gmail.com
davidbeeba@gmail.com
wbeyanca@gmail.com

moniquelombeida@gmail.com
chayc.cha@gmail.com
fapgrocha@gmail.com
zileonela@gmail.com
jonaula@gmail.com
alexisbwana@gmail.com
maprilpena@gmail.com
porcellimarianna@gmail.com
chanshanazara@gmail.com
fernandapereiramadeira@gmail.com
logavisa@gmail.com
duarte69sousa@gmail.com
vencedorusa@gmail.com
evaldoatlanta@gmail.com
nancysilhua@gmail.com
1988renesilva@gmail.com
sinlimitesavanza@gmail.com
suporteplanobbc@gmail.com
wcm777soporte.ec@gmail.com
kennedymamanic@gmail.com
javiermaximolc@gmail.com
ramondeleonc@gmail.com
kouvuepc@gmail.com
edy.dance@gmail.com
muluyakibudde@gmail.com
reglamide@gmail.com
seeveelee@gmail.com
cerqueirafree@gmail.com
tice.telexfree@gmail.com
mmalves.telexfree@gmail.com
rosalindatelexfree@gmail.com
viniciussuethtelexfree@gmail.com
idultelexfree@gmail.com
yurisantiagotelexfree@gmail.com
hessotelexfree@gmail.com
lidertelexfree@gmail.com
autoferraztelexfree@gmail.com
rahiniza.tlxfree@gmail.com
bmrs7.inetese@gmail.com
uriashenrique@gmail.com
a.belikov.staff@gmail.com
guayamisg@gmail.com
lvmrktg@gmail.com

olimpogarciah@gmail.com
cedric.gh@gmail.com
oupalavanh@gmail.com
saory.shiiki@gmail.com
paulaprak@gmail.com
edimaxnetwork@gmail.com
licinioamaral@gmail.com
ebpmultnivel@gmail.com
sasha1986m@gmail.com
carmenelenaalban@gmail.com
bancofinan@gmail.com
epvalentin@gmail.com
rievsamann@gmail.com
sofiefashion@gmail.com
cabezon.tibilon@gmail.com
poloamon@gmail.com
sophannarithhorn@gmail.com
galocampiao@gmail.com
gnkalubo@gmail.com
handryusmoresco@gmail.com
renatopenedo@gmail.com
dkfigueiredo@gmail.com
williamamaral.stgo@gmail.com
alejandro3jo@gmail.com
godie.banjo@gmail.com
telexrebelo@gmail.com
kilokilo@gmail.com
aguillentello@gmail.com
reyesdeleoncastillo@gmail.com
sandroscamporlino@gmail.com
uplinebruno@gmail.com
alfeganacimento@gmail.com
luiskasimoto@gmail.com
carlos.jacome.p@gmail.com
julio.cesar.prado.jcp@gmail.com
zacarias.raq@gmail.com
wcmgaspar@gmail.com
gennerbolivar@gmail.com
adolforober@gmail.com
jcedillor@gmail.com
rodrigo.realtor@gmail.com
friasr@gmail.com
rogercaracas@gmail.com

paulorjfreitas@gmail.com
robertrodricks@gmail.com
nelioferreiranegocios@gmail.com
telexfreesuccess@gmail.com
comunicazioni3plus@gmail.com
denissonnet@gmail.com
misionsietevmt@gmail.com
maxsacogu@gmail.com
elenatirnovanu@gmail.com
r.ikematsu@gmail.com
drummond.a.telex@gmail.com
squadratelex@gmail.com
rizatelex@gmail.com
armandotelex@gmail.com
roseeffelix@gmail.com
mauricioalarconmartinez@gmail.com
lauralopezalvarez@gmail.com
paulina.arias.perez@gmail.com
janserrodriguez@gmail.com
palmira.diniz@gmail.com
wetferg@tubaraomail.com
brayanvas00@hotmail.com
isaiasleite2010@hotmail.com
jrjr0320@hotmail.com
beliza_love_30@hotmail.com
nohemo50@hotmail.com
elfenix2280@hotmail.com
joaocamacho2011@hotmail.com
fongxie11@hotmail.com
brazilian.angel21@hotmail.com
angelvr1971@hotmail.com
doctornico71@hotmail.com
lu_depaula1@hotmail.com
enrriquedante1@hotmail.com
wiliamax1@hotmail.com
skio.12@hotmail.com
jair0_12@hotmail.com
aricar22@hotmail.com
percy.1972@hotmail.com
francia_1303@hotmail.com
michael_1803@hotmail.com
abc123@hotmail.com
ilse1073@hotmail.com

jossandra1993@hotmail.com
suany_3@hotmail.com
faranibar3@hotmail.com
gutcris2004@hotmail.com
espesuarez-1234@hotmail.com
enriquet444@hotmail.com
navivan54@hotmail.com
marcel21964@hotmail.com
rdb94@hotmail.com
eduardo_alves4@hotmail.com
marullo_05@hotmail.com
zeguedes_075@hotmail.com
vicente_091195@hotmail.com
franklin_444_5@hotmail.com
sofia_luis5@hotmail.com
ester.ev-2006@hotmail.com
school_69_6@hotmail.com
lissethmarcela6@hotmail.com
moshe107@hotmail.com
jrodriguez6507@hotmail.com
helo977@hotmail.com
Luisfernandes-87@hotmail.com
jessyestefany_1997@hotmail.com
lueno7@hotmail.com
geo.-9@hotmail.com
gladysala29@hotmail.com
osvaldovaldez29@hotmail.com
carlos_perez79@hotmail.com
dipagiba@hotmail.com
gissellericca@hotmail.com
hucajoca@hotmail.com
andresocaposada@hotmail.com
claudiociuffa@hotmail.com
lydamurcia@hotmail.com
markffaria@hotmail.com
angelexousia@hotmail.com
brunosalema@hotmail.com
e.rodriguez.joa@hotmail.com
jp_pa@hotmail.com
m.luis.madera@hotmail.com
felipeandreu_ccb@hotmail.com
edith_shande@hotmail.com
javiertelexfree@hotmail.com

fannitafe@hotmail.com
benecartage@hotmail.com
helengularte@hotmail.com
barbara_radloff@hotmail.com
mnghghg@hotmail.com
alizabelpg@hotmail.com
natalialiprandi@hotmail.com
hsiwani@hotmail.com
tony_falconi@hotmail.com
joseloparati@hotmail.com
ricardo_vivar_camavi@hotmail.com
japroservi@hotmail.com
almeganetwork@hotmail.com
pacohuichol@hotmail.com
renaldpaul@hotmail.com
fabioguimaraesbomfim@hotmail.com
pekmanbbom@hotmail.com
lop.asen@hotmail.com
ueslainevenilson@hotmail.com
alexandrareiscrrn@hotmail.com
emshego@hotmail.com
haggeo_rosario@hotmail.com
diegomavalo@hotmail.com
eljefemaximo@hotmail.com
delimafabioo@hotmail.com
maricita_aqp@hotmail.com
lopemar@hotmail.com
viniciuswetter@hotmail.com
jvillalvir@hotmail.com
jason.farias@hotmail.com
sebasjim_lds@hotmail.com
allinenovaes@hotmail.com
laragianes@hotmail.com
marcokeiros@hotmail.com
telexfree.moneynet@hotmail.com
myaperu@hotmail.com
irvins_gsv@hotmail.com
gilneyw@hotmail.com
rayzamarleny@hotmail.com
salusperez@hotmail.com
paulobeloni_foz@hotmail.com
bluevisal@ymail.com
trebol64@aol.com

promchiro@aol.com
brunogleal@globo.com
claireimage2000@yahho.com
ileanamartinez781@yahoo.com
hafiz92@yahoo.com
rubysong003@yahoo.com
sandyta1823@yahoo.com
bishopjma73@yahoo.com
todosganan55@yahoo.com
alba.castro95@yahoo.com
lfidalgo_28@yahoo.com
carmenduran9@yahoo.com
luis.suculanda@yahoo.com
alfredo_vlsc@yahoo.com
grrrace@yahoo.com
yumvang@yahoo.com
lalin.heng@yahoo.com
luolyeng@yahoo.com
zieffj@yahoo.com
yunusemail@yahoo.com
tintasbm@yahoo.com
eganndem@yahoo.com
thabit.salim@yahoo.com
sylvom@yahoo.com
violetegbadon@yahoo.com
joguiro@yahoo.com
wcsjunior@yahoo.com
samy.cortes@yahoo.com
hsuarezrios@yahoo.com
wlpmnv@yahoo.com
13563087333@qq.com
12345@qq.com
3589447775@qq.com
13487591499@qq.com
soporte@myonecallxpress.com
rhcnbmzg@snapunit.com
finanzas@fem-ingenieria.com.co
wgarcia@sidepsa.com.co
paulorobert156@bol.com.br
jerseilian@bol.com.br
cletocor@bol.com.br
marciogomesoline@uol.com.br
robpin1958@yahoo.com.br

nina.bbom@yahoo.com.br
cassianomarko@yahoo.com.br
elio.pires@yahoo.com.br
agroalba@mhnet.com.br
margarita.rengel@hotmail.es
mercedesrocio@hotmail.es
jennyherrera725@yahoo.es
fmorenodaza@yahoo.es
gustavo@sarasantojitos.net
cecgomes20@comcast.net
duster.l@cox.net
adautopinheiro@hotmail.it
pino.varratta@libero.it
mminorchio@libero.it
rosaangelacino@libero.it
isabel.freitass@sapo.pt

Exhibit 2.5

The *rep_cel* field values from *representante* for these 400 records, sorted alphabetically.

rep_cel
(809) 964-0979
(829)639 4590
+256712538522
+34627046884
+3518970862
+351967468465
+375257843934
+393479126620
+50499333582
+51 1 2642019
+6287854497098
+77015261730
+79170638522
.
00
0034629412629
0034657286949
00351912121505
00351912301180
00351916428707
00351918814581
00351926102236
00351961840174
00351965550681
00351968343440
00393204811016
00393454638036
0051959369858
0059179454547
00593988161297
0061405810099
01112455783
011962797197051
012818713
0148299303
0167734679
016807871
016981041290
0351962969869
0351963159315

03533711623
0512611293
051963751971
051999595113
0554299748209
0573105777173
0573173757607
0701396396
070215889
07388347867
0772436082
0781093781
07908432781
08023374003
08032092699
08032691565
08037037160
08047222021
08054318685
08095398656
081355203254
087088
089987858
09051188255
09064631070
091211606
0934561789
0959230639
0965051886
0973852882
0984619495
098868696
0993034631
0997476702
099761661
1
1
1111111111
1111111111
11111111111
12349973
1281015840
12981488666

13081029083
13083626044
1310000
131231
13209496565
13311883817
13473589713
13504559512
13512607447
13604507219
13605339874
13704518112
13761279502
13801633366
13891278831
13937805296
13945567987
13945874178
15046380954
15084105696
15089332888
15213125413
15279989002
155770888245
156156156156
15811916535
15960133009
17173302810
17184382292
17991853555
17992179898
18-996551669
18050898223
18088771801
18095445555
18199052424
18511060629
18660377988
18678869979
18765343728
18996242626
19-34526600
19783023434

19788769629
203 275 7753
2092334176
2096145465
213
2274822368
231488
235678
2398229236
2405155154
2414201637
250788590092
2672352798
2814274456
3013177872
3016930631
3108634173
3116087504
3125192104
3127481519
3144674854
3145825126
3147910652
3165890172
3175005899
3176165383
3194066658
3204950505
3208411664
3208431196
3213717436
3214210768
3331398595
3334357995
33395926
3343158136
3356481793
33603458392
3399273139
34342324
3454911746
34646090274
34676362107

34722317313
3472663376
3479824991
3495652865
351261937300
351913958066
351925534038
351967043776
351967876352
375297415137
393206565660
3970909
4079286403
414754245741
4156374504
4169939547
42323871932
4288032008
43545656
441212706722
442056986547
44347443335
447876558572
456457
45667
459898477896
4599157777
4632322806
46874654845
4806888158
50250486529
50495060089
50499957388
5086151681
5088603665
5108606309
5148935434
51959345479
51968084164
51968694270
51983754785
51991287674
51999056655

51999080254
523524246
5511992851262
5516981711961
5519982686928
5521994086384
553175983820
553199844030
554399352030
554599646613
554599991187
555
555591004560
5557779999
556182072016
556596038070
556699542801
556892298749
556934714151
557182094025
557999665426
558100612647
5584-88672797
558796834374
5593334957
5593993279679
559492817557
559899935406
5599444546
5615063846
573004191340
573006918601
573014942381
573108169326
573112082396
573114608493
573134208892
573207214497
573214570978
59322437118
59322504150
59366557892
593984999360

593990413417
593992162854
593993951668
593994536919
593994734538
593995721476
593998792632
593998812240
593999462340
59899951676
60123904755
60133662
60195944575
6124082185
6146320875
617 386 9689
6172067515
6172704232
6174130243
6176987267
6177109238
6177410002
6177784040
6179554744
6313352714
6466664
6599883386
661829310
6622569250
6641612812
6696103380
678262473
6788552995
6825529465
6862094330
689958478
689958478
689958478
691684848
6984593205
7025040232
7043607312
71112520

7182169951
72007700
7324060098
7543675913
7705497709
7742791900
7791719844
7813001513
7813304744
78191874
7873996719
79128701598
8014659283
8058869666
8092585655
8094493735
8095654857
8096133626
8097107067
8099109686
8134439434
8293067550
8293759656
8293819106
8293923776
8294326954
8299046324
8299161605
8299410223
8299943445
8328944357
84 96071106
8487482872
8498629761
85226714333
85516957695
8559294123
8559898
8573122571
8602092045
8613487591499
8613563087333
8613589447775

86344407
876543908
9
9046541972
914264638
9165393054
917601158
9254217242
941249125
942845624
943958087
944635541
946621498
951137080
951409268
952917617
9546874023
954729758
9548512332
9565298110
956545460
961063303
961265353
962501314
964308633
964851693
964953007
965135327
965456922
966380983
968546719
968665387
968708635
973505754
973539904
974219466
976107701
9782103090
9783330861
9786015573
979232336
981461007
983428883

984999180
986547231
986866244
987564
987600049
98793280
989649964
99072736
99314089
9931679450
995795449
9991310303
Exceden
VB CF y DG u d
dfghjkghjk
ewrewtre

Exhibit 2.6

The *rep_fone* field values from *representante* for these 400 records, sorted alphabetically.

rep_fone
(809) 964-0979
(809)472-0404
+256712538522
+351967468465
+351968970862
+375257843934
+393479126620
+50499333582
+51 1 2642019
+6287854497098
+77015261730
+79170638522
..
00
0034657286949
00351212488727
00351224670384
00351912121505
00351918814581
00351926102236
00351961840174
00351965550681
00351968343440
00393204811016
00393454638036
0051316021
0059179454547
00593988161297
0061405810099
00982545193
01112455783
011962797197051
01239332197
012756571
012818713
0148299303
0167734679
016981041290
0276569157
032792321

0351962969869
0351963159315
03533711623
0512611293
051963751971
051999595113
052638078
0554234465427
05714564323
0575133574
0701396396
070215889
07388347867
0772436082
0781093781
0782043285
07908432781
08023374003
08032092699
08032691565
08037037160
08054318685
08055091804
081355203254
08u087
09051188255
09064631070
091211606
0934561789
0965051886
0973852882
098868696
1
1
1111111111
1111111111
11111111111
123
12349973
12548453486
1288394867
13081029083
13083626044

1310000
13209496565
13311883817
13473589713
13504559512
13512607447
13604507219
13605339874
13704518112
13761279502
13801633366
13937805296
13945567987
13945874178
15046380954
15084730554
15089332888
15213125413
155770888245
15811916535
15960133009
1713302810
17184382292
1733431478
17992179898
18-36423154
18050898223
18088771801
18095445555
18299052424
1836445990
18511060629
18660377988
18678869979
18734526783
19-34526600
19783023434
19788769629
203 275 7753
2092334176
2096145465
2222561
2274822368

231488
234141
23456789
2398229236
2405155154
2414182970
2486749
250788590092
2672352798
2712389132
2726531
2732440203
2814274456
291960000
3013177872
3016930631
3061012
3125192104
3134278516
3137212032
3145825126
3147910652
3165890172
31920172
3208411664
3208431196
32277072
326127
3310256
3331398595
3334357995
3343158136
3356481793
33603458392
3399273139
3454911746
34646090274
34676362107
3472663376
3479824991
3495652865
34981634675
351261937300

351913958066
351925534038
351967043776
351967876352
375297415137
3810703
3863123
393206565660
4079286403
4156374504
4288032008
441212706722
442056986547
44347443335
447876558572
4530389816
45677
459846229385
4632322806
464646656
47333767
4806888158
4941191
50250486529
50495060089
50499957388
5086151681
5088603665
5088722268
5108606309
5114232467
511968084164
5148935434
5174497230
51959345479
51983754785
51991287674
51999080254
523524246
54546545645
5456456456
551120745126
551636302223

5519982686928
552126701321
553136240711
553199844030
554399378003
554535243184
554599124621
555
555591004560
5557779999
556182072016
556596038070
556699542801
556892298749
556934714151
557182094025
557999665426
558199612647
5584-20106105
558796834374
5593334957
559492817557
5597232254
559899935406
5615063846
5646
56776
5712893264
5714832952
5714936899
5718824435
5723965844
573007056185
573143297019
5745969874
577949516
593-2955016
59322437118
59322699199
59323741018
5932490644
5932824177
59332874567

59372421505
593990367147
593993951668
593994536919
593995721476
593998792632
593999831134
59822083814
60123904755
60133662
60341019493
6124082185
6146320875
61571428
617 386 9689
6172067515
6172704232
6174130243
6176987267
6177109238
6177410002
6177784040
6179554744
6313352714
65454654546
6599883386
661829310
6622803192
6646806455
6699812366
678262473
6788552995
6825520465
6862530968
689958478
689958478
689958478
691684848
6984593205
7025040232
7043607312
71112520
7182169951

72007700
7324060098
7543675913
7705497709
7742791900
7791719844
7813304744
7816200168
78191874
78327848
786543
7873996719
79128701598
8014659283
8028806
8058869666
8092336874
8093801727
8094493735
8095322990
8095566891
8095654857
8095964381
8096133626
8097107067
8097368776
8099109686
8134439434
8225702
8231814
8293067550
8299161605
8299410223
8299943445
8328944357
84 96071106
8487482872
8498629761
85265996128
85434321
85516957695
8559294123
8559898

8602092045
8613487591499
8613563087333
8613589447775
8627357
86344407
87890936
9
9046541972
905 8214181
914264638
9165393054
917601158
9254217242
9407502
941249125
942845624
943958087
944635541
946621498
951137080
952823647
9546874023
954729758
9548512332
9565298110
956545460
958554555
961063303
961265353
962501314
964851693
964953007
965135327
965456922
966380983
968546719
968665387
968708635
973505754
973539904
974219466
976107701

9782103090
9783330861
9787106268
979232336
981461007
983
984264269
984999180
986866244
99072736
9931679450
GN g UF DCC gg
U tvt
Ybańez
s/n
trruyt

Exhibit 2.7 The *rep_end* field values from *representante* for these 400 records, sorted alphabetically.

rep_end
#12,st2
#35E, St.378
.
.
07 home st
1
1
1 calle, 3 ave. Bo. Medina No.22
1031 Mchenry ave
104 villa drive
1102 angelo ave.
11218 evans trail
115 Brookline St
1165 E Zach St
1207 middlesex st
123 Crysler
12312 parkwood dr #110
12345 main
1285 Garcia street
13 Burkeley St
1305 kelston pl apt107
1310 PARTRIDGE CLOSE
133 Garfield Ave
133 irvin st
1383 Summer Ave
14 bruce st
14 shadowbrook lane
144-51roosevelt ave apt 8D flushing ny 11354
153
15310 Beecham dr
1818 Granby Way
185 foal court
1918 coliseum drive
20 bonito dr.
21 silver lane c1
22 Malin Court
220 Union Street
22222222222222222222222222
22union st Lynn mass
2313 cajun dr

23425435
2459 jacaranda dr
258 vine st
26 beacon st
26 stirling Road
2912 NW 67TH COURT
2° REP CAMP RAFFALLI CCL
3 odudu eleyiwo oniru estate
30 franklin st
3356,Old Line Avenue
3550 blve dr#103
3764 East Bart Street
37A Ashburton
40-letia Pobedy,8
403 branch wood ln
437 Captain Eames Circle
4392 Candlelight Drive
44 alhekma st
454 parkwood way
5 elm dr
540 D ST Apt 9
5514
5618 48th ave
56799
586 W Stacie Avenue
6237 crafton st
6367 Hanger Park Dr
647 brittany n
654
66 Harrington Farms
671 47 street,3a
6985, boul. des Laurentides, Laval,QC
7564 chateau ct
825 aventura drive
878 broadway
9
9 walden
97 bellevue avenue
AV 3 #15-87 EDIF ORIPAYA
AV NARCIZO MENDOZA
AV. Duque de Caxias
Alberto
Alfonso XIII

Antonio Sanchez s/n y Francisco Ecija
Av Alcantara Navarro
Av Juan Pezet 1765 - San Isidro
Av. Camino Real Nº 476
Av. Generoso Marques, 1562
Av. Marco Polo1461
Av. Sarasota 20
Av. Tnte Hugo Ortiz y Moro Moro
Av. Universitaria Km
Av. dos Caiapós
Av. españa edif 23
Av. francia
Av: Brasil
Avenida 17 de julio # 8-65
Avenida Laranjeiras
Avtozavodskaya 68
Barrio El Porvenir
Block 223plot 2165 nabwojo
Bond Streeet
Buinickogo
C-21-1-12, Pangsapuri Sri Abadi
C/ Sant Jaume
CALLE 1 # 12-58
CALLE 22A N° 16A - 05
CALLE VIRGEN DE GUADALUPE,8
CAMINHO DAS COURELAS
CAMPECHE 270
CLL 54 16-18
COSTA DE SOL
CUSCO - CANCHIS - SAN PEDRO
Calle 10
Calle 1ra, Res Don Jose II
Calle 72B No. 70F-45
Calle Eugenio Von Boeck
Calle Las Gardenias 232 dpto. 401 Valle Hermoso
Caminho Velho da Palmeira
Caminho dos Saltos
Caminho dos Tis
Carlota Jaramillo
Carrera 47 # 55-30
Carvajal 195 y la Gasca
Cll 12b 32A 86
Col. Sandoval sorto

Col. Santa Monica segunda calle bloque 2 casa 37
Cra 136A 145 30 Casa 36
DESA VIEW TOWER
Desa Sumberwudi Karanggeneng RT.001/001
Dona Marina dechamps
Dresden Way
E1- 89, Street 55, Sangkat Boeung Reang, Khan Daun Penh,
Phnom Penh City, Cambodia
Estrada do Pedregal
Ex Calle Union 107
Flat A, 3/F., To Li Garden
Flat E,11/F King Win Fty. Bldg.,
Gurabo Calle 20
Joaquin Requena
Josefina Reis Assunção
Jr Jose de San Martin 150
Jr Nazca 656 dpto 303 torre 1
Jr. Roberto Barriga
Jr. jose antonio encinas
Juan Paullier
Juana de Oriol
Kigali
Kunimoto2254-2
LIRIÑAÑ
La Concordia
Lavrinovicha
Livermore
Llano Grande:calle Carapungo y Psje: Princesa Pacha
Loural
MZ B LT10 AA HH NUEVA JUVENTUD
Main ST
Manzana 25 No.5 el eden
Meseta
NBA
No.41, st.143 Phnom Penh
No.57, Lorong Bentara 16A,
P.O. Box 21940 Kampala
PP
PRLG.JOSE GALVEZ 1360 EL PROGRESO
Plaza Arturo Piera
Po Box 3232
Potosi
Psje. Talara Pueblo Joven el Porvenir

Quito
R Dr Flauzino Mendes
R: Mariana
RES EL DORAL
RUA DA ESPERANÇA
RUA JOÃO MATEUS DE MORAES
Repardo 13 #17
Rua 13 de maio
Rua 405 Qda 124
Rua Augusto Haller
Rua Barreira Vermelha nº2A
Rua Dom Antônio de Almeida Lustosa
Rua Dom Dinis
Rua Guarani
Rua Helio Campo Dallorto
Rua Ismael Nery
Rua Java
Rua Jo
Rua Jo
Rua Jo
Rua Luiz Catharin
Rua Marcelo Terceros Banzer
Rua Nova Quinta Deão Edif. Cuibem 1 2º E
Rua Rio Pardo
Rua Sao Martinho
Rua das boas noites
Rua do Barreiro
Rua do Damasqueiro
Rua do Juncal
Rua francisco franco
Rua x1
Sha cj 06 ch 15 cs 19
Stepnoy4
Suit 3, Mammy, Lungi Barracks
Tercera av Norte y Novena Calle Poniente Varrio El
Travessa da Paragem. Moradia D
Trías de Bes
URB LA NUEVA GRANJA
Urb. Ciudad Jardin
VIA PERTINI
Via Strada Ciosa
Viale Tito Labieno 70
Wood Dale Rd

Zone 11 gariki Abuja
a barreira
adfaf
agrupacion ciudad nueva Mz 19 Lt 06
america
antonio josé moreira
asd
av independencia
av. 6 de diciembre 20
av. antisuyo
av. dsadwqdqw
av. revolucion
av. sucre 1125 dpto. 44
av.las parhuas
avenida calouste gulbencan
avenida curralinho
b
bancofinan@gmail.com
battambang
bawujiu
bayan
beijing
beilin
beisan road
breton
c
c/primera #9 res. mendoza 2
calle 12 # 18 b 42
calle 13 No 24-70
calle 13 Nº 8 - 30
calle 14 #11_44
calle 4b # 53 f 19
calle 5 de junio
calle acosta ñu
calle azua
calle bermudez #128
calle del maestro c cero cora
calle miguel negrete
calle san nicolas
calle teniente amado
calle3 No. 17 a 16
calle6 No 4AE 20
calle64A #50-24

capitol ave
carrera 25
carrera 52 B 75 50 apto 210
carrera 83B #15-122
carretera mella 102
carretera veron punta cana
casa
castrolazaro-portomarin
cdla jaime roldos
center st #1
centro poblado huarichaca
chongqing
chunxi road,chengdu
cl 41c n 62a 17 int 401
cll 152c 133-09
cll 20 6 53
cll 9 20-25
comunidad sencca chectuyoc
condor ñan
conjunto residencial campoy block l 515
coop. santa Eufracia Mz. D , Solar 1
cooperativa el triunfo
cr72 22d - 54
cra 4 este 30 21
cra 42 No. 18 a 08
cuatro torres
entebbe
estrada Padre A gostinho João Cardoso
estrada da camacha
estrada nacional 360
fairfield road
fdhfzfgigjuggf
fdtyfug
francisco chira y j paredes
gao qing qu
gaungdong
gjgugygyg
gui tang
henan
henan china
henanluoyang
huangheroad
irb. ttio la florida

jarison
jilin
jiqirod kuansahn
jln balau 4
joseverissimo
jositan12@gmail.com
jr huallaga 434
jr las palmeras
jr. miraflores
jr. miraflores
jr.pierola
kawanda
khjdsjh
kiev
kjh
km30
ksjdfseu
l
la primavera
la sirena Oe4-51
laizhou
lanxi
larrañaga
las palmas 3381 depto 4 lan-A Hualpen
lot 55 D kg teluk baru
manahuañunca
manz 21
micaela bastidas
mno.bustamante
montalegre
mz 4 casa 15
nangang
nangang
odonnell lane
peaton 6 #50 Hato Mayor
phnom penh
phnom penh
pobox 570372
pouk chas
praca bom jesus
r: Rio Solimões
rua 1 de maio
rua Diogo Cao

rua Pref. Antonio Witchemichen
rua airo
rua alberto torres
rua cel pillar
rua clarissas
rua da felicidade
rua da olaria ed alta vista bl2 2 i
rua do concelho
rua do triunfo
rua doze
rua luis epaminondas
rua macaé
rua mateus vicente 11
rua mato groso ap.22
rua nelson lameira pinho
rua porto velho
rua senador pinheiro ramos
rua:das turmalinas
rue
rue degrennb
san francisco de macoris
sasagawa 4 chome 101 - 12
shang hai
shangdongshengjinanshizhongqu
shanxi
shenlanlu
shijiazhuang
shiyan
sps
strada del salice vecchio km 4
tarakan
tarqui
taunton ave
teide
tiedong
tl
topacio
tra
truyuyt
urb, guardia civil
urbanização vista alegre
via Degli Stretti
via a la loma

via bartali
via cavourrina
via kennedy
via lequile
via murolo
via pinelli
via rosa luxemburg 13
weihai
wertyuio
westside av.
wilbahi estate
xian
yananxilu1764hao606
yorikido
yorikido1649-1miura koopo 105
yura cho 1620 10
zhangzhou

Exhibit 2.8

The *rep_numero* field values from *representante* for these 400 records, sorted alphabetically.

rep_numero

.
.
.
.
.
.
..
0
0
0000
001
001
001
00100
00393
00852
00852
01112
01281
02
03
06
07
07
07125
1
1
1
1
1
1
1
1
1
1
1
1
1-22

1-B
10
10
100#
1000
101
102
102
1050e
11
11
11
11
11
111
112
1125
115
118
12
12
12
12
12
12
123
123
12301
1231
1231
12312
125
13
13
1305
1310
132
1322
1342
136
1376-
1383
14

14
14
14
1485
1485
1485
149
15
15122
152
153
15310
15465
1598
16
16
1603
17
17
172
1739
17765
178/c
18
18
18
18.
185
19
1918
194
2
2
2
2
2
20
20
20
20
20
20
200
2025

205
2063
21
21
215
215
217
22
22
22
22
22
220
2254
2254-
226
23
23
23
23
231
233
2356
2370
2374
24
2446
2459
25
25
2558
256
256
26
26
260
270
291
292
2L
3
3
3

3
3
3
3 E
301
303
31
31
311
32
32
320
324
32a86
33
333
336
3381
339
34
345
350
353
37
37
379
39
394
3ºI
40
401
407
41
41
41
43
45
454
46 a
468 N
47
47
48

4AE 2
5
5
5
5
5
50
51
51 b
515
52
53
53-4
546
55
56
572
59
5b
6
6
60
6036
62
6237
63
633
6367
653
656
656
68
680
692
69785
6985
7
71
72
72
72
73/77
76-93

7699
77
772
777
779
786
8
8
8-65
86
88
89
9
9
9
905
99
990
995
CASA
Ih y
LOTE
PP
S/N
S/N
S/N
S/N
S/N
S/N
VB g
a 3
bloq
c-11
gkjhj
j-2
l
l-2
l-7
ll
miura
mjb
n 11
o

qweri
s/n
s/n
s/n
s/n
sn
sn
xxx
xxxx
None
None
None
None
None
None
None

Exhibit 2.9

The *rep_cep* field values from *representante* for these 400 records, sorted alphabetically.

rep_cep
.
.
...
00
00
0000
0000
0000
0000
0000
0000
0000
00000
00000
00000
00000
00000
00000
00000
00000
00000
00000
00000
00000
00000
00000
0001
0009
001
001
001
001
001111
00174
00254
00591
0062
00852
00929
00987

01
01
01
01
01
01
01545
01603
01701
01721
01749
01749
017571144
01803
01841
01851
01902
01930
02026
02126
02136
02148
02149
02149
02149
02150
02151
02176
02215
02368
02411310
02740
02914
02914
03001
030450
03201
03400
043
050000
051
051
051

051
051
051
051
051
051
057
06118
063
06606
070650
070650
070650
073
08225180
084
084
084
084
084
08440
08861
08980
090610
10
10000
100000
100000
10100
10100
10100
10147
11000
11001
11001000
11201
11204
11207
11220
11354
11377
114000
11800

11800
11900
11937
12060
12233220
12235500
130103
13480000
136000
14055320
142352
14270000
14702114
15
1500 445
150001
150001
15011
15047147
151500
15169
15172
151800
152000
153000
156300
157400
160150
16201-235
170105
17011
170127
170137
170143
170146
170146
170155
17602
18100
182
19003
19149
200050

200940
20260
20705
20724
20765
21000
21002
21040
21257
21345
21702
22000
22000
22000
222304
22240
23000
23000
23000
23000
23000
230101
234
23409
234567
234574
23666
247351
250
2500
2500-609
250000
2500155
250023
2510-049
2520-042
255000
256
256
256
25641
26040
261426

2625-622
264200
26440590
2750-159
28033
28212
29052303
30066
30236
304
31000
31100
31730725
32256
32821
33064
33064
33073
33309
334
33446
33822770
33971
35053340
363000
3700 384
3700535
3700535
37122
3730036
3788
40000
400000
400001
40059
41000100
41200
4121
41705155
4260257
426068
432072
43556

4370012
4485777
45000000
4505-505
45995000
46018
471000
472
4730
475000
48471
50000
50001
5001000
5001000
5001000
504
504
504
504
504
51000
51000
5100944
51032
511
515000
516000
518000
52120-090
53100
537499
540006
54001
54340095
5470-000
54753110
55109
55337
55828-868
56640000
57910-000
58100

59068360
59073110
60191
610000
610081
6102
62254
64130
6500000
66835
68520000
6890000
6890000
6890000
69900205
69918272
7000
70000
70088611
710000
71122
719000
72000000
73100
74180
760041
76119
76875828
7691600
77065
77520
7800
78572
78575000
78965
80039
809
81750
82001
83190
83706
83900000
84

84
84400000
85295
855
855
855
855
855
855
855
855
855
85550-000
85816020
85819750
85868150
86150
89157
9000-011
9020080
90453
9050-071
9050-219
9060-010
9125-012
9125-043
9125-128
9125-194
9125-902
92024
9230-112
9240-216
9300-304
93036
9325-024
9325-172
93257
9350-033
93776
94440
94806
94901
95212

95350
95828
95828
98005190
98804710
EC230101
H7H3A1
HP3 9AQ
LIMA 36
São Paulo
b169bg
e16 4eg
jghgjhj
kjb
l
tyuiuy

Exhibit 2.10

The *rep_login* field values from *representante* for these 400 records, sorted alphabetically.

rep_login
0511lisslujan87
06zenon
07wjp303
10voiccer
11dede08
1234vi1
12redgigi8
14hm7v3
168peterc1
17jdnstl
1982pao
19noemonterovp
1topcc1
20140320272
23majolivi01
2503sa7
2aikidoc
2i9xwxqejjbf910
2kalexcmartins7
2sousafn101
3145825125
3183365469
3santato1
40mds
5voipsenax341
76enchufex
7aguiasroze7
7iylx8gx17pt59c
8006607713
80991096867
8434
8497012235f
8q5tnw4on8yon83
96495300727
KIRIN02
Luckyspirits01
a331cli
abrahaml001
abregos7s
adaco31

aech09
afidalgo71
ag928
akijuh13c4
alambre2
albacastro14
alexandrareis1
alexpaqu3
alfi1
alicia2030
allinenovaes32
alvarez53c2
am9008
ameriquito911c
anaclaudia2c1
angelia01
anglada9
anka9999
anthonygiraldo
asgluis35
aubrisalvador_tlf
autof38
avsc9
azebeloni3
bafrm222c2
bann8
barakah05a3
bbbb11
bebelonita32
benedi99
betty567voip1
bhinvest114
bjavierv6
bonfin205voip
borges07voip8
bouaphet
buja26
burusac1
bvp059
bx27373
c17043607312
c21paluka4
c2csudk7ixrs935

carlacsueth7
carlosalegro01
carmenelena1
cartania710
casacambulos3
cece007
cecgomes016000
cedillo119
cedricgh
cel2voip08
celoso2
cesaraugusto298
cgh114a
charmed17
chefaf07ip
chelcho001
chouathao1
christian8943
cjacome1
cl12laragiane
claraher5
cleoren61
clientejsa03
cnoe1
color115
comandante0016
compostela20
cosmemh30
cubalibre38
cubillas4vp3
cvt04
danielasilva03
danielcuba52
daniellotero16f
dcf
delavegamaximo
dgbr0501
diegof114
dologg03
dorcasmar
drandpool5
dripa37
dxh06

edgardaniel02
edimax56
edna0010
egs010
elizamor10
elpalomo
en87503
entavarez9
equipo185
erikakaty6
eris1234
ermos4
escolero2
eser8
esterfone09
evalika22
fabrizio79
fannydes007
fdbs
feraatlanta46c
ffonos8
fgb3voip7
fiocco3
flagusc02
fo533ccda358bca
fo53420618c702a
fo534b36e7a106b
franciacliente
freenight2cl1
frex26174c3
fulberti47
gamaya827
gegre7
gfs09
ghl15
gionela11
gmq042
gpsiter1126
graceofchrist3e
greenleadafrik6
grosacf09
gruporedevix
gtt308pro2

guaytas4
gucall8
guillentello3
gyanni
hafizs10
hajiya
heang567418
hechos8
hegu07voip
hernando310
hgq404
hhkhmer5
hkt0001
hog5m5awshdoplp
honey09a
horto037
hsj1948
hug005
ianalejandros2
icce9
inte5
interc38
iranzi6
jaah84
janserrodriguez
jaq168e
jason1f2
javierml7voip1
javitorias8cl6
jcsantosusa9
jerseilian999
jesusmamani6
jesusmiluz34
jindiya1tb4
jishb63c
joaquinvoip22
joguiro
jolvi1cl6
jordangil05
juliocc5
jvillalvir01
kacr1
kalabaya9

kash1
kaydita18voip
kehilah40
kenv32
kikasplinha2
kikebcn
kolik3
kpqq555
larapy5
lbf1623013
ldh43
lduster5
lebazi3p6
lh0rb1ymfondifx
lhy5882
liderpires
lilanacimento02
lnd2jgapftbcm5s
londonx65
loperro52
lorecast
loren0901
lqx760b7
lubolfone7
lucky4phone5
lucypoc4398036
luisfernandes5
luk26105f
lx51125
ly5009
maa1voip2
maake8
mags0003
malidy4c8
mapy02
marcelotelexba2
marianna36
marinobelen
mariyy08
marlibgi1x0
massimilianom
mavalo10
maxsavoip45

mayola199
mgomesgroupcl
milces62
milingo3
minudiana10
mllk4
mlml123
mov18006
movini7v1
muckluck4
myaperu3voip8
mybless81voip6
mzc0001
n10c9
na1
neath1234
newmoneywill7
newtonusa772
neydrica7
niko1955eua8
njmy001009
nkirote110
obbehi11
omeufuturocl
omonus43
oresty06
os8
osmarinaw2a
osvaldovaldez02
pabkuv31
pachamar15
pan20
pascoa3006
pasha358
pauliariasp
peka2
percychm
perpame2
pesas29
peugeot78lc5
pinovar
pras99
preciosom15

pscheng7d
qkq598
qrb88010
raffa1209
rahiniza10
ramses8824
raysa0806
rebelo116
renenatal5
rexal29
richbeauty3c
richmarkvangb
ridlc7f
riky0044
rkike7voip5
rlmartins5c
rob8
robert7878a
robpin20168
rosalba1234
rosalindatv187
rosefel4a
rosette404
rosiva1005
roxysteff15
royh8
rry022a2
rubysong04s3
s2e
sa1140
sagara3vp03
saifs2
samaem24
samann0001
sanlis2
santosbalvino09
saory2013jp6e
sara986
sasha1986m09
sbl0208
scasab
schamyp145
sebasjim18

sebasp3
serna04
shaareey1
shafici4
smarttlx4
solac
sonso32
stacecilia
sty026
successbiz42a7
successcall101
swane06
swj
sxh5670
taguaexitoso02
tcvision38
teamarraza13103
telexcameca
th3b3st
theav10
titacontreras7
todopoderoso810
tonlucca205
trome023
turbod2lwxe0
urias2015
usnewying007c4
v7pife001
vace1
valeusa100
valexa09
valmeidavoip07
vanesas05
vangie05
vegapo26
vinnytel10
visionariodeluz
voipmarcoplus
voipsantiago7
voipvalmm35
voipwhitee
vopiu103
voz1046

walfigueiredo
wendsonfonseca1
whq0006
wichol4tres
williamamaraluy
wilsonr691
wlp09
wsd1681
xagas7109
ygl666a8
ynwcu4tfwl1rr5i
yulys107
zeguedesmacvel
zouzhi004
zwh788803
zzy6679a4

Exhibit 2.11

The *rep_pwd_secondary* field values from *representante* for these 400 records, sorted alphabetically.

rep_pwd_secondary
00c66aaf5f2c3f49946f15c1ad2ea0d3
03435e70b2d275b16191f12f3b41df2f
043db66e8d0fc4512a0bc9f840f1554d
04973d1800f7c29d2b5daccdb01f31a7
0dec527c556b0354a28dafffc266063b
10f87c89f3fb79f8e86435548376a4a5
1106ef9ecd422a4e62912c68ac8d2df0
1200addfe20e301049ec29bec4d40331
149815eb972b3c370dee3b89d645ae14
1849c91d7b856fa80fb45b4130d6fd3b
235b567112edea1e02589e9a734126fb
263bce650e68ab4e23f28263760b9fa5
2ad5495bb5022419d894270f829b5125
2d95666e2649fcfc6e3af75e09f5adb9
2dcdd05ece6efa19f17274bbc1364330
2f4e613b3796bb305942e81621f1f088
2f8339d149a1ea2e112e47d9e29e74d4
2ff287f725a4e9d9d7c04b7a819c8f29
381ee61c5844233b757c8dbff6ba975d
382facfdafa0c490ec81be9b93b9dd1b
3d863b367aa379f71c7afc0c9cdca41d
3e106f801582b714ab41d5fa18e00b87
3ebcf2ea126331c42b12da5abb3da6b0
3f1ab41efddee120ba34f197705d1a01
42aa6053bc606f9645d3832265e9f548
45e1cc9e358bb3f36f7cbe74e5fc0eec
4919433b44e074dceda655cf5788a76e
4c91dd5327d6d3867566013261ac73a6
4ecc04528d462a27230d8334e79b24b6
509c5eca2b6523ce3719ddb1c2ab883d
56eb2fed43098c5f44e1a37ea007251b
575ea7987311a647196c8669b1c57b70
58c8221746f2c81a450c45f48adec438
58ca638d0bc81650ed425b23fa0a0567
59e020e1b521106aec6cdaf41f5b9ee4
5a01f0597ac4bdf35c24846734ee9a76
5e8667a439c68f5145dd2fcbecf02209
5f14615696649541a025d3d0f8e0447f
5ff2f237f9e624088563a70efe2528ca
611de0b10a10b9fae5f3a9bdabeb04fb

62e201f3d00d21ca737fde0e4ed9b545
6649c7f2602777662f075f0d14d84b35
677ff5f13ed72ab9580c7a61af6f1645
680f53fc72a58404c6e6256c48bf753a
68d47877655c857fedb09e7b7afe7af0
6b9db1f28f1e2ed917a49b3f6dcbd42b
6e0d0cdf9e48538255bc003a8587e6d0
71d6c631e36a3432a5899bb09c51d52a
734c085fb7ebb2b95b7e0e82d452d837
749c6b091e2d50b185c593c1c679634e
78c4318cd553838269207ece540b906f
7dadbe5a37682c50d0600552bc389b33
800e365301349f2b62fc5bcbf6ac2b39
802d8c9b92946f3e6b368312cd1383e1
81e3b96f6932b28a4074797685dbd8eb
838ef0e61e11ca0469110b10d157ca81
8ce87b8ec346ff4c80635f667d1592ae
8e1ee624dc8676881fdc12cb002d8708
90ce1abf15835bcdf2326b31f8913bda
92f632fea36a4ddaebe79564dbe6029f
934b535800b1cba8f96a5d72f72f1611
935f099f5e344faddbae2ad5d08b681a
9551f54626799ab0f93c314eaadaa1c0
9fe6f4d1ef8fab724cc1779b732a042d
a08e91f1a15579e233ce4854ed66efe2
a14523e002a82deb3b41ab9ab2a7f01d
a3eda7c758497f437c166ab6f568983b
a431d70133ef6cf688bc4f6093922b48
a947d4cd95e5d0c30f0420b21bca3b27
ac802e53178ac650a139574c534f872c
ade7c1dc609408b28708a9d241f17889
b802b8d91fd69363661f43ae17fca18f
b918cc34caecf0777981414c38908144
c0bb9bbaf56198988ba3a662ff7450d3
c1f2c71ded85e3a3eb6fe1c5b690417e
c54f739fe53a1f36ae7c15b21e6bda85
ca6949e70eabe8adbc8632697f6b988f
ca81257d0a041dd986e9972ec6f4e225
d0970714757783e6cf17b26fb8e2298f
d43070b938bdae4eda7f980f67a44e30
d5e29feb654a6681f1e29c3ff36a3cfa
d7f3c343f886ee24f320aa8a737747d1
d87febe46d5519ef1874c36eacd46f2e

d93591bdf7860e1e4ee2fca799911215
daffe63d7b852b29209b0ecb668c8525
dceeb85ee6ef9808a39800fe69f64114
e02676412b538fe93f4a90805848bdf8
e10adc3949ba59abbe56e057f20f883e
e489562a6c8de9769278c6cee58c47e3
e6f2e66a147445dfb2f26048e4781e29
e83e154707cd1a7f351a5ac05026ad8e
eaca99ab01438061f06b232dde801fc5
ec038f4aef0cdd8eb081faf735e1ab89
ec5bac90e1f795e4f0ef8d140e24e34a
ed07470d794db05903a5de1e1cc64043
ee8b9658fd8e7133494ec5be4dbe897f
f1fb1ea9a408b1f321062e86ec4bae3e
f1fb1ea9a408b1f321062e86ec4bae3e
f9fbe35b8769c57b7b324c8ec0fc59cd
fb5f2f27be2de104ac2b192f3e874dda
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None

None
None
None

None
None
None

None
None
None

None
None
None

None
None
None

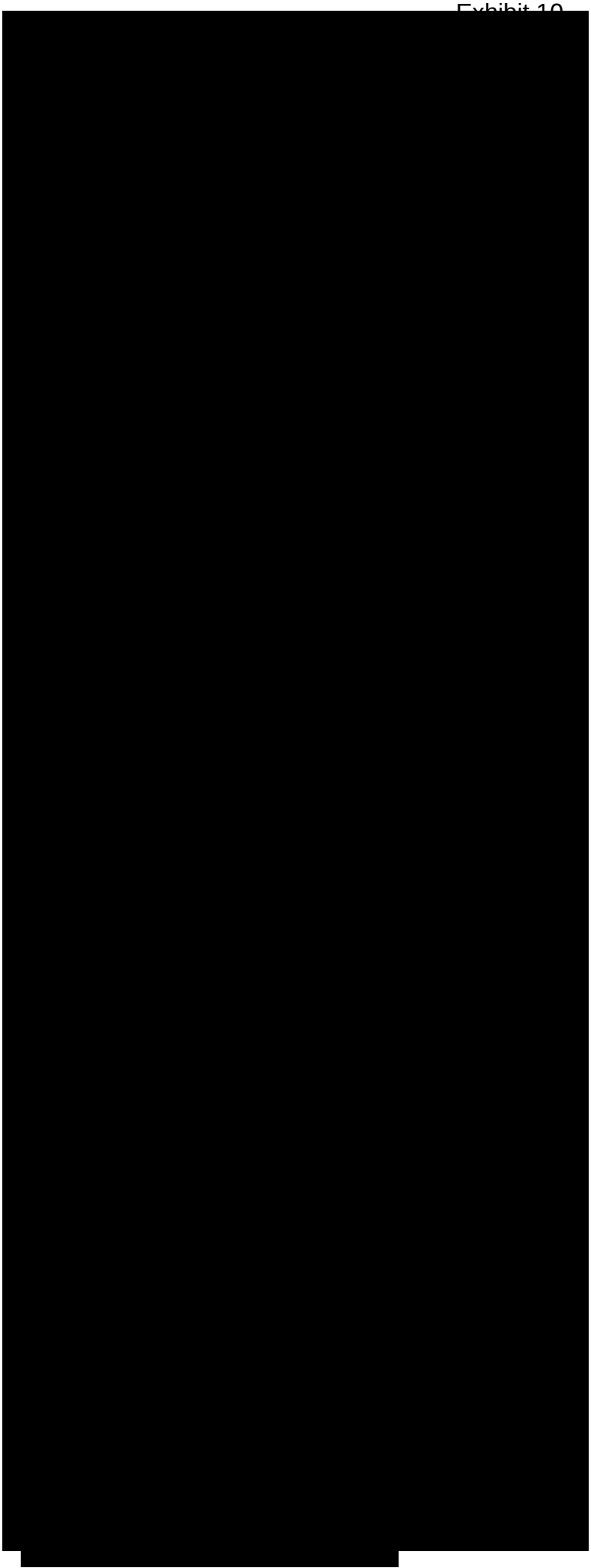
None
None
None

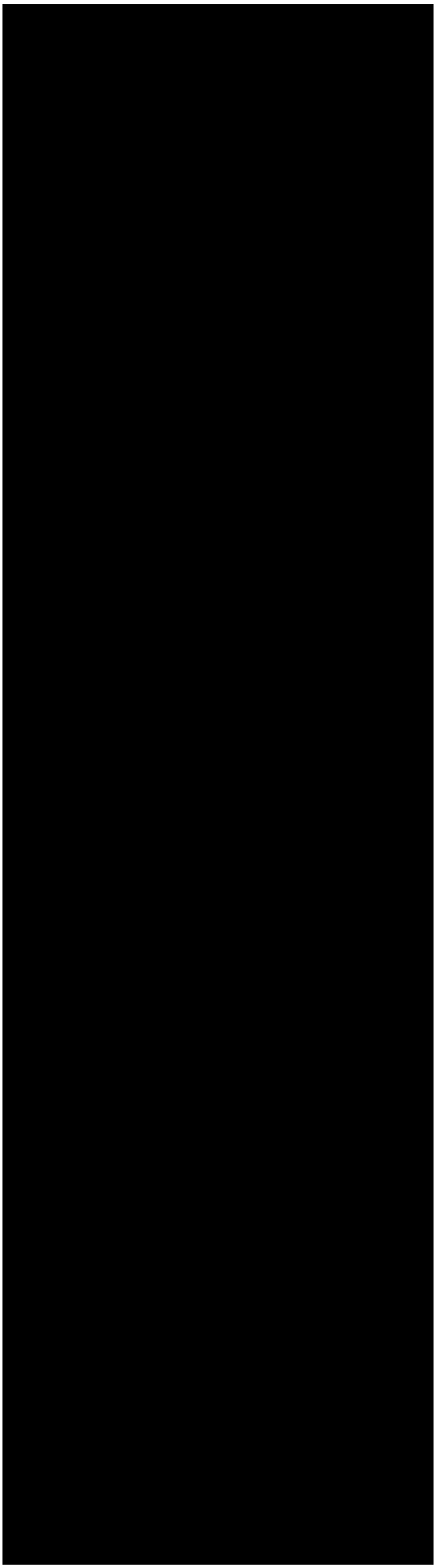
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None

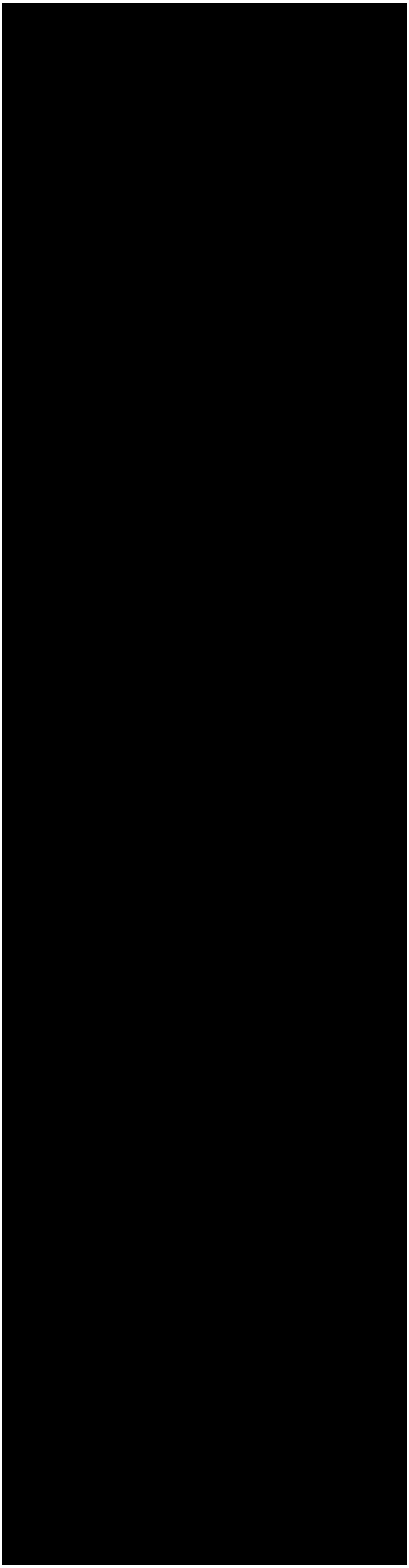
Exhibit 2.12

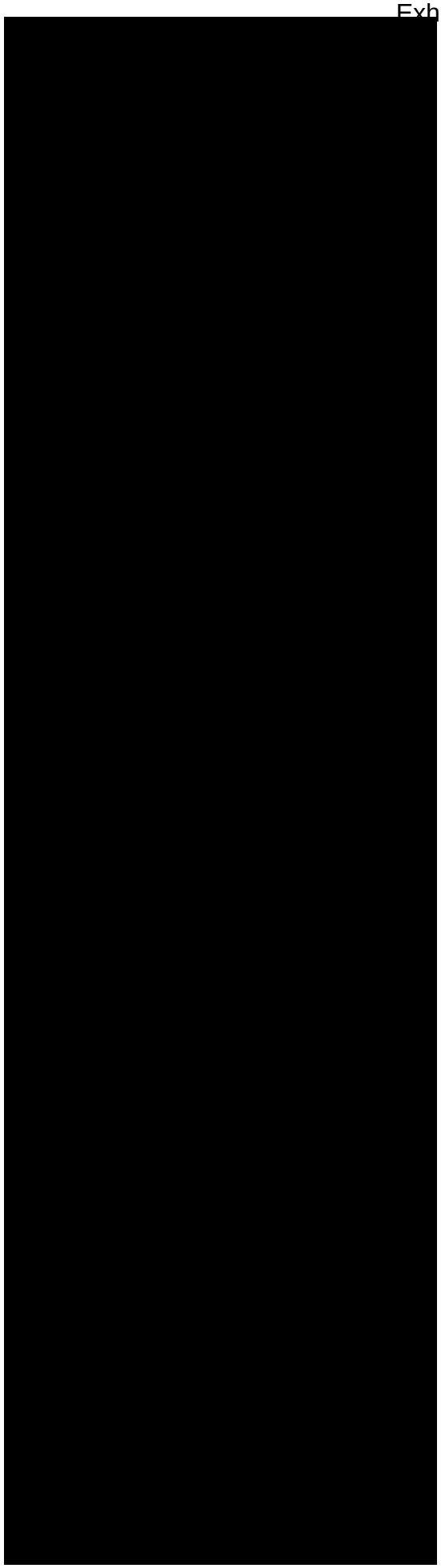
The *rep_cpf* field values from *representante* for these 400 records, sorted alphabetically.

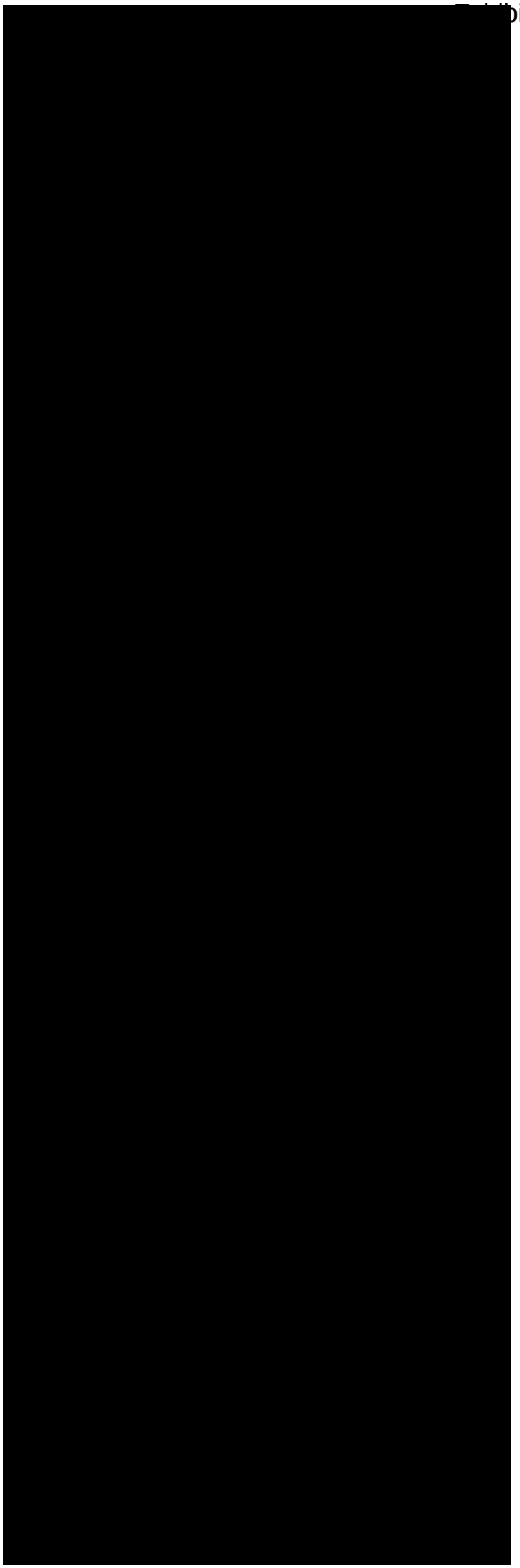
rep_cpf
[redacted]
********6
,...,.,..
.....[redacted]

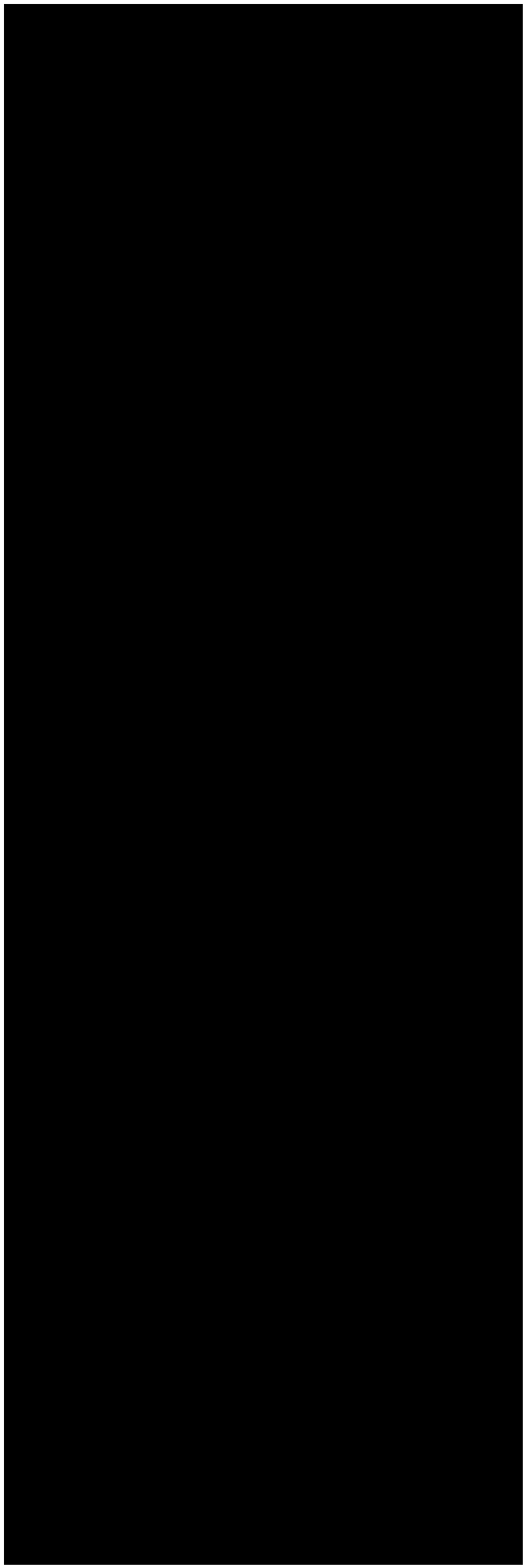


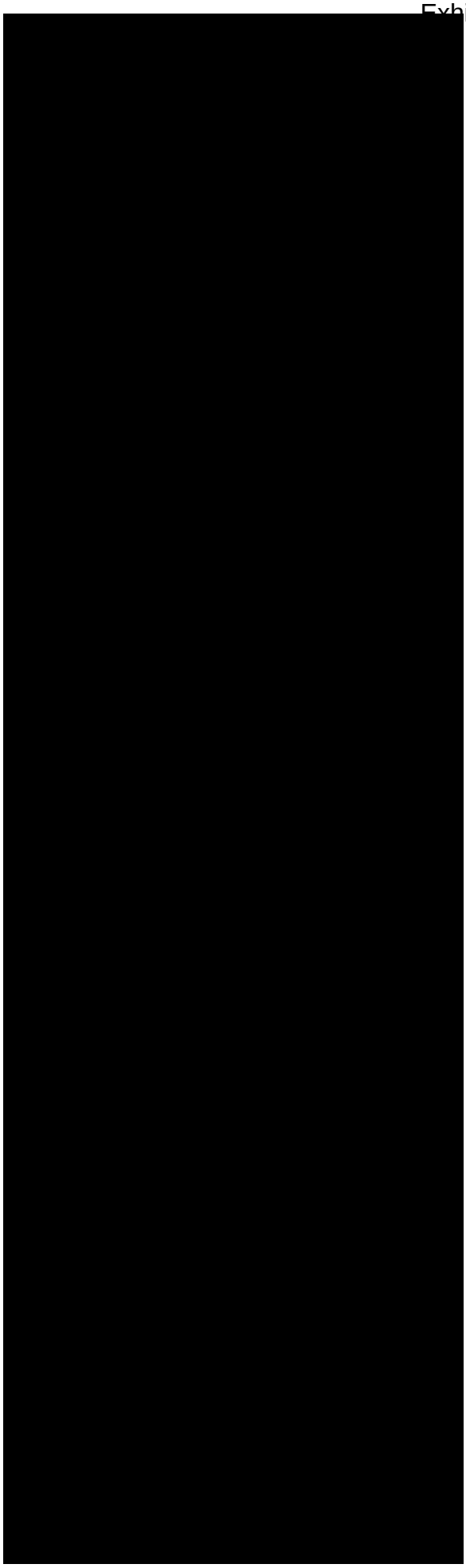














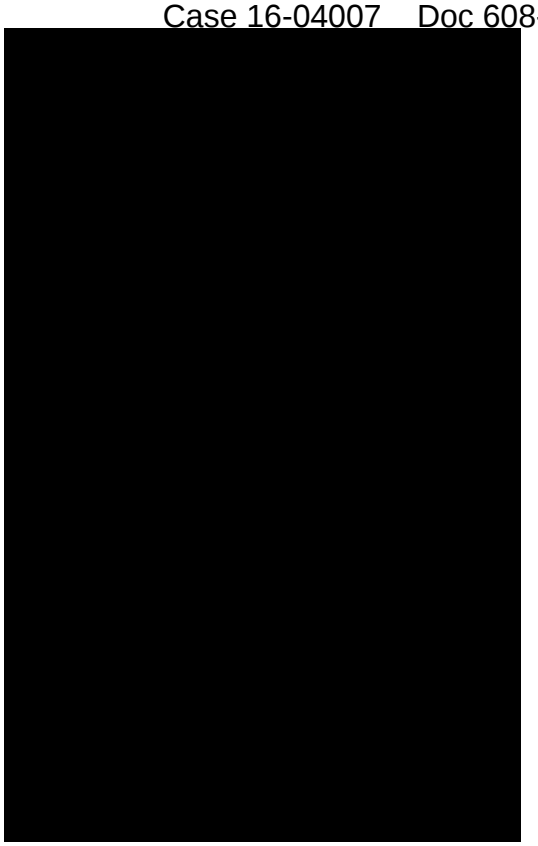


Exhibit 2.13

The *rep_datanasc* field values from *representante* for these 400 records, sorted chronologically.

rep_datanasc
1947-10-26
1948-03-06
1954-06-19
1957-05-09
1961-02-23
1961-07-15
1962-01-06
1962-08-26
1963-05-05
1964-06-24
1965-02-09
1965-05-18
1966-11-10
1967-01-16
1967-06-14
1967-09-06
1967-09-09
1968-05-06
1969-02-03
1969-03-02
1969-05-15
1969-08-24
1969-08-29
1969-09-23
1970-10-05
1971-03-13
1971-08-05
1971-11-15
1972-01-15
1972-02-05
1973-02-08
1973-10-23
1973-12-30
1974-08-24
1974-11-02
1975-01-01
1975-04-04
1975-04-19
1975-05-30
1975-07-10

1975-11-09
1975-11-21
1976-05-09
1976-06-23
1976-07-19
1976-10-12
1976-12-01
1976-12-17
1977-01-12
1978-01-01
1978-03-09
1978-07-12
1978-08-24
1979-04-25
1979-10-01
1980-05-01
1980-06-30
1980-07-22
1980-08-20
1980-09-04
1980-12-18
1981-02-28
1981-10-24
1982-04-14
1982-06-19
1983-02-04
1983-08-19
1983-10-23
1985-05-18
1985-11-07
1985-12-22
1986-04-24
1986-07-03
1986-07-13
1987-03-25
1987-07-28
1987-09-22
1987-12-06
1988-02-17
1989-02-02
1989-03-07
1989-08-08
1990-08-12

1991-03-29
1991-04-13
1992-02-22
1993-02-09
1993-03-08
1994-08-16
1996-01-26
1996-10-04
2013-01-28
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None

None
None
None

None
None
None

None
None
None

None
None
None

None
None
None

None
None
None

None
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None

EXHIBIT 3

Exhibit 3

The *rep_nome* field from *representante* for 300 records chosen randomly with fixed random seed, selected randomly among those 97% of records for which rep_nome is 5 characters or longer and contains only the alphabetical characters or period, comma, hyphen, space, or apostrophe. Highlighted values of *rep_nome* were deemed to not look like a name.

rep_id	rep_nome
27827474	jack zheng
21639603	alcebiades de souza neto souza
11143629	olger hancco condori
13347935	JESUS PEREZ
12439637	jose correia da silva
26556190	Francisca Orlanda de Lima Ferreira
21123269	long fang zhang
29846120	lizardo juarez apaza
25261112	Gustavo Maroso Gessi
24654528	alex malo
26833110	Rosa Urena
14643515	SERAPIO ACURIO RAMIREZ
27588446	javier Patino
12873725	Clemente Geronimo Viasalla
30211184	Nina Mazhaylova
14748375	ariel hoyos lozada
29975304	gilmar de jesus silva
10858007	eraldo bolan
17801155	wilson lima
14030641	Natalia Vasilyeva
30989782	John Williams
25346644	eloy flores
27204416	luis g jimenez
32305954	enrrique guizado
31917132	iracema laudelina da silva
21664499	cua cua xia
17298997	Juan Cuesta
33146908	sd lkjhgf
17197597	osorio mesut
25726838	claudia potgieter
13488927	marlene polanco
19547891	gnjli zhang
22189353	decio andrade
16158397	yefer turpo llanos
30483498	Claudio Emmanuel Brito Gomez
20750909	yanping li
33826742	michel p
26727212	jeferson daniel freitas
12554677	patricia salazar valenzuela
11365003	Orestes Boggiano
28545600	Igor Kotenko
31314930	gregori santer
33723788	deniz judy mamani ccasa

14451789	Javier Villanueva
29677522	Efemia Carmen Cruz Cahuana
15265797	Nelson Santos
19468545	jianzhe huo
13315929	Wanderley de Carvalho Silva
10530005	Carlos Abel Salas Davalos
28518052	Carolina Calderon Gil
10722643	Yuldash Gafarov
27442684	Phayfer Souza Sutil
30111168	tinoco rodrigues menezes
22851811	evaristo mamani mamani
21045251	carlos perez
23671674	rifei teng
10446609	Robertas Rimkus
21729879	Eulises Heriberto Monserrate Bravo
15846211	fuyun ji
14994545	wendy Arteaga Cruz
26411026	HUBERT CORAZAO TEVEZ
14253205	Claudio Bueno
24260454	daju li
32010776	nelson paniagua
23644630	renee montalvo medina
10616271	Ainomugisha Diana
17330205	isabel dominguez leon
23777936	kuiying liu
25041340	yong jiang
29945982	Sergei Ulyanov
23118721	denny pacompia zapana
20614081	nelson navas
19832919	wilmer escobar
22118085	Roberto Lacouture
19702065	dava chanta
18726741	daniel rogelio suca jara
28780038	JOSÉ EDUARDO SEDOSKI MENDES DA SILVA
23804438	kelun tian
26403090	aleida aragon cocoma
28694740	marco carvalho
11102417	Concepcion Monarrez
26138348	ian kimani
14216493	João da Silva Junior
28132220	MARIA NEVES
27193598	sergio damian
22579909	ricardo oliveira
23884394	mario ribeiro
18772447	luis cruz
20766673	agnaldu fritz
23242129	Virginia Reyes
15856237	Berto Genao
10370785	oscar Hernández Jiménez
32307940	John Williams
18904695	rocio ydme torres
28172270	francisco divino tavares

13112507	gelvane pereira da silva pereira
19333869	tyhour pheng
10476541	welliton melo
12696947	ruixian wang
24719078	kefeng zhao
11136703	pascualita jaen segarra
23578064	nixon aldaz
33416008	alexandre souza
25668562	Nelly Jluissa Giron
19151285	Eulalia Huanca Quilla
33934712	leo da cunha
33468562	Joshua Omaruic
25935690	JULIE HER
30182938	valentin barrera
33944108	mario alexander melo
16628125	clariza pareja flores
27818372	julio fenandsdes
21455879	angel meza hidalgo
27553662	JOHN AFEDZIE
30853710	luis nunez
30128138	Don Yang
10843919	Matt Modica
31804596	reyna trujillo layme
13111583	gina pierre
28101274	Houdy m
31828564	elizabe sanchez
32314920	Luz Gloria Uribe Vélez
18656911	jorge gallardo
23650506	DONG HAE
15292719	Viktor Cherepanov
33836416	Sunday Ayodele Ajayi
21048495	dora escobar
21507093	mingyue qu
16920799	Paula Espinal
28454984	Jon Wing
24027588	Jenny Olivares
26561370	Victor Mendoza Zavaleta
13883327	lala sasa
33931384	João Melo
10911121	hernando gallo
18186615	MEI XIAO
13579601	dave saka
21706639	du du
12781019	igor morais avelino matias igor
27073276	indalecio ñahui calle
27952704	ling shan
31822950	marcos rodriguez
24278798	nadezhda kalugina
24857390	Adriana Bettencourt
14272457	lidia avila
23709932	Herika Patricia Guzmán Gómez
10856257	Luis Antunes

30284334	Renato Domingues
31613346	jorge rivera pereira
18633033	Waydene De Paula Abreu
22112953	CINALDO MENEZES SOUZA
26038110	carlos estacio
15116109	Alexandra Ortiz Ruiz
21530799	jihua xia
16247449	AIDA ORTEGA ZHUNIO
33720942	QUIRINO NUNES
30985344	nuri vargas reyes
15152611	walter blaudt da vitoria
33039288	diego cahui quispe
21081955	Lin Wong
30309446	roberto cadeddu
22366803	Ruben Solano
29498504	bin he
30794978	xxxxxxxxxxx xxxxxxxxxx
17351141	yankson arevalo
11808979	clemilson
23577506	Thiago Andrade de Souza
24486278	Antonio Pereira
17631747	Khun Teck Yap
11465347	delia murrieta lopez
25599514	xixi la
15878913	jesus botelho
27374278	Thiago Ricardo Nunes Silva fe Oliveira
32549964	julia ramos
30086112	luiz carlos da silva
26667364	daeana souza
29220588	John Williams
18823371	mika monn
13945879	rosalie awat montero
25699416	jose natanael gomes
23388041	Silvia Padilla
11035363	Soraya Ferreira
22404535	bin li
31862488	betulia flores
32563234	marcone borges da fonseca
14060779	Luis Dumitru
6386218	Jarolin N De Jesus Hannia
30422776	Lucia Vargaya Paiva de castillo
12440603	Nelson Pereira Santana
21451229	eugenio mercedes gomez
29753068	orlando dadalt
30847112	Joao Baiense
15215641	Ana Clara Noberto Galdino de Araujo
33035076	Barbara Sousa
27999502	ming hua
20260595	laura santafe
15040785	Abraao Santos de Jesus
29721888	John Williams
32772146	lazarom Mejia

25951124	andre FRANCISCO RIBEIRO MENDES
29671862	mengqing ma
11425687	Noemi Kisorssy
18175887	Oa Chamroeun
26244836	juan stiven giraldo
31620552	LORENA PAOLA soriano sanchez
22300533	Fernanda Paredes Calatayud
19061079	fen li
19575369	Allan Valeriano Chavez Mejía
29902974	Bobby Brown
24611576	fabian carrion
15266363	mihai vlad
14109257	gustavo raniele fontes silva
25228946	anselmo vicente castillo
26907736	yongqin li
13173065	Maria Alvaro
22627643	luisa braganca
26528936	antonio goncalves
11722069	kule ibrahim
21205459	elio escalante
29015204	Monica Martelo
18976735	maria guillermina huisa calloquispe
26186662	alan apaza mamani
28865286	maria teresa barreto pita sousa
14817049	elson ferreira
10342005	Sharon Couch
15075999	CARLOS EDUARDO HUAMANI NOA
11144859	jinmei wei
30915146	Murilo Fernando Batista
18936763	Marcelo Yol Ajuchán
31006768	John Williams
17740871	Jose E Landaverde
27990486	baihe guo
15590039	flavio torres
12244139	juan ceron
13381807	Carlos Chavez Mazache
23822232	jian chen
27503384	Ralph Njau
19729071	Chhay Rat
19944259	EVELYN REBECA PASCACIO EGUSQUIZA
21591595	David Ornelas
24089362	Pamela Alejandra Diaz Osorio
13043869	Maria Matos
14130403	Manuel Álvarez Miño
24705260	hangbin yu
12775181	Leandro Germano Mendes
18782041	diego armando bautista esquivel
16568641	Paulo Santo
23518442	Marioza Menezes Farias
30099918	santo soto
18967203	Hugo Dalto Sanchez Leon
21625397	uvestor Ojobin

21964629	lijuan tong
20846267	silvani brais de souza
20562689	Marcia Meira Sellmann Moreira
26792974	jó santos
25069386	ngu deru
13302037	Constantino Meliton Ruiz Maldonado
26545916	cipriano dias
23136141	ernesto antonio duran polanco
14930517	YOHON RONAL QUISPE FERRO
15984601	Minea Ngam
33386438	Victor Sosa
12628197	TABULA HASSAN
28884248	dina viveiros
25064818	abel pilco huahuacondori
19815891	LARRY KING LIVE
15665339	aaaaa aaaaa
20459057	Jvan Di Stefano
32503118	pharrell bass
19186679	Aleksandr Domilovskyi
19692883	Ly Fang
17629131	Yevgeny Chernov
31638120	MARILENE PAULINO VIEIRA
17253715	Rosa Ortiz Lopez
31647898	mayra romero
12528333	mayker hidalgo sanchez
32116632	tatiana vitreniuc
16230235	Donald Rachels
16283257	sonia capinha
30256280	beijun ma
24349640	antonina jaime
14081611	maggy nehi
13189637	pukul dua
23562972	ronald nogales
30023688	erasmo barroso
5842880	Darrell Hunter
33089548	VOIP VOIP
18618529	marcos carrasco bustamante
25217686	svgb wang
20213561	hans sosa leon
14098359	Lizbeth Huerta
17982383	lightning carrier
14015255	MARINA ROIK
28219782	shirley katherine napa paz
22966851	Aleksey Badmaev
18571375	John Gutierrez
13784085	Fernando Avila Barros
33711134	olga aguirre
17796921	Natalia Karume Hayashi
26024538	Jack Zheng

EXHIBIT 11

Originally posted: 2022-10-14. View source code for this page here.

The Intuition Behind the Use of Expectation Maximisation to Train Record Linkage Models

How unsupervised learning is used to estimate model parameters in Splink

Splink is a free probabilistic record linkage library that predicts the likelihood that two records refer to the same entity. For example, what is the probability that the following two records match?

first_name	surname	dob	city	email
lucas	smith	1984-01-02	London	lucas.smith@hotmail.com
lucas	smith	1983-02-12	Manchester	lucas.smith@hotmail.com

The underlying statistical model is called the Fellegi Sunter model. It works by computing partial match weights, which are a measure of the importance of the different columns in computing the probability of a match. For example, the match on 'Smith' is less important than the match on email in the above example.

However, there's a paradox at the heart of the problem of training a record linkage model: To estimate the parameters of the model, you need to know which records match. But finding which records match is the whole problem you're trying to solve. Manual labelling of comparisons is error prone and costly so isn't a great option.

The Expectation Maximisation (EM) algorithm provides a solution to this paradox. This post provides an intuitive explanation for how this works, and some details of how it's implemented in Splink, our software for record linkage at scale. For a more formal treatment, see the FastLink paper.

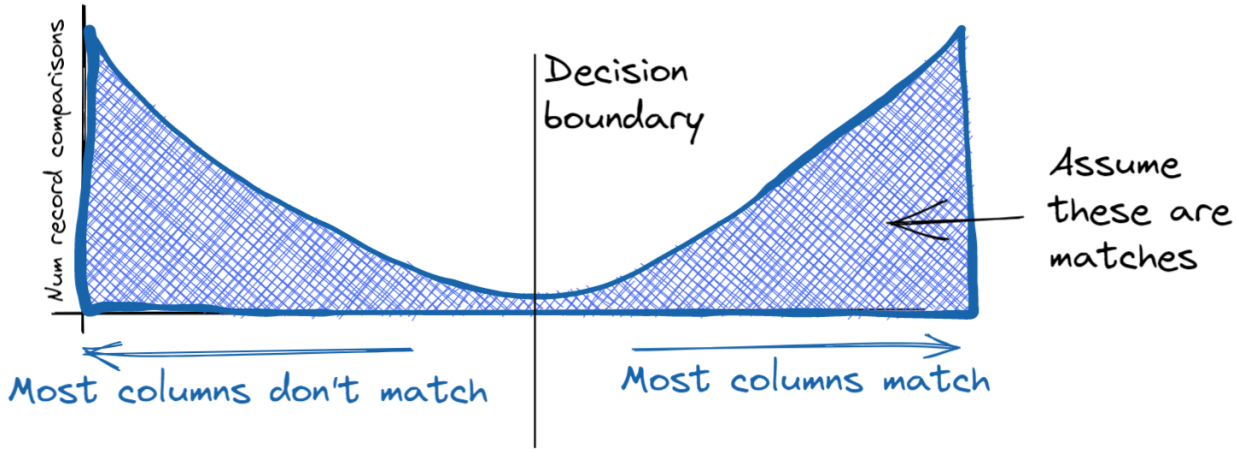
The EM Approach

The parameters of a record linkage model — the m and the u probabilities — can be calculated from the aggregate characteristics of matching records and non-matching records respectively. (For a fuller treatment of m and u probabilities, I recommend reading this blog post.) Once these values are known, the model is usually able accurately to predict which records match.

The key to the EM approach is leveraging the fact that matching and non-matching record comparisons form two fairly distinct 'clusters':

- Amongst matching record comparisons, the information in most or all columns in the two records usually matches.
- Amongst non-matching record comparisons, the information in most or all columns in the two records usually does not match

This results in a bimodal distribution of record comparisons:



This separation between matches and non-matches means it's relatively easy to come up with a rough decision rule that is able to correctly identify most matches and most non-matches. For instance, we could simply count the number of columns that match (first name surname, and so on), and predict a match for any record pair where more than half of them match.

This is fundamental to resolving our paradox: We've just found a way to designate matches and non matches without any labeled data (i.e. unsupervised learning).

Now that we have designated each record comparison as a match or a non match, we can use this information to estimate the implied m and u parameters of the model. And this enables us to iteratively improve this decision rule, using the following algorithm:

Expectation step: Use the estimated model parameters to predict which record comparisons are matches

Maximisation step: Use the predictions to re-estimate the model parameters

In this example, our initial decision rule worked by simply counting the number of columns with matching information. As such, we gave equal weight to every column. Clearly there is room for improvement in this rule — in reality a match on a postcode column is more important than a match on gender.

In the first maximisation step, the expectation maximisation process will immediately begin to learn these weights. For example, it will recognise that, amongst non-matches there are many record comparisons which match on gender, but few that match on postcode. This will be represented in the re-computed m and u probabilities, and in the next iteration, greater weight will be placed on a match on postcode than gender.

At each iteration, the decision rule will be refined as the model makes fewer mistakes, and the weights therefore become more accurate.

Convergence occurs when the parameter estimates stabilise, and there are no longer any changes to parameter estimates between iterations.

It turns out this algorithm is equivalent to maximising the likelihood function of the Fellegi-Sunter model (although note that it is only guaranteed to converge to a local maxima, not necessarily a global one).

A caveat around the decision boundary

In the above, to build intutiion, I describe concrete decision boundary that classifies each record comparison as either a match or a non match. The m and u values calculated in the maximisation step are based on the counts of matches and non-matches that result from this binary classifier.

In reality, we can improve on this, because using a decision boundary throws away some of the information in the match probability: it treats an uncertain match in the same way as a highly probable match.

There's a simple modification to the methodology that accounts for this. We can use the match probability directly, rather than applying the decision bondary. For example, if a record comparison has a 70% match probability, in the maximisation step this can be treated as 0.3 non-matching records and 0.7 matching records. So when the m and u probabilities are recomputed in the maximisation step, we use the sum of probabilities rather than counts of matches and non matches.

How the EM approach is implemented in Splink

In Splink, a few modifications to the above process are made that, in practice, result in faster convergence and a greater likelihood of convergence to a global maxima.

Estimating the u probabilities

Whilst it is possible to estimate u probabilities using the EM approach, we recommend an alternative estimation procedure in Splink which is usually more accurate.

Recall that the u probabilities represent the probability that we observe matching columns amongst truly non-matching records. For example, how often do we observe the coincidence that two different people have the same first name?

We can estimate the u probabilities by taking random pairwise record comparisons, assuming they do not match, and computing how often these coincidences occur. Since the probability of two random records being a match (representing the same entity) is usually very low, then this will yield good estimates of u values.

An additional benefit of this approach is that if the u probabilities are correct, it 'anchors' the EM estimation procedure and greatly improves the chance of it converging to a global, rather than local, maximum.

Starting values and the initial decision boundary

In the above, we suggested a simple decision rule for the first step of EM: count the number of matching columns, and designate as a match any comparison where more than half of fields match.

In reality, it's possible to begin with a decision rule which is usually more accurate, leading to faster convergence. Specifically, the initial decision rule uses the u probabilities that have been estimated using the random sampling procedure described above, and 'sensible default' starting values for the m probabilities .

You can see all of this at work in practice in some of our example notebooks.

Conclusion

EM is a powerful tool for estimating the parameters of probabilistic linkage models in the Fellegi Sunter framework. In our experience, the EM approach works best when it is used to estimate the m probabilities of the model only, following estimation of the u probabilities using random sampling, and estimating the prior using a direct estimation techique.

Probabilistic linkage home Next article >

Back to homepage.

This site is built using Observable HQ and Gatsby.js. Source code here.

EXHIBIT 12

Deposition of Joshua W. Dennis

May 9, 2023

VOLUME: 1
PAGES: 1 - 209
EXHIBITS: 1 - 11

UNITED STATES BANKRUPTCY COURT
DISTRICT OF MASSACHUSETTS

Chapter 11 Cases
14-40987-EDK
14-40988-EDK
14-40989-EDK

Jointly Administered
Adversary No. 16-4006
Adversary No. 16-4007

In Re:

TELEXFREE, LLC, TELEXFREE, INC., and TELEXFREE FINANCIAL, INC.,
Debtors,

STEPHEN B. DARR, TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC., and TELEXFREE FINANCIAL, INC.
Plaintiff,

vs.

FRANTZ BALAN, A REPRESENTATIVE OF A CLASS OF DEFENDANT NET WINNERS,
Defendants.

STEPHEN B. DARR, TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC., and TELEXFREE FINANCIAL, INC.
Plaintiff,

vs.

MARCO PUZZARINI AND SANDRO PAULO FREITAS, REPRESENTATIVES OF A CLASS OF DEFENDANT NET WINNERS,
Defendants.

CURRAN COURT REPORTING

DEPOSITION OF JOSHUA W. DENNIS, a witness called on behalf of the Plaintiff, Stephen B. Darr, taken pursuant to the Federal Rules of Bankruptcy Procedure and Federal Rules of Civil Procedure, before Jacqueline M. Curran, (CSR #134293), a Registered Professional Reporter, Registered Merit Reporter, and Notary Public in and for the Commonwealth of Massachusetts, at the law offices of MURPHY & KING, P.C., 28 State Street, 31st Floor, Boston, Massachusetts on Tuesday, May 9, 2023, commencing at 10:00 a.m.

CURRAN COURT REPORTING
21 Rowe Hill Road
Stoneham, Massachusetts 02180
(781) 279-8400 (telephone)
E-mail: currancourtrep@yahoo.com

CURRAN COURT REPORTING

APPEARANCES

Representing the Plaintiff, Stephen B. Darr, Trustee of the Estates of TelexFree, LLC, TelexFree, Inc., and TelexFree Financial, Inc.:

MURPHY & KING, P.C.
By: Daniel J. Lyne, Esq.
Alexandra M. Papas, Esq.
28 State Street, Suite 3101
Boston, MA 02109
(617) 423-0400
Dlyne@murphyking.com
Apapas@murphyking.com

Representing the Defendants, Frantz Balan, the Class Representative in the 16-4006 action, and Sandro Paulo Freitas and Marco Puzzarini, Class Representatives in the 16-4007 action:

MILLIGAN RONA DURAN & KING, LLC
By: Ilyas J. Rona, Esq.
28 State Street, Suite 802
Boston, MA 02109
(617) 395-9570
ijr@mrdklaw.com

CURRAN COURT REPORTING

INDEX

WITNESS: Joshua W. Dennis

Examination By:	DIRECT	CROSS	REDIRECT	RECROSS
Mr. Lyne	5			

EXHIBITS

No.	Description	Page
No. 1	Notice of Deposition of Joshua W. Dennis	5
No. 2	Supplemental Rebuttal Expert Report of Joshua W. Dennis	5
No. 3	Borelian Aggregation Methodology Code Execution Instructions	15
No. 4	Joshua Dennis Curriculum Vitae	33
No. 5	Page 43 of Dr. Freer's Report Entitled Steps in the Aggregation Methodology	67
No. 6	Expert Report of Dr. Cameron E. Freer	94
No. 7	One-Page Document	131
No. 8	Document Entitled Exhibit 2	132
No. 9	Document Entitled Exhibit 16	158
No. 10	Document Entitled Exhibit 13	178
No. 11	Document Entitled Exhibit 9.5	189

(All original exhibits retained by court reporter and scanned and e-mailed to all counsel; hard copy mailed back to Attorney Lyne)

CURRAN COURT REPORTING

5

(Deposition Exhibit No. 1, Notice of Deposition of Joshua W. Dennis; Exhibit No. 2, Supplemental Rebuttal Expert Report of Joshua W. Dennis, premarked for identification)

JOSHUA W. DENNIS, having been satisfactorily identified by the production of his driver's license, and duly sworn by the Notary Public, was examined and testified as follows:

DIRECT EXAMINATION

By Mr. Lyne:

Q. Mr. Dennis, my name is Dan Lyne. I represent the trustee in bankruptcy in the Telex proceedings -- multiple proceedings. I assume you've had your deposition taken before?

A. Not strictly a deposition. As part of the -- earlier we had the evidentiary hearings --

Q. Correct.

A. -- that was not truly a deposition. In addition, I've done trial testimony, but

6

not a deposition like this.

Q. Interesting. So let me give you the ground rules. I'm going to ask you a series of questions. If at any time you don't understand my question, please tell me, and I'll try to rephrase it so you can understand it. To the extent you answer my question, I'll assume that you understood it and that you answered it fully, completely, and truthfully. Agreed?

A. Agreed.

Q. One of the things that's important is to wait for me to answer or -- excuse me. Let me try that again.

One of the things that's important is to allow me to finish my question completely before you start to answer. This is a courtesy to the stenographer, and when you answer, please answer with words, not with gestures or Uh-huhs or Uh-Uhs. We're trying to make sure that at the end of the exercise, we have as clean a transcript as possible.

MR. LYNE: Ilyas, do you want to have your client read and sign?

7

MR. RONA: Yes.

MR. LYNE: Thirty days?

MR. RONA: Do you think -- is 30 days enough time?

THE WITNESS: For the errata?

MR. RONA: Yes.

THE WITNESS: Sure.

MR. RONA: Okay. And can we waive notarization?

MR. LYNE: Waive notarization. Do you want to waive all objections, except as to form, or do you want to waive objections generally, or what do you want to do?

MR. RONA: We'll do, I think, usual. Reserve all objections, except as to form, and motions to strike until the time of trial or evidentiary hearing.

MR. LYNE: Whatever it may be.

MR. RONA: Whatever it may be.

8

Q. Sir, directing your attention to Exhibit 1, are you appearing this morning pursuant to the notice of taking deposition which I've just shown you?

A. I am.

Q. And am I correct in understanding that you have been engaged by Mr. Rona's office to provide some form of analytical work in connection with this case?

A. I think that's fair.

Q. Okay. And can you describe for me what the scope of your engagement is in this case?

A. Sure. So the scope of my engagement, as I laid out in my expert reports, was to review and assess the methodology and the calculations that were put forth by experts on behalf of the trustee in terms of the determination of the net winners and net losers.

Q. And who assisted you in connection with this exercise internally at StoneTurn?

A. Sure. So do you want me to kind of go back in time here? We were engaged quite

some time ago, and I've had two reports here, so I just want to make sure I'm clear on --

Q. All right.

A. -- which report you're referring to.

Q. So why don't you -- why don't you tell me in connection with the 2020 report, and then I'm going to ask you the same question with respect to the current report.

A. Understood. And that's what will test my memory here --

MR. RONA: I'm sorry. "2020 report." Is that the Martin report?

MR. LYNE: That's his report --

MR. RONA: Okay.

MR. LYNE: -- commenting on the Martin report.

MR. RONA: Okay.

A. 2020 sounds about right for that.

Q. Yes.

A. So assisting me in that one would have been Patrick Ryan and Gladys Leung, L-E-U-N-G. I believe those are the two primary people.

Q. And were you either of those people trained in the area of data science?

A. Yes.

Q. All right. Who of those two, or were both of them data scientists?

A. Patrick Ryan had his master's in data science.

Q. What about Gladys Leung, what's her background?

A. She helped more on the intellectual property, the economic damages side of the house, so she helped in preparing more of the exhibits. You know, reading the report for typographical errors, things like that.

Q. Okay. And the current 2022 report that you prepared, who worked with you on that?

A. So that would be Charles Soha, S-O-H-A, as well as James Hubbs, H-U-B-B-S, and Nicole Stedman, S-T-E-D-M-A-N. I think I got that right.

Q. And are any of those individuals data scientists?

A. So just to be clear on this -- I can

give you their background in terms of -- so I'm thinking to answer as data scientists. Let me give you their education.

Q. Okay.

A. So Charles Soha, I believe, has his undergrad degree in math -- mathematics, I believe, and had continuing education, I believe has his MBA. James Hubbs, I believe his undergrad degree was in math statistics and has continuing graduate coursework in the same.

Q. And Ms. Stedman?

A. She's similar to Gladys on the economic damages side of the house, and, again, assisted largely with, you know, reading the report for confirming footnotes, typographical errors, and things like that.

Q. So am I correct in understanding -- well, let me back up a bit.

Do you have any formal training as a data scientist?

A. I do have formal training in the areas of data, yes.

Q. Okay. And what is your formal

training?

A. So as part of my undergraduate degree -- well, my major was in management business, my minor was in computer science, and since that time, I've had continuing coursework as well in the area of data science.

Q. Okay. So I'll get to that. Is it fair to say then that you've had technical support from data scientists or somebody with statistical or mathematics background in both -- in both of the reports that you have signed?

A. Yes. That's true.

Q. All right. And did you rely upon, for example, Mr. Ryan in doing analytical work with respect to the 2020 report?

A. What do you mean by "rely upon"?

Q. Well, you tell me how you used Mr. Ryan in two thousand -- in the 2020 report.

A. So going back a ways now, my recollection is what we received from Mr. Martin in here on the time involved some

13

fairly large databases. I think we received the original large copy of the SIG, S-I-G, database. I recall at least he helped me load that database onto our servers. I think he also helped me review some of the code I wrote to help QC and confirm it.

Q. You mean to QC and confirm whether the SIG database was loaded correctly or otherwise?

A. No. No. I'm sorry. I'm sorry. He helped me load the data, and then separately, I performed an analysis based on that data, and he reviewed the code I wrote that was in support of my original report in response to Mr. Martin's.

Q. Okay. And that was written in Python?

A. No. That was in SQL.

Q. SQL.

A. S-Q-L.

Q. Got it. All right. And then, Mr. Soha, did he provide data science or -- well, you tell me. What kind of support did he provide you specifically with respect to

14

the 2022 report?

A. So in the 2022 report, Mr. Soha helped in the re-performance of certain of the code that Dr. Freer provided.

Q. "The re-performance of certain of the code." Can you be more specific?

A. Sure. So we received the underlying code that Dr. Freer relied upon in forming his opinion. You know, as part of that, we reviewed that code, and then we took small portions of the data and reran and walked through that code step-by-step to make sure we understood how it worked.

Q. Directing your attention to Exhibit 2.

MR. LYNE: (To Attorney Papas) Did you hand it to me? Oh, I have it.

Q. Is this your two thousand -- your current 2022 report, absent the tables?

A. 2023. Sorry.

Q. Sorry. 2023.

A. You said '22 before. My apologies.

Q. Yup. My mistake.

15

A. It's my current report, yes.

MR. LYNE: Okay. (To Attorney Papas) And I need the code instructions.

MS. PAPAS: Here.

(Deposition Exhibit No. 3, Borelian Aggregation Methodology Code Execution Instructions, marked for identification)

Q. This is Exhibit 3. Let me just make sure I understand with respect to your report. Together with the tables, which I have not yet marked, but together with those tables, the Exhibit 3 plus the tables would constitute your entire report?

MR. RONA: Objection. You said Exhibit 3.

MR. LYNE: Oh, I'm sorry. Let me ask the question again.

Q. Let me make sure I understand. Exhibit 2, together with the tables, constitute your entire report; is that correct?

A. By "tables," you're referring to the

16

exhibits I attached?

Q. Yes.

A. That would be my expert report, yes.

Q. Okay. And was it your intention to include in your report all of the opinions that you formed that you deemed to be relevant to your engagement?

A. I certainly attempted to put in all my opinions in some fashion I could. Obviously, there was a lot of thoughts in my head, and I did my best to get them down on paper.

Q. All right. It was your intention to include in your report all of the opinions that you deemed to be relevant and important to your engagement. Is that a fair statement?

A. I think that's generally fair, based on the information I had available to me at the time.

Q. Okay. Have you subsequently obtained information which caused -- has caused you to change any of the opinions contained in your report?

A. I think in going back and, you know,

preparing for the exhibit -- for this -- excuse me -- this deposition, you know, I think there are things that I would -- maybe additional critiques or things that I feel I could have added, but nothing that substantively changes the ultimate opinion and conclusions.

Q. Did you omit any opinions to your report that you deemed to be relevant to your inclusion -- or to your engagement?

A. I don't believe so.

Q. Okay. Thus, to the extent that you formed opinions that you deemed to be relevant to your engagement, those opinions are disclosed in the four corners of your report and its exhibits; correct?

A. I think necessarily a summary of my opinions.

Q. Okay. But nothing substantively different?

A. Not that I can think of sitting here today.

Q. All right. The exhibits that you attached to your report include, for example,

cluster tables; right?

A. Can you be more specific?

Q. Do you remember attaching, for example, a 430-line table from Maria Neves, a cluster table?

A. I don't remember the exact number of records in that table, but I certainly included the user accounts that Dr. Freer had aggregated under the name Maria Neves --

Q. Okay.

A. -- and other names.

Q. And the tables that were generated by your office, did you include and attach all of the tables that you generated in the course of your investigation or just selective ones?

A. Selected ones.

Q. All right. And can you give me an order of magnitude how many tables StoneTurn generated in the course of its investigation associated with Dr. Freer's report?

A. What do you mean by "generated"?

Q. Well, "generated," you created, either digitally or in hard copy tables, some form of analytical tables. Is that a fair

statement?

A. It depends. Certainly, there's data exploration, so I can query a database and on the screen will show me some subset of that data. I didn't necessarily take that data and then paste into something else or create a table add that is something that lives momentarily until I query something else, so that's why I'm trying to understand when you say "generate," what you're referring to.

Q. Well, let's go with queries. Did you generate a number of queries against the Freer database?

A. What do you mean by "Freer database"?

Q. The clusters that Mr. Freer generated as a result of his work.

A. So what I'm trying to think through when you say clusters he generated, so generated groups of user accounts. There's a lot of information that goes along with it across many different tables, so I just want to make sure what tables you're asking me in terms of how many queries I did.

Q. Well, let's then go to Exhibit 3. Let me show you what's been marked as Exhibit 3. Directing your attention to Exhibit 3, do you recognize this document?

A. Certainly. The answer is I recognize -- I wouldn't say a singular document. What we received from Dr. Freer was various different things, different files.

Q. All right. Well, this says "Borelian Aggregation Methodology Code Execution Instructions." Do you remember ever reviewing this?

A. Trying to recall. We certainly received a lot of the code. I'm trying to remember reviewing this particular one or not. This does look familiar. I think this had to do more with setting up the database itself.

Q. Okay. Did you or somebody under your direction then set up the database consistent with the code execution instructions set forth in Exhibit 3?

A. Not in full, no.

Q. All right. You said previously that Mr. Soha --

21

A. Uh-huh.
Q. -- created and reviewed small portions of the code, ran small portions of the code; is that correct?
A. Yes.
Q. All right. And how did he determine which portions of Dr. Freer's code to run?
A. We had discussions about certain aspects of it, about how the blocking or what kind of -- how the cleaning worked, and so we would take samples of the data and run it through portions of the code to see what the out would look like.
Q. With what intention?
A. Simply making sure we had a full understanding of how Dr. Freer's code worked, reading it in connection with his expert report to make sure we had a clear picture of what the inputs and assumptions were.
Q. And what portion of the code as a percentage did you actually run?
A. I guess I don't know how to -- how to measure that.
Q. Well, he has ten steps in his

22

analytical process. Do you know that? Do you remember that?
A. Yes, I do.
Q. Okay. Of the ten steps, did you run code for any of those ten steps, blocking or otherwise?
A. Yes. Certainly.
Q. All right.
A. Certainly.
Q. How many of the steps did you run code for?
A. I don't recall.
Q. Do you have any recollection of whether it was more than half of the steps or less than half of the steps?
A. I don't recall. We didn't look at it in the concept of -- the context of steps.
Q. How did you look at it?
A. We received a number of files, and the files that we received the code didn't necessarily comport one to one with the steps.
Q. Well, did you attempt to run validating checks with respect to any blocks of the code that were associated with a

23

discrete step in his process?
A. Can you ask that question again?
Q. Sure. Did you attempt to run code with respect to any of the discrete steps of his process as an entire step so that you could validate his output in that step?
A. I don't recall.
Q. Well, you mentioned blocking; right? That was one of the things you ran code against?
A. For a portion of it, yes.
Q. When you say "a portion of it," what are you referring to?
A. I'd have to -- so there's a lot of code that goes into any one of those steps. I think we looked at certain pieces of that code, to my recollection.
Q. Okay. But let's take blocking as an example. What portions of blocking code did you run?
A. I'd have to go back and look.
Q. And where would you look?
A. I'd probably go back and talk to Mr. Soha and go back through and try to recall

24

what we specifically did, what we ran, and what we discussed.
Q. Is Mr. Soha the person who would have the most knowledge with respect to actually what portions of the code were run?
A. Between him and I. We both worked on it together.
Q. Okay. But you're not recollecting any specific sets other than looking at some portions of the code -- blocking code, and I'm trying to determine, as you sit here today, are there any other portions of the Freer code other than a portion of the blocking code which you or Mr. Soha ran?
A. Certain of the steps related to the cleansing of the data.
Q. Which steps?
A. So the step itself is cleaning -- under Dr. Freer's ten steps, my recollection is cleaning was one of the steps. Not the first step. May have been the second step.
Q. Okay. And what portions of the cleaning code did you run?
A. I wanted to look at and see and

understand particularly that was related to the name field. He had made certain statements in his report about how he had cleaned the actual string -- that wasn't the name -- such as trimming off (inaudible).

THE STENOGRAPHER: I didn't hear you.

A. Oh, I apologize. Trimming. So removing leading and trailing spaces, converting everything to lower case, as well as setting anything that, I think he said, was one character to null or blank. I believe he had also said and we had looked to confirm that he had also eliminated certain names that he considered to be entirely garbage, so things that were only punctuation and in numbers is my recollection.

Q. And in running those portions of the code that you ran, did it appear that Dr. Freer's description of how his code operated was accurate?

A. I didn't see any inaccuracies.

Q. All right. Would the same be true with respect to the blocking steps?

A. I didn't see any errors in the coding itself.

Q. Okay. Other than cleansing and blocking, did you run any other portions of Dr. Freer's code?

A. I think those are the two primary steps that we focused on that we wanted to understand.

Q. When you say, "those are the two primary steps," were there any subsidiary steps that you focused on after looking at blocking and cleaning?

A. So I also looked at the production or how the final tables that were provided were created or produced. What we received from Dr. Freer in terms of output were two tables. One that was kind of a more detailed output table, and one that was titled summary.

Within the output table, the larger of the two, we saw that there was significant duplication of user accounts where the name that was used across the duplicated records sometimes differed, and we wanted -- and I wanted to go and look and understand what may

be creating that duplication and those differences.

Q. And did you form an opinion as to what was creating the duplication or differences?

A. Ultimately, I didn't. I didn't see in the code where that -- what step was creating that potential duplication.

Q. When you say "potential duplication," I want to go from the abstract to the concrete.

Was there duplication or potential duplication?

A. So there was definitely duplication in that table. In the summary table, which was a summary of the output, it didn't appear that the duplication existed, so that's why I say "potential." There was clearly duplication in the detailed output table.

Q. And the duplication was in what table fields?

A. At least within the name field. It may have been in others.

Q. And are any of the specific

duplication detail included in your report?

A. I'm trying to remember if I -- if I inserted a footnote to that -- that effect. I may have. I'd have to go back and look.

Q. Do you have a recollection of what percentage -- when you say the name field, are you talking about rep_id?

A. No. No. No. The rep_nome, N-O-M-E.

Q. Okay. And besides the rep_nome field, were there any other fields which you determined had duplication?

A. Well, to be clear, it's not just -- the records -- the entire records were duplicated; not just the field.

Q. Okay.

A. What was different was -- about the name field was that they were duplicated because the -- likely because there's distinction or difference between the name field and what would otherwise be the duplicate record.

Q. You mean in the clustering process, it was -- they were -- there were clusters

with more than one cluster with rep_nome?
A. No.
Q. What do you mean?
A. So, by way of example, within the output, rep_id should be unique. Rep_id represents a unique user account, and there should be one and only one. Rep_id should appear once and only once in that output set. What we saw is that rep_id was appearing more than once, and so when we looked across two records that had the same rep_id, what we saw was that there was a difference in at least the name field, and maybe others, across those two names.
What appeared to have been happening is that at some point in time, the string or phrase "NAN," N-A-N, had been excluded. "NAN" in Python is another word for null or blank. And in doing so, it appears in certain ways to be including -- causing certain duplication. So, for example, Fernando would be Fer-Do. I think I got that right.
Q. Is that the extent to which you located duplication, when a name, for example,

included N-A-N?
A. There seems to be what was causing the duplication, yes. It may have extended to the address and other fields as well, but I remember specifically seeing it in the name.
Q. And as a percentage of output, what percentage of the records implicated this duplication?
A. I don't recall doing that calculation.
Q. Was it a significant amount or a de minimis amount? I'm trying to look at order of magnitude here.
A. Again, I'd have to go back and do the math. I'm not sure I calculated that.
Q. Was it in the thousands, the hundreds?
A. I would think it would be more than a hundred, but I don't -- again, definitely more than a hundred. I don't remember the exact magnitude though.
Q. All right. To the extent that you determined that that was a -- that was a -- strike that.

Did any of the opinions in your report rely upon this duplication based on the NAN, N-A-N, field issue?
MR. RONA: Objection.
A. It was certainly something I had to account for when utilizing Dr. Freer's output.
Q. And how did you account for it?
A. I accounted for it by, you know, consolidating his output to make sure there was only one rep_id within the output table to avoid duplication within my analysis.
Q. Okay. We were talking about which portions of the code you had run, and you had run the portions of the code on the cleansing step, on the -- on the out -- well, strike that. Did you run it on the output table?
A. I'm sorry?
Q. I'm sorry. You received an output table. I want to focus just on what portions of the code you actually ran. So you ran it on blocking, you ran it on trimming; right?
A. Cleansing --
Q. Cleansing. Okay.
A. -- was the name of the step.

Q. Yup. Anything else that you can remember?
A. I think those are the two we primarily focused on.
Q. Okay. And, for that, you relied upon Mr. Soha?
A. In part.
Q. All right. Did you write the code yourself for any of those analytical steps, or did Mr. Soha?
A. Well, we were relying on Dr. Freer's code.
MR. LYNE: Okay. Before I turn away here, the -- I need the materials considered portion of his report.
MS. PAPAS: The list?
MR. LYNE: The list. Do you have that? It's Exhibit 2 to his report. Can you go get that?
MS. PAPAS: Yup.
MR. LYNE: And while we're doing that, let's go with Mr. Dennis's CV.

(Deposition Exhibit No. 4, Joshua Dennis Curriculum Vitae, marked for identification)

Q. Sir, directing your attention to what's been marked as Exhibit 4. Is that a copy of your current CV?

A. Yes, it is.

MR. RONA: It's current down do the necktie.

THE WITNESS: Wow. Look at that. Very observant.

MR. LYNE: And including the shirt, it appears.

THE WITNESS: Maybe. I have multiple blue shirts, but it's certainly possible.

Q. So you have a BS in management and business from Skidmore; is that correct?

A. It is.

Q. And management and business is an undergrad MBA-type study? What is it exactly?

A. If you're comparing it to MBA, it was like an MBA to the extent it concentrated, I think, on various areas of business, so

within that, it involved multiple accounting courses, economics courses, finance courses, marketing courses. Things of that nature.

Q. And you had a minor in computer science; correct?

A. Correct.

Q. And the minor in computer science required how many credits?

A. I don't recall.

Q. Do you remember approximately how many computer science courses you took at Skidmore?

A. Between six and ten probably. I'd have to go back and look at my transcript. It's been a while.

Q. Did you take any data science courses at Skidmore?

A. As distinct from computer science?

Q. Yes.

A. I think it could be considered one and the same.

Q. Well, I would consider computer science to be a broader area than data science. Data science being the manipulation

and aggregation or disassociation of data -- datasets. So if I can use that term, did you take any data science courses that concentrated on manipulating or either aggregating or disassociating datasets?

MR. RONA: Objection. You can answer.

A. Yes, I did.

Q. Okay. And, in that context, what courses do you remember taking?

A. I certainly at least had coursework related to at that time it was Microsoft Access.

Q. Anything else?

A. I think that there may very well have been, but, again, this is quite some time ago. That's the one that comes to mind.

Q. Okay. And then subsequent or -- strike that.

In college, did you take any statistics courses?

A. I did.

Q. Which -- how many courses do you remember taking?

A. At least one.

Q. Sort of an intro to statistics type course?

A. Statistics is offered in various programs. There was a business statistics. There was math statistics. I recall certainly taking the math statistics. I don't recall if I took separately a business statistics course.

Q. Okay. Was the math statistics course a half-year course, a semester course, or a full year?

A. I think it was a semester course.

Q. Okay. So let me move now. After you graduated, did you continue your education after you graduated, either formally or informally?

A. Yes.

Q. All right. And can you tell me, to the best of your recollection, what coursework or study you engaged in relating to, I'll call it again, data science?

A. So certainly I took continuing education as it related to SQL. That would be

the primary focus of continuing education -- formal continuing education as opposed to experiential.

Q. Okay. And the SQL coursework was done through what entity?

A. That was a number of years ago now. I don't recall.

Q. Was it a continuing education like a night course, or was it --

A. No. It was -- it was a daytime course.

Q. Multiday course --

A. Yes.

Q. -- multiweek course?

A. I think it spanned a couple weeks. Yes.

Q. Couple weeks. Okay. And do you remember generally just it was sort of SQL power user type stuff?

A. What do you mean by "SQL power user"?

Q. Well, maybe I'll ask you to tell me. What was the subject matter of the SQL course that you took over several weeks?

A. So -- sure. It was learning about programming in SQL. It looked at complex ways to join, manipulate, aggregate, filter data.

Q. All in SQL?

A. In SQL, yes.

Q. Okay. And besides the SQL course that you took, was there any other formal continuing education you took, leaving aside experiential education?

A. Related to data science?

Q. Yes.

A. Not that I can think of.

Q. Okay. And, now, you refer to experiential education. What do you mean when you say "experiential education"?

A. So having been in this business for, it says more than 15, it's becoming a lot closer to 20 years now, a large portion of the work I do involves data analysis or assessing data analysis generally in the area of compliance litigation or investigation matters. So in doing that, I use a wide variety of programs and approaches to perform data analytics, and doing that on a daily

basis has led to a great degree of education.

Q. All right. First, let's start with when you refer to compliance and investigation, what are you referring to?

A. Sure. I'll take that in two parts, if that's okay.

Q. Of course.

A. First, maybe it relates to compliance. A big portion of StoneTurn's work is in the area of things like monitorships, so that is where we might come in and assess the compliance program of a company to make sure it's reasonably designed and implemented to detect and prevent fraud. That may review-- that may include reviewing how they approach things like transaction testing.

In the area of litigation, that's, obviously, a very big area. A lot of my work relates to economic damages, whether it's through things like contracts disputes or intellectual property, patent infringement trademark. That frequently involves financial modeling and data analytics. I think those are the two you asked me to describe.

Q. And your compliance experiential education, is it primarily directed towards testing the sufficiency of a business's internal compliance systems?

A. I think the word I'm thinking about there is "sufficiency." That's only one aspect we look at. And efficacy would be another.

Q. Efficacy and sufficiency. Would that that be a fair --

A. There are likely others, but those are the two big ones.

Q. Okay. And then on the litigation side, your experiential education includes looking at and creating models for economic damages analysis?

A. That's certainly part of it. I mean, creating models is a wide, big area. A lot of that includes taking large or complex datasets that need to be joined or integrated or compared or distilled or visualized, and using that to usually come up with an opinion in the form of economic damages, although there are certainly other opinions, I'm sure,

we've offered.
Q. This case involves large-scale data aggregation or very large dataset. Is that consistent with your understanding?
A. That's certainly one aspect of this case.
Q. Okay. Have you had previous experience with data analysis and manipulation on a dataset of equivalent size?
A. Yes.
Q. Okay. And --
A. I'm sorry. When you say "dataset," you're referring to -- there's a lot of different datasets that are in this case. I just want to make sure that I'm clear on what you're referring to by the "dataset."
Q. Well, the transactional dataset. What's the term we use for that? SIG?
A. You mean the original source data that Huron analyzed?
Q. Yes.
A. Yes.
Q. Okay. Do any of those -- those engagements appear in your "Select

Professional Experience" description on the second and third pages of your -- actually, it's pages two, three, four, and five?
A. So I imagine some of them would, yes.
Q. Which?
A. So, certainly, one, two, three -- maybe the fourth one down certainly would.
Q. The class action matter about potential wages due?
A. Yes. And, again, I think it's important to make the distinction that when we're talking about data, it's not that there's a single table that encompasses all the records. That's rarely the case, and wasn't the case here. Here, the primary tables that there were, you know, four or eight, depending how you look at it, each contained a subset of data from a different perspective in terms of what Dr. Freer relied upon as well in conducting his analysis.
Q. Okay. Are there any other of your stated select professional engagements on your CV that you consider to be akin in size to the

size of the Telex dataset or sets, plural?
A. I think focusing on the dataset in terms of what was analyzed by Dr. Freer, I think a lot of -- again, it's hard for me to say 'cause I don't remember the total sum -- I'm talking about purely on size, and it depends how you define "size." Are we talking about number of rows? Are we talking about the file size? I guess I don't have all that data in front of me to know one way or the other.
Q. Well, there were how many -- how many user accounts in the SIG?
A. I think my recollection was there was, I think -- and I have it in my report, but in the neighborhood of 17 million, of which 15 million had at least one paid invoice.
Q. All right. And if you take those and you exclude the Brazilian accounts and just focus on the domestic Telex accounts, that would be about 11 million accounts?
A. I don't recall if that's domestic or domestic and international.

Q. Okay. But taking that number of about 11 million accounts and using that as the metric against which to look at your other engagements, do any of the other engagements on your CV involve datasets that large?
A. They very well could. I rarely -- sometimes I think it's interesting to put in number of records, but it's rarely to me a meaningful statistic, other than simply of interest, so I don't know, sitting here today, the number of records associated with each of these engagements.
Q. Well, regardless of whether you have a specific number of records, are there any of these that you considered to be akin in size to the Telex SIG database?
A. Again, I'd have to go back, to the extent we even still have them. We tend to not retain data beyond what's necessary for the engagement. So I'd have to go back and look at each of these cases to see what data we received, what we analyzed, what we did. It would really depend. I would have to add up all the data received across multiple

45

tables for different purposes. It's really a hard comparison to make.

Q. Okay. Do you have experience using Python notebooks?

A. Some.

Q. When you say "some," what experience do you have?

A. We have certainly utilized Python for certain analyses. It really depends on the use case.

Q. When you say we use Python notebooks, do you -- are you conversant in Python?

A. I have some base knowledge.

Q. When you say "I have some base knowledge," what do you mean by that?

A. So we certainly use Python when necessary. As part of that, I'll go through and walk through the notebooks. I think I generally have a stronger understanding of literacy in it, reading it and understanding how it works. I do some basic analysis in Python, but generally use SQL.

Q. Okay. How about Pandas? Do you use

46

Pandas?

A. Pandas data frames?

Q. Yes.

A. I have in the past.

Q. And do you consider yourself to be fully conversant in Pandas packages?

A. I'd say partially. It's not something I use frequently.

Q. Something you use -- you've used a handful of times or --

A. I think a handful is a good estimate.

Q. All right. And would the same be true with respect to Python?

A. Probably more -- a little bit more than a handful.

Q. But not -- it's not your go-to?

A. Generally not. I tend to work more in SQL.

Q. How about Splinc? Have you used Splinc?

A. No, I haven't.

Q. How many of your -- how many -- I'll call them data aggregation projects. So how

47

many -- how many projects that you have been involved in have been data aggregation projects that require cluster analysis of the type that we are dealing with in the Telex case?

A. When you say "cluster analysis," can you be more specific?

Q. Attempting to aggregate accounts across a large database assigned to individual clusters.

A. I'm still not sure I understand your question.

Q. There are 11 million -- let's use that number -- 11 million accounts of interest in this case, whether domestic or domestic and international. Of those, there are a number where every one of those accounts has to be assigned to a cluster, right, to determine net winners and net losers?

MR. RONA: Objection.

A. I don't necessarily know that's true.

Q. Well, it could be down to a cluster of one. That would still be a cluster, or it

48

could be a cluster of one plus N, whatever that number is.

Taking that as your guide, of the 11 million accounts, to determine who's a net winner and a net loser, you have to determine, for individual participants, which one of those accounts gets assigned to the participant; correct?

A. To a participant.

Q. Yes.

A. Yes.

Q. Okay.

A. My understanding, that was part of the analysis that needed to be performed.

Q. Okay. And when I talk about a cluster analysis, I'm talking about that exercise, taking a series, an N number of participants and an N number of accounts and determining which accounts belong to which participant, like we're doing in the Telex case?

MR. RONA: Objection.

Q. Okay?

A. I think, in my mind, there's a

distinction between a cluster and a participant. Those things are not necessarily -- often they're not the same, so I think that's why I'm having a hard time answering your question.

Q. Well, you'll agree with me that the exercise in this case for determining who is a net winner and who is a net loser is looking at the 11 million accounts, individual accounts, and then determining whether or not those accounts represent a single participant, or whether a single participant may have opened multiple accounts which should be aggregated into a single cluster. Is that a fair statement?

MR. RONA: Objection.

A. No. I don't think it is.

Q. Why?

A. I think they should be aggregated to a single participant.

Q. All right. So using the term "participant" as opposed to "cluster," let's go there. So the exercise is to take N number of accounts, which in this case is 11 million,

and assign to all -- to all of those accounts -- all of those accounts two individual participants, such that no account is assigned to more than one participant.

MR. RONA: Objection.

Q. Would you agree with me that's the exercise here?

A. I think it is, with the caveat that I don't know factually or legally the basis for exclusion of Ympactus. Certainly, it's what's been done, but notwithstanding that fact, I think that's accurate.

Q. Okay. Now, that same concept, that same process, have you had other engagements which have required you to go through this same process, whether you call it a cluster analysis, whether you call it something else, have you had other engagements that require this same analytical process?

A. Yes.

Q. Okay. And can you describe for me what engagements those would be?

A. Sure. So one matter was -- it originated as an anti-price fixing matter.

The bank needed to notify -- they went back a few years now -- notify certain participants of the issue. The data we received related -- had the names of businesses and addresses, and part of what we were asked to do is essentially deduplicate, which is one aspect, and essentially not similar to what was being done here, but deduplicate those addresses to minimize a number of mail-ins the company ultimately had to do.

Q. All right. Well, that's a portion of what happened here, deduplicating?

A. Correct.

Q. But it was a little more than that at the end; right, in this -- in the Telex case? You weren't -- people weren't simply deduplicating records?

A. I guess --

MR. RONA: Objection.

A. -- it depends. I don't know I agree with that.

Q. All right. What analytical process did you use to deduplicate records in the anti-price fixing engagement?

A. Your question is about process?

Q. Yes.

A. So we went through and looked at the data. We analyzed it in terms of looking at the nature of the different fields we had at our disposal. In that case, we went through, if I recall correctly, and used a combination of fuzzy matching and parsing to try to split out and find addresses that were likely in common and should be brought together and aggregated.

Q. And what fuzzy matching tools did you use?

A. I'd have to go back and recall. I don't remember sitting here today.

Q. How long ago was this engagement?

A. Three, four years maybe. Maybe more.

Q. I'm sorry. Was it three or four or more?

A. I said three or four or maybe more. I'd have to go back and look.

Q. Okay. But you can't remember what fuzzy matching tools you used?

53

A. I think we used a combination, so that's why I'm trying to remember which specifically ones we used. We may have used some probably likely just right in Excel, Excel fuzzy being one. I think we may have used other approaches as well. Probably Levenshtein-based approaches as well at that time.

Q. So Levenshtein approaches? Jaro-Winkler? I mean, which approaches did you use?

A. I'd have to go back and look. It's been a number of years.

Q. And what size dataset are we talking about here?

A. I don't have a specific recollection.

Q. Okay. Other than the anti-price fixing case, were there any other cases or engagements which you considered to be akin to the process that was used here?

A. And by "process," again, you mean general aggregation or deduplication?

Q. Sure.

54

A. I think, as a conceptual matter, again, aggregation is something we do. Grouping things is summing (inaudible).

THE STENOGRAPHER: I'm sorry. I didn't hear you.

A. I'm sorry. Aggregation. So generally summing things up or grouping things is something we do in a lot of our matters.

Q. All right. Well, let's then focus on deduplication.

A. Sure. So for deduplication, again, we got a lot of datasets that are very messy, so I would have to say as a deduplication, a lot of our matters contain data that we need to present in a more summary fashion. We have multiple records that relate to a single entity or company, and we need to aggregate and bring that down to show something at a more summary individualized level.

Q. Okay. Well, that sounds more like a data dashboard, but I'm really focusing on deduplicating otherwise dirty datasets.

You referred to the auto price fixing case where you used deduping tools.

55

Were there other engagements in which you used deduping tools?

MR. RONA: Objection.

A. I have no doubt that there are. And, again, deduplication is part of the process. Generally understanding what you have when you have, say, a dataset, and really deduplication is a function of what it is you're trying to summarize or aggregate or show. In this instance, if the goal was to look at things on a rep_id level, then there would be no deduplication needed. It's, certainly, a function of what you're trying to get at.

Q. I'm really focusing on your engagements. And you said I'm sure there may have been others, but really what I'm trying to do is go from the abstract to the concrete. Are there others that you can recollect as you sit here today?

A. May I go through my list I have here?

Q. Of course.

A. Yup. So in the -- in the first one

56

here so --

Q. The 500 hardcopy bank statements?

A. Yes.

Q. Okay. Is that listed first 'cause it's been your most recent?

A. No.

Q. Okay. Anything else?

A. The third bullet is the one we discussed with respect to anti-price fixing -- antitrust and price fixing. Yup. So, certainly, the fourth one as well, and I believe the second to last bullet under the complex data analysis, the one related to the unbundling of products. This is not to say there aren't -- I have to keep going -- not to say there aren't other examples within the complex business disputes or intellectual property or valuation subcategories.

Q. Well, that says, if I'm reading the right one, retained by a national owner and franchisor of travel centers?

A. Oh, no. No. I'm sorry. I think we're on different pages.

Q. Hence, my confusion. Which page am

I supposed to be on?
A. The third page. The second one before "Complex Business Disputes."
Q. Reimbursement pursuant to federal drug rebate program?
A. Yes.
Q. How about -- well, strike that.
Are you familiar with the modeling -- strike that.
Are you familiar with the term "Fellegi-Sunter"?
A. I am.
Q. All right. And what is Fellegi-Sunter?
A. It is one approach to probabilistic record linkage.
Q. And it's F-E-L-L-E-G-I, hyphen, S-U-N-T-E-R.
In any of your prior engagements, did you use Fellegi-Sunter?
A. I have not had cause to use that, no.
Q. All right. Is it fair to say that the Telex engagement is the first time that

you've been exposed to the Fellegi-Sunter model?
A. It is.
Q. How about Charles Soha, do you have an understanding as to whether he ever had any prior experience with Fellegi-Sunter?
A. I don't believe so.
Q. What are the inputs, as you understand it, to the Fellegi-Sunter method, and what is the output sort of generally?
A. That's a very big question. I mean --
Q. Uh-huh.
A. -- in that Dr. Freer relied on a lot of inputs. The entirety of what's in his documents is considered solely.
Q. No. I'm talking sort of generally speaking, in a Fellegi-Sunter model, what are the inputs to the method, the Fellegi-Sunter method?
A. I think it would depend on the purpose to which it was being performed.
Q. Well, in what context is Fellegi-Sunter used as you understand it?

A. It's a part of probabilistic record linkage.
Q. Okay. All right. Let me try it a little differently.
How many prior engagements have you had in the field of probabilistic data analysis or data aggregation?
A. Can you ask that question again?
Q. Sure. How many engagements, prior to the Telex engagement, involved probabilistic computing?
MR. RONA: Objection.
A. Again, engagements usually have a goal that is not -- the approach isn't determined necessarily by the engagement. The engagement helps determine how you approach it, so that's why I'm struggling with answering your question specifically.
Q. In any of the Select Professional Experience descriptions, did you use probabilistic computing in any of those?
MR. RONA: Objection.
A. I believe I did.
Q. Which ones?

A. Certainly, at a minimum, the price fixing matter.
Q. And when you say "probabilistic computing," are you referring to your fuzzy deduping?
A. Fuzzy matching can be one aspect of it, yes.
Q. All right. Well, I'm really interested in what you actually did. And you said in that engagement, you used fuzzy matching, which has a probabilistic component to it.
So other than your fuzzy matching experience, have you ever had any experience using probabilistic computing tools beyond simple fuzzy matching?
A. I'd argue with the concept of fuzzy matching is simple.
Q. Okay. The Levenshtein ratio, whatever you want to call it.
A. There's multiple different techniques we use, but, yes.
Q. Okay. Whether it's using Excel tools or the Levenshtein ratio or whatever you

61

want to call it, that's one aspect of probabilistic computing that you have some experience with; right?
A. There are other methods of fuzzy matching and things we use, but, yes. That's one aspect.
Q. Okay. What other tools have you used in prior engagements for probabilistic computing?
A. When you say "tools" --
MR. RONA: Objection.
A. -- what are you referring to? Sorry.
Q. Whether it's a software package or whether it's a native code. Whatever it is that you consider to rise to the level of a probabilistic computing system as opposed to a deterministic computing system.
MR. RONA: Objection.
A. Generally, when we look at probabilistic record linkage, it's usually a combination of things. It can be -- some aspects might be deterministic; some aspects may be probabilistic. We often do use fuzzy

62

matching as part of that. Whether it's TF-IDF or Jaccard distance or Levenshtein or some combination thereof.
Q. But that's all data cleaning?
A. No. I'm sorry. I thought you were --
Q. Okay. I'm really -- so you've identified TF-IDF, Jaccard, and you said Levenshtein?
A. Yes.
Q. Any other what you considered to be probabilistic tools which you have used previously, models, tools, however you want to describe it?
A. Again, those fuzzy matching are inputs to an overall methodology and process.
Q. Understood. Any other tools or models that you've used in that process?
A. Each one is case and fact specific, and we applied them in different ways, so I guess that's why, you know, it's hard to put an individual name on something when each thing is so case and fact specific.
Q. Well, for example, we can agree that

63

you've never used the Fellegi-Sunter data aggregation method; correct?
A. Correct.
Q. All right. And you've used Levenshtein distance?
A. I have.
Q. Okay. You've used Jaccard?
A. Jaccard, J-A-C-C-A-R-D.
Q. You've used TF-IDF?
A. Correct.
Q. "TF-IDF" stands for what?
A. I'd have to go back and look. I don't recall.
Q. Is it a -- is it a fuzzy matching model?
A. It is, yes.
Q. Do you have any understanding as to how the Fellegi-Sunter model with expectation maximization improves upon the base Fellegi-Sunter method?
MR. RONA: Objection.
A. I've certainly read that characterization in Dr. Freer's report.
Q. Well, I'm asking have you -- do you

64

have an understanding as to how FS-EM improves upon or differs in any way from the FS model?
MR. RONA: Objection.
A. Again, I'd refer to what was in Dr. Freer's report.
Q. Well, other than referring to what's in Dr. Freer's report, do you have any understanding, as you sit here today, as to how FS-EM differs from FS?
A. That's not something I've independently studied.
Q. Okay. Well, whether you've independently studied it or not, can you describe for me, as you sit here today, what the difference is between Fellegi-Sunter and Fellegi-Sunter with expectation maximization?
MR. RONA: Objection.
A. I certainly went back -- I want to go back and look at the way it was described. I have a general high-level understanding.
Q. All right. Tell me what your general high-level understanding is.
A. That the expectation -- expectation maximization looks to go and help determine

65

the relative weights of the various fields of agreement within the -- within the comparison.

Q. And how are the relative weights of the fields determined in FS-EM?

MR. RONA: Objection.

A. You'd have to go back and look at the mathematics of it. That's not something I've studied.

Q. Did you form any independent opinions with respect to whether Dr. Freer's weighting of fields using the FS-EM model was done correctly or incorrectly?

MR. RONA: Objection.

A. I didn't find any issues with the code itself, which is what performed the process.

Q. Well, beyond just looking at the code, did you form any opinions with respect to whether Dr. Freer applied the FS-EM method appropriately against his dataset?

MR. RONA: Objection.

A. I did.

Q. Okay. And with respect to the relative weights which the FS-EM model

66

ascribed to each of the fields, did you form an opinion as to whether those relevant weights were appropriate or inappropriate?

MR. RONA: Objection.

A. It's hard for me to aggregate that step from everything that came before and look at that just in isolation. I disagree with a lot of the assumptions and the approach that went into it, so it's hard for me to say that this single step is right 'cause it's a culmination of everything that's led up to it in case-specific facts.

Q. Okay. Well, let's go a little deeper. With respect to the relative weights of the fields, can you describe for me what inputs you considered to be most important in the FS-EM model?

MR. RONA: Objection.

A. What inputs that were most important in the FS-EM model. I don't think I looked at it in that perspective in terms of most to least important.

Q. I'm just trying to determine, did you form any opinions with respect to whether

67

the relative weighting process used by Dr. Freer in his FS-EM model used appropriate inputs and, therefore, the weights were correct or inappropriate inputs, and, therefore, the weights were incorrect?

MR. RONA: Objection.

A. I didn't really look at his model from that perspective.

Q. As you sit here today -- let me try to look at --

MR. LYNE: (To Attorney Papas) Can I have the chart of the ten steps?

(Deposition Exhibit No. 5, Page 43 of Dr. Freer's Report Entitled Steps in the Aggregation Methodology, marked for identification)

Q. Sir, directing your attention to what's been marked as Exhibit 5, which I will represent to you is page 43 of Dr. Freer's report, have you seen this table before?

A. I have.

Q. Okay. And it lays out ten steps in his -- in his process. Do you see that?

68

A. I do.

Q. And which of these steps involve data cleaning using fuzzy matching?

A. I'm not sure I agree with the source of the question. Fuzzy matching isn't inherently data cleaning.

Q. Okay. Well, which of these steps includes fuzzy matching?

A. So, at a minimum, certainly it's described in at least the blocking and the clustering processes. They're elements of fuzzy matching.

Q. And field cleaning?

A. I don't believe there was in the field cleaning, at least I don't recall that there was, sitting here today.

Q. Okay. Step one, filtering 17 million users down to 11 million users, do you have any issue with Dr. Freer's decision to filter the 17 million down to 11 million?

A. I mean, that comes back to a legal determination about whether or not the Ympactus accounts should be included or not. Ympactus, Y-M-P-A-C-T-U-S. I think I got that

right.
Q. But leaving aside the legal question, which we can all fight about later, do you disagree with the process used to reduce 17 million total accounts down to 11 million user accounts by disaggregating the Ympactus? The math is right; correct?
MR. RONA: Objection.
A. I -- I didn't see any issue with the filtering process from my mathematical perspective.
Q. Okay. And step one, the field selection and transformation, did you see any issues from a mathematical perspective in the work that Dr. Freer did in step one?
A. In terms of the code itself? I certainly disagree with a lot of the assumptions and the process.
Q. All right. Let's start with code.
A. I didn't see any issues with the code, the mathematics of it.
Q. Okay. And you said you disagreed with the assumptions and the process. What was the disagreement with respect to the

assumptions?
A. So I think that would go back to -- looking at my report in a lot of the inputs and the reliance on the data, the lack of validation, the understandings he has that are unsupported and unsubstantiated with regards to how data was entered, the consistency with which data was entered.
Q. Okay. So we'll get to those. I just want to go through all of these steps. Then he engages in what he refers to as field cleaning. Did you have any issues with respect to the process he used to clean individual data fields?
A. I think there are elements of it that were unclear and not supported.
Q. Which elements?
A. So, for example, he elects to remove and make null or blank any -- at a minimum, names that were either -- I believe it was one character in length, and I believe he also eliminated any names that were comprised entirely of numbers or -- well, he called it, I believe, garbage characters.

Q. All right. Why do you -- do you disagree with that decision, that assumption, or however you want to characterize it?
A. I think it's simply an unsupported assumption. I think it's just an unsupported assumption. There's no basis for removing things in one character, but why not remove things with two characters or three or four, as Mr. Martin had done.
Q. Did you attempt to run any code that removed two characters or three characters or four characters?
A. For what purpose?
Q. Exploration in connection with your engagement.
A. I guess I don't know what the output of that would be. Simply a column that had fewer values.
Q. Well, I'm asking did you determine how many fewer values there might be, what the size of the impact of that process would be on --
A. Yes. Certainly. Because Mr. Martin had elected to remove -- I don't recall if

it's three or four -- I think there's actually some ambiguity in Mr. Martin's report. He had a different determination of whether to discount or remove the name field based on the number of characters, so it's certainly something we looked at differently when we reviewed Mr. Martin's report.
Q. Okay. But I'm trying to determine whether you, yourself, ran any code which removed accounts that had more than one character, say, two characters as opposed to one character?
A. We certainly quantified, and I certainly looked at the number of names that had certain lengths of characters as part of the exploration of the data.
Q. Well, when you say, I looked at an exploration of the data, how many had two characters versus three characters?
A. Oh, I couldn't recall, sitting here today, exactly.
Q. All right. And did you form any independent opinions with respect to what the appropriate field cleaning process was if not

73

removing null/one characters and numbers or garbage?

A. No. That wasn't part of the scope of my assignment.

Q. Okay. Step three, with respect to the deterministic step, did you uncover any errors, either in the code or otherwise, in Dr. Freer's deterministic process? And by "deterministic," I mean his data matching process.

A. I think there's questions about the validity of the data on which it's based.

Q. Leaving aside the validity of the data on which it's based, did you uncover any issues from your perspective?

A. From a code perspective? No.

Q. No. And leaving aside code, but talking about just the algorithm that was used, did you have any issue with respect to the algorithm that was used from a deterministic phase?

A. I wasn't distinguishing the algorithm from the code. I was using those terms synonymously.

74

Q. Okay. There wasn't anything that jumped out to you or was exposed in your data analysis which suggested to you that the deterministic step was performed incorrectly; is that correct?

A. From a purely mathematical perspective.

Q. All right. And then the next step is blocking. I believe you did run blocking code; right?

A. We ran portions of it.

Q. Okay. And did the portions of your blocking code that you did run, did they confirm Dr. Freer's blocking process, or did they expose some differences?

A. Again, I disagree with the approach to blocking and how he chose to -- what and how he chose to do the blocking.

Q. Have you ever had a prior engagement which requires you to perform blocking against very large datasets?

A. I'd have to go back and think. I'm not sure I've thought about --

Q. Well, take your time. Think about

75

it.

A. -- thought about that. We've certainly had matters that have involved large data where we looked to distill it down to that which in those facts we thought were most relevant and pertinent.

Q. Well, that's not blocking. That's sort of an ad hoc determination as to what things you want to look at; right?

A. I disagree that it's ad hoc determination. The purpose of blocking is to go in and try to only look at and perform subsequent analysis, at least initially, on those pairs that Dr. Freer in his determine -- in his opinion thought had a chance of potentially having linkage.

Q. Right. But I'm trying to determine in a -- did you have any -- have you had any prior engagements which required you to engage in blocking large datasets to reduce the computer run time to make it realistic?

A. I don't know that I've done it for purposes of minimizing computer run time. I'd have to go back and think on those.

76

Obviously, there's a lot of cases in the last 20 years.

Q. Well, why do you understand Dr. Freer was blocking in this engagement?

A. Based on what he said, it was to minimize (inaudible).

THE STENOGRAPHER: I'm sorry. I didn't hear you. "Minimize"?

A. Minimize the amount of comparisons he had to make.

THE WITNESS: I'm sorry. I think my mouth is getting a little dry. We've been going for maybe close to an hour and a half here.

MR. LYNE: Would you like to take a break?

THE WITNESS: If that's okay with you.

MR. LYNE: Sure. Of course.

(Recess was taken)

Q. Have any of your prior engagements involved transitive closure?

A. I'm trying to -- I apologize -- I'm trying to think. There have certainly been

77

engagements that have required things to be bucketed and only input in one and only one bucket.

Q. Okay. When you're describing putting things in one bucket, did you use transitive closure in those instances, or was that just bucketizing individual bits of information?

A. My understanding is -- with Dr. Freer is that transitive closure was putting something in one cluster and only one cluster.

Q. Right. But using a transitive process to do so; correct? If A equals B, B equals C, then A equals C? Well, broadly speaking.

A. I'd have to go back and think. It's possible.

Q. Okay. Well, can you think?

A. It's certainly possible. I can't think of anything sitting here today. I'll continue to think on it and let you know.

Q. Okay. Please do. And do you have an understanding how to appropriately choose a

78

sample size to estimate how frequently some property holds in a dataset?

MR. RONA: Objection.

A. Can you ask that question again?

Q. Sure. How do you appropriately choose a sample size to estimate how frequently a given property holds in a given dataset?

MR. RONA: Objection.

A. When you say "property holds," what do you mean?

Q. I mean that a property -- a given property that is true in a dataset, that is -- that is valid in the dataset.

MR. RONA: Objection.

A. I guess I'm not clear on what you mean by "given property."

Q. Are you familiar with the concept of sample size?

A. Yes.

Q. In statistics?

A. I am.

Q. Okay. How do you set sample size in statistics -- in a statistical way?

79

MR. RONA: Objection.

A. Generally, you want to look at a number of factors, such as how many standard deviations you want to have, the size of the dataset. There are other factors that go into it. You have to look at how the -- how disparate the dataset is, how diverse it is. A number of things, depending on what you're trying to look at.

Q. In this engagement, did you do any work to determine appropriate sample sizes for any purposes?

MR. RONA: Objection.

A. I think I certainly compared certain things that were viewed to aggregate size of the dataset. I didn't do an independent determination of an appropriate sample size.

Q. Why not?

A. I wasn't asked to.

Q. So you formed no opinion on that subject as it relates to this engagement?

A. What subject is that?

Q. Calculating sample size.

A. Can you be more specific 'cause,

80

again, I certainly looked at certain transactions or amounts of transactions that were reviewed by Dr. Freer and calculated percentages. I didn't come up with an alternative correct answer, if that's what you're asking.

Q. That is what I was asking.

A. Correct.

Q. Are you familiar with Cochran's Formula. C-O-C-H-R-A-N apostrophe S.

A. No.

Q. Do you understand the difference between false positives and false negatives?

A. Yes.

Q. And can you describe it for me?

A. Sure. So a false positive would be two things that are determined to match that are not true matches. False negative would be things -- two things that the outcome would suggest are not matches but truly are.

Q. I just realized -- let me go back one step. You said you compared some of Dr. Freer's work against the aggregate data size of the dataset; is that right?

81

A. That was probably -- yes, but I think I could probably be more specific in terms of aggregate size of the dataset. I'd have to go back to my report and look specifically what I compared. It was looking, I believe, at the number of user accounts and specifically names that Dr. Freer reviewed in the context of, I think, in the denominator it was the total number of TelexFree user accounts, but I'd have to go back and confirm.

Q. All right. You have your report. Can you take a look?

A. Yes. Certainly.

Q. I'll see if I can find it and we'll race.

A. Yes. I can tell you the answer was about .003 percent. I don't remember if it was in a footnote or in the body.

Q. No. I believe it was in the body.

A. Thank you. I'm on page 31, and I have suspicion it's probably not far from that. Here we go. It's footnote 105, which is on page 34.

MR. RONA: Of Exhibit 2.

82

A. Of Exhibit 2. Correct.

Q. Okay. And this was reviewing the 300 name field values randomly selected?

A. That was the context, yes.

Q. Okay. And did you form any opinion as to whether or not that was an adequate sample size?

A. Not specifically. I certainly would and have reviewed far more.

Q. Okay. But I'm really just focusing on this engagement, and what I think I heard you say was no, you didn't form an opinion as to whether that was an adequate sample size in this instance; is that correct?

A. I would have certainly wanted it to be more than that.

Q. Well, did you do any analysis to determine how much more you thought it should be?

A. No, I didn't.

Q. Okay. Does minimum sample size grow as the population size increases?

MR. RONA: Objection.

A. I think it would depend on the

83

dataset.

Q. In what sense would it depend on the dataset?

A. I think it would depend on the nature of the dataset, what characteristics it had, how similar or different it was across that dataset.

Q. Well, generally speaking, does the required sample size grow in some linear fashion as relates to the overall data size -- dataset?

A. Not linearly, no.

Q. Well, how does it grow?

A. Well, I don't recall the math behind it.

Q. Is it logarithmic? Is it sublinear? Do you have any recollection?

A. I don't recall them.

Q. Besides the name sampling done by Dr. Freer on page 37, I think you referred to it in your report, did you do any other work with respect to sample size and the adequacy of sample size in your engagement?

MR. RONA: Objection.

84

A. I think in certain contexts, yes.

Q. What context?

A. So Dr. Freer makes a comment that a single aggregation, a single cluster is representative -- not going to get his exact words right -- of the general quality of the data, and I saw no indication as to why this particular aggregation selected by Dr. Freer was, in fact, representative.

Q. Did you do any analytical work to determine if, in fact, the single cluster chosen by Dr. Freer was, in fact, generally representative of the quality of the data?

MR. RONA: Objection.

A. I think I did.

Q. What did you do?

A. I looked at lots of different aggregations.

Q. All right. And is that referenced in your report?

A. The fact that I looked at lots of different aggregations?

Q. No. I'm focusing really on the single cluster that you referred to that

85

Dr. Freer relied upon as representative of the general quality of the data overall. Did I hear you correctly?
A. I'd have to go back and see exactly what I said, but that was the gist.
Q. Okay. Can you find it in your report?
A. Where I reference that particular issue?
Q. Yes. Sure.
A. Certainly. I'll race you again if you'd like to find Maria Neves.
Q. Okay.
A. It was towards the end of the section with Dr. Freer.
Q. Page 47?
A. Yes.
Q. Okay. So can you point on page 47 or around this area what specifically you're referring to?
A. I'm sorry. With what regard?
Q. In regard to the single cluster which Dr. Freer said you say referred to as generally representative of the quality of the

86

data.
A. Right. Now that I have it in front of me, the specific language which Dr. Freer contends is, and I quote, "a visual illustration of the general quality of the dataset."
Q. Sorry. Where are you?
A. Paragraph 141.
Q. Okay. And did you do any analytical work to determine if, in fact, the Neves dataset or cluster was, in fact, representative of the general quality of the data in the overall dataset?
MR. RONA: Objection.
A. I believe so.
Q. Okay. And what did you do?
A. I certainly reviewed that particular cluster, and then I certainly looked at many, many other clusters, some of which are summarized as exhibits in my report.
Q. And did you calculate what you deemed to be the error rate for the Neves dataset?
MR. RONA: Objection.

87

A. How do you define "error rate"?
Q. Something included in the Neves cluster that should not have been included in the Neves cluster.
A. I certainly address user accounts that are included in there that appear to be overaggregations.
Q. Okay. And there's, I think, what, 430-ish accounts listed in the Neves cluster; is that right?
A. I'd have to go back and look. It may be in the report here. 466 user accounts.
Q. And of the 466 user accounts, how many did you determine were, in your opinion, overaggregations?
MR. RONA: Objection.
A. I'd have to go look and look in detail at the entire aggregation.
Q. The aggregation is an exhibit to your report; isn't it?
A. It is.
Q. Do you know which number it is? What exhibit number?
A. I have an index of exhibits in the

88

front of my report which should help guide us.
Q. Very good.
A. Let's go with Exhibit No. 16.
Q. Exhibit No. 16.
MR. LYNE: (To Attorney Papas) Do you have that?
Q. And while that's happening, did you do any work -- well, are you familiar with the terms "precision" and "recall" in statistical work -- in statistical analysis?
A. To some degree.
Q. When you say "to some degree," what do you mean?
A. I certainly read Dr. Freer's report and how he described those terms.
Q. Other than reading Dr. Freer's report and how he described those terms, do you have independent experience calculating precision and recall?
A. That's not something I've been asked to do.
Q. I mean generally speaking.
A. There are specific definitions, and they're, again, not something I typically

focus on or I've been asked to focus on.
Q. I'm not asking whether you typically focus on. I'm asking if you have any experience whatsoever calculating precision and recall as it relates to error rate?
A. Generally not.
Q. When you say "generally not," as you sit here today, is it fair to say you can't think of any such experience?
A. Agreed. I certainly look at how right or wrong things are and things that we think are incorrect, but I'm not usually using those terms or necessarily specific definitions.
Q. Well, how do you calculate the error rate if you're not looking at precision and recall?
A. It depends on the case.
Q. Well, let's take it from the abstract to the concrete. In this case, did you calculate error rate, what would be an appropriate error rate for the output of Dr. Freer's cluster methodology?
MR. RONA: Objection.

A. I didn't opine as to what an appropriate error rate would be in this case, no.
Q. Whether you opined or not, did you attempt to calculate what would be an appropriate error rate?
A. No, I did not.
MR. RONA: Objection.
A. It was not something I was asked to do.
Q. Okay. Do you have an understanding as to what an F-measure is or an F1 score is?
A. Not beyond what I read in Dr. Freer's report.
Q. Do you understand what the connected components of a network are in the sense of graph theory?
MR. RONA: Objection.
A. I think I have some sense as Dr. Freer described them.
Q. Other than as Dr. Freer described them in his report, do you have any independent experience dealing with calculating connected components?

A. No. That's not something I've been asked to do.
Q. Generally? No?
A. That's fair. It's not something that I've been asked to do or studied before.
Q. Okay. Let's move on to transitive closure. Can you describe for me or define what is transitive closure?
A. So transitive closure, the purpose was to look at things and make sure that things end up in one bucket and only one bucket. So, for Dr. Freer's purpose, he was looking to place user accounts in one cluster and only one cluster and did so by connecting components, you know, such as A is connected to B and B is connected to C as he defines it, and he applies it such that A and C would end up in one and only one cluster.
Q. Okay. And do you have any opinion as to whether transitive closure was something that was a required step in this process that Dr. Freer engaged in?
MR. RONA: Objection.
Q. I mean, in other words, if you don't

do transitive closure, is there another way to skin the cat?
MR. RONA: Objection.
A. I think it depends.
Q. Depends on what?
A. Depends on the approach taken.
Q. Well, did you consider other approaches beyond what Dr. Freer did?
A. I wasn't asked to come up with an affirmative determination.
Q. Whether you were -- whether you were asked to come up with an affirmative determination, did you consider alternative determinations?
A. No. That was not something I was asked to do.
Q. Okay. How does transitive closure fit into the Fellegi-Sunter model, if you know?
A. So how Dr. Freer applied it was after -- and we have it in the steps in front of us -- how he applied it was part of the clustering to bring all of those user accounts together, all his pairwise matches, to bring

93

that together and try to form individual clusters, each of which are composed of user accounts, and each user account appearing in one and only one cluster.

Q. You identified certain clusters from Dr. Freer's report in your report; correct? Like the Neves cluster and various other clusters?

A. I certainly included certain clusters as exhibits.

Q. Okay. And did you use certain search techniques to isolate those clusters and determine that you wanted to use them in your report?

A. What do you mean by "search techniques"?

Q. Well, could be query strings or --

A. I certainly used queries to extract that -- those clusters from the data.

Q. Okay. And the query strings, how did you determine what query strings you were looking for? What were you searching for?

A. I think it depends on the purpose of the illustration.

94

Q. Well, pick one of your illustrations. Let's take Maria Neves. What queries were you running for purposes of the Maria Neves cluster?

A. That was the cluster that Dr. Freer stated was representative of the general quality of the dataset -- I think I might have forgotten that word sort of imperfectly -- so that was one that I elected to look at.

(Deposition Exhibit No. 6, Expert Report of Dr. Cameron E. Freer, marked for identification)

Q. I believe if we go to -- let me direct your attention to Exhibit 6. Can you identify that document?

A. I can. It's the Expert Report of Dr. Cameron E. Freer.

Q. All right. And if you turn to Exhibit 3 of Dr. Freer's report, do you recognize that as the Neves cluster?

A. It is Cluster ID 132785.

Q. Okay. And is that the Neves cluster, Maria Neves cluster?

A. It contains names, including -- some

95

of which are Maria Neves.

Q. Okay. And did you attempt to determine or identify any accounts which you considered to be overclustered in this -- in this exhibit?

A. I certainly looked at certain of the user accounts and questioned when they had different names, different addresses and things of that nature, different contact information.

Q. Okay. Well, when you went through that process, did you attempt to determine how many accounts out of the 471 listed in this Cluster 132785 were, in your opinion, overclustered?

A. Not fulsome.

Q. When you say "not fulsome," what does that mean?

A. So I think there are certain accounts in here that more clearly appear to differentiate themselves from other accounts, and there are certain accounts in here that have some different characteristics. I focused on a discussion of those accounts

96

which more clearly differentiated themselves, but that's not to say there aren't other accounts in there that could potentially not belong or not have been created by whoever created the majority of these.

Q. Okay. Well, did you do any analytical work to actually come up with a number that you say, in your opinion, are overaggregations?

MR. RONA: Objection.

A. Not specifically -- I'm sorry -- not specifically. I provided examples of certain accounts that appeared to be more greatly differentiated.

Q. Well, did you form any opinion as to whether or not even those accounts were overaggregations?

MR. RONA: Objection.

A. I think that they differentiate themselves and oftentimes don't comport with the stated -- stated goal of Dr. Freer's aggregation process.

Q. Do you have an understanding as -- did you attempt to calculate any error rates

97

for any individual clusters including the Neves cluster?
MR. RONA: Objection.
A. I attempted to -- I attempted to demonstrate illustrations in the aggregation process itself in the methodology.
Q. That wasn't my question. My question is did you calculate the actual error rate, what you considered to be an overaggregation in any cluster?
MR. RONA: Objection.
A. No. That would not have been possible without determining affirmatively first what I thought a right answer would be -- was supposed to be, and that's not something I was asked to do.
Q. So with respect to all of Dr. Freer's clusters, is it fair to say that there are no clusters in which you can actually say this cluster has an error rate of X --
MR. RONA: Objection.
Q. -- in your opinion?
A. Again, my goal was to demonstrate

98

issues with the process, and it would not have been possible to determine an error rate without first determining affirmative analysis, which is not something I was asked to do.
Q. Did you ever perform an error rate calculation for overclusters generally in Dr. Freer's -- in Dr. Freer's model?
MR. RONA: Objection.
A. I don't even know what that statistic would mean. Again, you'd have to have known what the right answer is, which would require an affirmative analysis and opinion and knowledge of what was right in the first instance.
Q. I'm really just asking you did you perform any error rate calculation overall for which you considered to be overclusters in Dr. Freer's clustering?
MR. RONA: Objection.
A. I certainly provided plenty of examples of instances that appear to be overclustering or instances where his clusterings don't comport with the stated

99

goals.
Q. Did you perform any such calculation for underclusters in your opinion?
MR. RONA: Objection.
A. Same comment. There are plenty of examples in instances where I show two separate clusters that share a high degree of similarity, which calls into question whether or not that was -- those individual user accounts were, in fact, created or the ultimate participant wasn't the same economic actor.
Q. Well, but I'm really focusing on actual error rate calculations, which -- and I'm just trying to confirm that, as you sit here today, you're not in a position to identify what the error rate for overclustering is or the error rate for underclustering?
MR. RONA: Objection.
Q. Is that a fair statement?
A. Again, I think I demonstrate and show instances where the clustering process lacks reliability, where the aggregation

100

methodology is inconsistent with Dr. Freer's approach and stated goals, and I provided indicia of that.
Q. Those are examples. They're ad hoc examples. I'm asking if you went from ad hoc example to a statistical analysis of whether the Dr. Freer's clusters overall are within an acceptable error rate or not?
MR. RONA: Objection.
A. That would not have been possible without first knowing and coming up with what I believe the right answer to be, and in this case, part of the biggest issue is that there isn't a golden standard. There isn't a known set of data that says these are exactly the right accounts that are owned by this person. So without that information, it simply wouldn't have been possible, nor was I asked, to come up with a statistical error rate.
Q. Is it your understanding that data aggregation requires a golden set?
MR. RONA: Objection.
A. What do you mean by "requires"?
Q. That it requires it or otherwise --

101

or the -- or the data aggregation process and output are inherently unreliable.
MR. RONA: Objection.
A. I think it depends on the case facts.
Q. There's no hard-and-fast rule?
MR. RONA: Objection.
A. I think it depends on the case facts. I think within a given circumstance, it certainly could be necessary.
Q. Okay. What about in this case?
MR. RONA: Objection.
A. Potentially.
Q. When you say "potentially," what do you mean?
A. So I think one of the issues here, and this is what I discussed in my report, is that the data itself was not of high quality. There was no testing done of the accuracy or truly the real consistency of the data. So absent knowledge about what is a true match, it's simply not possible from the data itself to know how right or wrong you are.
Q. Are you familiar with the fact that

102

all transitive closure processes have some kind of error rate? It's either over -- a little bit overclustered, a little bit underclustered. You can be very, very conservative, but you run the risk of being underclustered, or you can be very, very liberal, but run the risk of being underclustered. Are you familiar with that, generally speaking?
MR. RONA: Objection.
A. The concept that Dr. Freer's approach is inherently imperfect, sure.
Q. No. No. I'm saying, generally speaking, in transitive closure models, there is always a trade-off between being overinclusive in your transitive rules or underinclusive in your transitive rules. Is that a fair statement?
A. I think it depends on the dataset.
Q. Well, that rule wouldn't apply to certain datasets if you're transitively closing the dataset?
MR. RONA: Objection.
A. Your suggestion is that inherently,

103

as I'm hearing your question, is the transitive closure process necessarily going to either over- or underaggregate, and, therefore, be imperfect, and that's potentially possible or not.
Q. Well, do you have an understanding from the literature as to whether transitive closure always is a balance between overclustering and underclustering and that there is an acceptable error rate that is accepted by data science generally?
MR. RONA: Objection.
Q. Do you have any understanding about that?
A. I think it would be case in fact specific.
Q. Well, did you attempt, in your engagement, to determine whether Dr. Freer's calculation of a .98 accuracy rate was correct or incorrect?
MR. RONA: Objection.
A. What do you mean by "correct or incorrect"?
Q. That it fell within acceptable

104

statistical norms.
A. I think it depends on the case. In this case, the reason he had to do that was because he didn't find or didn't look or didn't search for or didn't attempt to look to real data, real real-world information, so in the absence of having real-world information, that's the approach he chose. The issue with that though is you don't have and didn't account for the fact that the data quality, the accuracy and the consistency were unsupported and unquestioned, so it was a self-fulfilling prophecy.
Q. And when you say real-world data, what are you referring to?
A. True knowledge of what economic actor opened and benefitted which user accounts.
Q. All right. And what true knowledge are you referring to?
A. It would depend.
Q. Well, in this case, what is it that you say Dr. Freer should have done that would qualify as true knowledge?

105

A. I think there are things he certainly could have attempted to do. I'm not aware of him talking to anyone. I'm not aware of him looking at the EPOC data. I'm not aware of him looking at E-wallet data, any indicia, quite frankly.
Q. Well, you've said three things. He didn't talk to people?
A. Uh-huh.
Q. He didn't look at EPOC, and he didn't look at E-wallet data. Anything else?
A. I think there was bank data he could have potentially looked at.
Q. Anything else?
A. Again, I haven't been asked to do this affirmatively, but those are certain things that come to mind, and the fact of the matter is that simply looking at the data itself or even looking at those things doesn't necessarily mean it's complete or accurate. Some datasets are simply not sufficient for a different purpose.
Q. I'm asking anything else?
A. I don't know. It's not a -- I

106

haven't been -- haven't given enough thought to determine what the universe of options are that he could have potentially looked at.
Q. Did you talk to anyone?
A. Not directly.
Q. Well, when you say "not directly," what does that mean?
A. No. I've simply looked at affidavits and other legal filings and depositions. I'm sorry. When you say "anyone," can you be more specific? I think I presumed what you meant.
Q. Any participants.
A. Thank you. No. Not live or remotely.
Q. You do refer to the Frantz Balan affidavit a number of times in your report; correct?
A. Yes. Is it Frantz Balan or Balan? I've always wondered.
Q. I don't know. What is it?
MR. RONA: The record is not going to clarify that so.
Q. I'm just going to say -- I'm just

107

going to say Frantz Balan.
A. Very good.
Q. All right. Did you take any steps to determine whether or not Frantz Balan was available for you to talk to?
A. No.
Q. Did you look at any EPOC data?
A. Yes.
Q. What EPOC data did you look at?
A. The EPOC data that was provided and maintained by Huron.
Q. Okay. And what specifically did you look at in the -- what were you looking for in the EPOC data?
A. So, originally, going back to Mr. Martin, Mr. Martin contended that the EPOC data was supportive of his aggregation, and in that, we looked to determine whether we believed that was the case, and we concluded that was not.
Q. Well, that was your work -- 2018, '19, '20 work?
A. That's why I originally received it and why I originally analyzed it, yes.

108

Q. Okay. And you determined what in connection with that?
A. That Mr. Martin's contention that the EPOC data supported his aggregation was an incorrect assumption and conclusion.
Q. All right. And did you then compare the EPOC data against Dr. Freer's aggregation?
A. To some extent. To some extent, I did, yes.
Q. To what extent?
A. So certain of my exhibits include columns that say -- I think the terminology was claimed, and that relates to instances where at least one user account within that cluster was part of an accepted or agreed claim within EPOC.
Q. And what did you do with that data?
A. I included it in my exhibits.
Q. Okay. But other than including it in your exhibits, did you do any analytical work with that?
A. It certainly was something I explored to look at instances where certain aggregations, particularly things as large net

winners included data that was claimed by somebody.

Q. Well, when you say "I explored," did you come to any conclusions?

A. I'm not sure I came to an ultimate opinion. I certainly saw things that caused me to look more carefully at certain things.

Q. Well, in 2018, '19, '20, you went through the EPOC data and formed an opinion that it did not support the Huron aggregation; is that right?

A. Certainly, it wasn't -- Mr. Martin's conclusion that it supported the Huron aggregation, I thought, was faulty.

Q. Okay. And did you take any steps to then compare the EPOC data against Dr. Freer's clusters?

A. I looked at some individual examples, yes.

Q. Which individual examples? Who?

A. Certainly, at least Benjamin Argueta, A-R-G-U-E-T-A.

Q. And with respect to Mr. Argueta, did you form any opinions with respect to whether

the EPOC data supported Dr. Freer's Argueta cluster?

A. That particular cluster -- we'll call it Argueta 'cause that was probably the name that appeared most frequently -- it did not entirely support it, no.

Q. How did it not support it?

A. So there were a number of accounts that had been claimed as part of that cluster, it may have been 55 or so, in that range, and I was curious to understand why that might be, so I went back and explored the EPOC data, and what appeared to be the case is that as potential or likely overclustering, Dr. Freer had included user accounts with the name -- gosh -- was it maybe Benjamin Oleo [phonetic] or something, something along those regards, it was a different Ben, and someone had gone in and claimed the EPOC, those 55 accounts, which is, again, you know, strong indicia that those 55 accounts represent overclustering.

Q. Well, who claimed the 55 accounts?

A. My recollection was someone by the name of -- I want to say Ben Oleo, but I may

not have that right. I would have to go back and confirm.

Q. All right. Other than this one example, did you do anything else with the EPOC data?

A. I did.

Q. And what else did you do, other than looking at Argueta?

A. I certainly went in and looked at certain commentary that people included.

Q. Did you attempt to sample this in any statistical way, or was this just sort of a random look at varying accounts, or was it targeted in some way?

A. It was probably more targeted. Again, it's for different purposes at different times.

Q. What about the E-wallet, did you look at that data?

A. No. Not specifically.

Q. How about bank data, did you look at bank data?

A. No, not specifically.

Q. Okay. Do you have any sense of how

many accounts had corresponding bank data available to it?

A. I don't -- I didn't look in detail at the bank data, no.

Q. Okay. Do you have any sense of how many E-wallet accounts were available to examine against the Freer clusters?

A. That's not something I examined.

Q. Same question with respect to the EPOC?

A. Yes.

Q. Okay. How many clusters did you examine? I mean -- strike that.

How many EPOC accounts did you examine? You've identified Argueta?

A. Oh, I analyzed the entire EPOC dataset.

Q. The entire EPOC dataset?

A. Yes.

Q. And when you say "I analyzed it," what did you do?

A. So I went through a process of loading that dataset, and in SQL, went and sought to match those -- the terminology of

the EPOC data, the unique identifier was more the login itself, so tie that login back to the aggregation and the user account detail originally in Mr. Martin's aggregation process and subsequently as part of Dr. Freer's aggregation process.

Q. All right. And what was -- what were the match statistics?

A. I'm not sure I understand your question.

Q. Well, you said you loaded the entire EPOC dataset?

A. Correct.

Q. And then you matched the EPOC login to the aggregation?

A. Correct.

Q. All right. And what did that provide you?

A. There's at least one of my -- actually, multiple of my exhibits that I described was claimed. I think one actually aggregates and totals the number of logins by tier that were claimed as part of EPOC.

Q. All right. And did you form any

opinion with respect to Dr. Freer's clustering as a result of this EPOC analysis?

A. I think it was one piece of evidence I considered.

Q. Okay. But what is it specifically that you determined as a result of this exercise as it relates to the Freer aggregation process?

A. I think what it did is it helped inform -- as an example of Argueta -- that there were potential over- and underaggregations.

Q. Okay. But did you attempt to then quantify that or qualify it into some statistical analysis that said there are this many that don't match between EPOC and Dr. Freer's cluster or some other analysis?

A. Not -- not what's beyond that was presented in my exhibits.

Q. So whatever you did is limited to what's in your exhibits?

A. Correct. Well, I guess beyond the more exploratory review that we talked about with Argueta, the quantification reside in the

exhibits.

Q. Okay. And as a result of the quantification exercise, did you attempt to calculate any overaggregation or underaggregation error rates?

MR. RONA: Objection.

A. Not in this case. Again, beyond the examples provided.

Q. Did you do it for the Huron aggregation?

MR. RONA: Objection.

A. In part, I did -- I focused on the analysis at least with respect to the way Mr. Martin had stated it in his opinion.

Q. Did you review the most recent deposition transcripts from participants in the February and March time frame?

A. I received all of those, and some I read more carefully than others, but I reviewed them all.

Q. You read Argueta?

A. I did.

Q. Okay. Did you -- well, strike that. What he said is what he said. I

just want to try to nail down, do you agree that transitive closure is inherent in any kind of aggregation methodology?

A. No. Not necessarily.

Q. You'll agree that Freer's methodology performed a transitive closure step; right?

A. He states it did.

Q. Well, do you agree that that's a fact looking at his code and your --

A. Yes.

Q. -- review? Okay. And do you believe that was a required step in his process -- in the process using the Fellegi-Sunter method, that you would have to go through a closure step?

MR. RONA: Objection.

A. I don't know that I've analyzed the alternatives.

Q. Are there any alternatives, as you sit here today, that you can think of?

A. There's certainly different approaches to how that's performed, assumptions within. Categorically? Again,

it's not something I explored, whether categorical options exist in the universe.

Q. Well, I'm asking you, transitive closure was performed here. Can you think of another alternative step besides transitive closure which could have been used here using the Freer methodology? The Freer methodology requires transitive closure; doesn't it?

MR. RONA: Objection.

A. Mr. Freer certainly used transitive closure, so inherently, his methodology required it.

Q. All right. So I'm trying to determine, are you saying transitive closure could have -- you could have used something other than transitive closure in this analysis, or would you agree that transitive closure is a necessary component to a clustering analysis in this dataset?

MR. RONA: Objection.

A. Again, I haven't sought to look for alternatives or performed my own independent affirmative analysis.

Q. As you sit here today, is it fair to

say you can't think of another analysis --

A. I certainly have --

Q. -- tool?

A. Again, using the term "analysis tool." There's a difference between an analysis tool and a conceptual process, so I'm having a hard time maybe disaggregating that.

Q. Let's go with conceptual process. As you sit here today, can you think of another conceptual process besides transitive closure which could be used in clustering this dataset? And by "clustering," I mean taking all of the individual accounts and assigning them to a participant.

A. Again, I haven't sought to think about the universe of potential approaches to bringing things together.

Q. So, again, as you sit here today, you can't think of one; is that a fair statement?

A. Can you ask your question in full again, please?

Q. Sure. As you sit here today, you can't think of an alternative to transitive

closure for this clustering process?

MR. RONA: Objection.

A. So Mr. Martin performed an aggregation, you can call it a clustering, if you will. He certainly -- I don't agree with how he did it and the approach he used, but he certainly never used the word "transitive closure" as part of that process, so that would be an alternative. I'm not saying it's a better or a worse one, but it's a different approach. It certainly didn't rely on the terminology of transitive closure.

Q. Well, I'm trying to determine if you want an aggregation output for the Telex dataset, as you sit here today, can you think of anything that does not require transitive closure to get to a full cluster, full set of clusters?

MR. RONA: Objection.

A. I guess I don't know what you mean by "a full set of clusters."

Q. "A full set of clusters," by that I mean assigning each account to a cluster and only one cluster.

A. So I go back to Mr. Martin, again, I don't agree with his output and what he did for the reasons I stated in my original report, but, again, that was a process he performed without ever using the word "transitive closure" anywhere in his analysis.

Q. No. I understand, but you disagree with his analysis, so I'm going to put that to one side.

As you sit here today, can you think of another analytical process which would -- which would aggregate every single Telex account to an individual participant but only one individual participant that would not require transitive closure?

MR. RONA: Objection. I think he's answered the question.

A. Again, I haven't endeavored to go and look at the universe of possible approaches. Certainly, the people that Mr. Martin attempted to do it without, you know, again, I don't agree with his output, but he certainly used an approach without using the terminology "transitive closure."

Q. I understand that you could argue that he used some form of transitive closure process. I'm taking the Huron analysis and putting it to one side.

A. Okay.

Q. I have the Freer transitive closure process right here in front of us, and I'm asking, generally speaking, when you are clustering data, when you're creating clusters of individual -- it could be accounts or individual medical results or individual whatever, datapoints, doesn't it always require some degree of transitive closure?

MR. RONA: Objection.

A. The goal of what Mr. Freer was asked to do was to put individual user accounts together and form a whole participant, an individual economic actor. I don't think he did that all the way. I think he stopped short the clustering level, but his stated goal was to take each user account and put it in one participant or one cluster and only one cluster.

Q. Right. And my question is doesn't

that necessarily require transitive closure?

MR. RONA: Objection.

A. Again, I don't know and haven't attempted to research what other options may exist in the universe. I don't know one way or the other.

Q. So you have no opinion on that?

A. I don't have an opinion on whether or not it was necessary to -- absolutely necessary to use transitive closure in every instance, other than the fact that Mr. Martin attempted the same goal and never said the words "transitive closure."

Q. Okay. And, remind me. Have you ever done -- or strike that.

Have you ever been engaged to perform analytical work that requires transitive closure?

A. I don't think I've ever used that specific terminology in a report.

Q. Okay. Regardless of whether you've used the specific terminology, have you used transitive closure in any of your prior engagements?

A. I think you asked me this earlier, and my recollection now, as it was then, is likely, but I can't think of anything specifically, sitting here today, over the last 20 years.

Q. When you say "likely," why do you say that? You have a general recollection of having used it?

A. Yeah. Yes. That's right.

Q. Okay. And your general recollection is in what context?

A. Certainly, I'm trying to think of examples where we have sought to group or aggregate or deduplicate things and have looked to try to combine in a way that's consistent with transitive closure to the extent there's A is attached to B and B is attached to C, such that A and C are now grouped. I feel like that's something I have done, but I'm just not thinking of -- I can't think of a specific example sitting here today. I'd have to go back and look.

Q. All right. And you wouldn't have done it unless you considered it to be a valid

tool; correct?

A. Well, I'm not sure --

MR. LYNE: Objection.

A. -- I'd call it a tool. It's a --

Q. Concept?

A. -- concept. It depends on the case in facts.

Q. Right. But you wouldn't have done it in whatever that engagement might have been unless you considered transitive closure to be an appropriate concept to apply to that engagement --

MR. RONA: Objection.

Q. -- fair statement?

MR. RONA: Objection.

A. Yes. I think that's fair. Again, how you apply it is important, but conceptually, that's true.

Q. Would it be fair to say that for the stated goal, applying transitive closure to assign every account to an individual cluster but only one cluster, that transitive closure will necessarily result in some degree of overaggregation and some degree of

underaggregation? It's inherent in the process, in other words. Is that a fair statement?

MR. RONA: Objection.

A. I don't know. It's a fair statement in every situation.

Q. How about in the context of Dr. Freer's model? Dr. Freer using transitive closure under these circumstances, would it be fair to say that it would be inherent in Dr. Freer's transitive closure process that there would be some level of overaggregation and some level of underaggregation?

MR. RONA: Objection.

A. Absolutely. I certainly agree with the fact that Dr. Freer's analysis has a level of overaggregation and a level of underaggregation. I think I've made that pretty clear.

Q. Okay. But what you can't say is what the acceptable level of an error rate would be in overaggregation and underaggregation in this instance --

MR. RONA: Objection.

Q. -- 'cause you haven't done the work; correct?

MR. RONA: I think we've covered this, but same objection.

A. As I said, it would not be possible to do that work without having come up with an affirmative solution, and, frankly, without knowing more about the data itself and imparting real-world information because the data itself lacks accuracy. Certainly, Dr. Freer hasn't demonstrated its accuracy nor its consistency.

Q. The accuracy is talking to people, looking at EPOC, looking at E-wallet, and bank data?

A. Those are certain things he could have -- could have done. I'm not sure that would have necessarily been sufficient, nor is it necessarily, you know, extensive on what else he could have done.

Q. Well, what else could he have done?

A. I don't know. I haven't been asked to look at that.

Q. Well, as you sit here today, can you

think of anything else that he could have done?

A. Again --

MR. RONA: Objection.

A. -- that was not something that I was asked to think about or opine.

Q. Okay. Does Fellegi-Sunter with equal -- strike that.

Does FS-EM require an external ground-true dataset, as you understand it?

A. Define "require."

Q. Is it part of the FS-EM process?

A. It can be.

Q. In what circumstances -- when you say "it can be," is it -- well, strike that.

I think I asked you previously can you explain how FS-EM differs from the Fellegi-Sunter model generally, and I thought you said you weren't sure.

MR. RONA: Objection.

A. My answer was in the context of the overall process, which would include testing its reliability.

Q. Well, did you attempt to determine

how the FS-EM model deals with the lack of a gold standard dataset or ground-truth dataset or however you want to define it?

A. Can you ask that question one more time?

Q. Sure. Does FS-EM require a gold standard --

MR. RONA: Objection.

Q. -- as you understand it?

MR. RONA: Objection.

A. I don't know if I've thought about whether or not it's a requirement. Certainly, you can do an analysis. It doesn't make the analysis accurate or correct.

Q. Okay. Well, you've given me four things that you say Dr. Freer could have done. He could have talked to some number of people. Is there a minimum number of people that he would have had to have talked to for it to be statistically significant in your opinion?

A. That's not something I attempted to compute.

Q. Is there a minimum number of EPOC accounts that he would have had to have

reviewed for it to be statistically significant in your opinion?

A. Again, that's not something I've attempted to calculate.

Q. Is there a minimum number of E-wallet transactions that he would have had to have looked at for you to believe that it was statistically significant?

A. Again, that's not something I attempted to look at.

Q. All right. How about bank data, bank accounts, would there be a minimum number of bank accounts that he would have had to have sampled for it to be statistically significant in your opinion?

A. Again, that's not something I attempted to look at.

Q. Okay. Can we agree that it's a standard technique in data science to perform a pairwise match assessment followed by transitive closure?

MR. RONA: Objection.

A. It's certainly a technique used. I think standard would require the case facts in

terms of whether it would be applied.

Q. Okay. But it's, in general, an acknowledged process in the data science world for data aggregation?

MR. RONA: Objection.

A. Generally, I think its application would be case and fact specific or its applicability would be case and fact specific.

Q. But it's an acknowledged tool in data science? Yes?

A. Generally. Certainly.

Q. Okay. And that process does require first a pairwise match and then transitive closure; right?

MR. RONA: Objection.

A. Which process -- which process generally are you referring to?

Q. I'm talking about using transitive closure. First, you got to start with a pairwise match, then you got to follow with transitive closure?

MR. RONA: Objection.

A. That's certainly the analysis that Dr. Freer performed, yes.

Q. And that's a generally acknowledged toolset in data aggregation science; right?

MR. RONA: Objection.

A. I think it would depend -- excuse me -- it would depend on the field of science, and Dr. Freer cites to various articles.

Q. All right. Can you cite to articles that says that's not a valid data aggregation methodology?

MR. RONA: Objection.

A. Not specifically, no.

Q. When you say "not specifically," as you sit here today, can you cite to any such?

A. No. No.

MR. LYNE: Okay. It's 12:30. Do you want to have a 45-minute break for lunch?

MR. RONA: Sure. Is that good for you?

THE WITNESS: Yeah. Whatever you think works best.

(Lunch recess was taken from 12:30 p.m. to 1:26 p.m.)

(Deposition Exhibit No. 7, One-Page

Document, marked for identification)

MR. LYNE: Mark that as an exhibit.

(Deposition Exhibit No. 8, Document Entitled Exhibit 2, marked for identification)

MR. LYNE: Okay. Back on.

Q. Sir, during the break, you've had a chance to reflect back on the answers you gave me prior to lunch. Is there any answer that you believe you need to change to make it either more accurate or more complete or to correct an error?

A. I think the only thing that was coming to mind was the discussion of error rates, and I tried to think about, you know, where in my report I'd addressed error rates, and I think the one thing that I was working through was, you know, the error rates and the things I looked at were on -- and I think this is the appropriate way to look at it -- on a specific cluster basis. Right? What is the impact, what is the error -- potential error rate on a given cluster? And I think there

133

are a number of examples in the report where I do specifically quantify more on the impact to net equity side of things, what the impact on specific, you know, clusters are of over- and underclustering.

Q. But that's only as to the specific clusters that you have chosen to include in your report; correct?

A. The examples I provided where I quantified, yes.

Q. Which are the -- that isn't every single cluster that you examined in the course of your engagement, however; correct?

A. No. There are certainly other clusters that I didn't -- that I saw potential problems in that I didn't include as examples. That's correct.

Q. And there are other clusters that had no problems; correct?

A. I don't know that's true at all.

Q. Well, did you look?

A. Did I look at what? I'm sorry.

Q. Did you attempt to quantify how many clusters in Dr. Freer's aggregate clusters had

134

no issues that you could discern?

MR. RONA: Objection.

A. I think that's part of inherently the problem is that the data itself is unreliable, so to look at clusters that are based on unreliable data without bringing in potential other information to help answer that question, it's not possible to determine with certainty that this is accurate or not.

Q. Let me ask you the question again. With respect to all of the clusters, did you attempt to determine, irrespective of the accuracy of the information, attempt to determine whether or not the majority or the significant majority or the minority of those clusters were clusters that, on your visual examination, showed no issues?

MR. RONA: Objection.

A. So in this -- in this hypothetical, you're assuming that the data -- the underlying data was accurate, that it had been tested and demonstrated to be accurate.

Q. I said irrespective of accuracy. This is the clustering exercise. Did you

135

attempt to determine how many of Dr. Freer's clusters showed no apparent error, at least with respect to the clustering itself?

MR. RONA: Objection.

A. Again, that would require me to have had a fulsome analysis of my own opinion about what the clustering should have been, which I wasn't asked to do, and it's not clear to me that -- whether or not it's even possible 'cause I wasn't asked to do it.

Q. You identified and discussed a relative handful of clusters where you point out what you considered to be overclustering issues or underclustering issues. And my question is did you attempt in any way to determine with respect to all of the other clusters that Dr. Freer created in his clustering analysis how many of those showed no apparent overclustering or underclustering --

MR. RONA: Objection.

Q. -- just from the cluster itself?

A. Again, that would require me to have done an analysis on the affirmative to

136

determine what I believe the appropriate clustering -- over- and underclustering would have been.

Q. No. I'm saying independent of what you consider to be accuracy questions. Assume that the data itself is accurate. Did you attempt to look at and quantify with respect to all of Dr. Freer's other clusters how many of those showed no apparent overclustering or no apparent underclustering?

MR. RONA: Objection.

A. Again, I would go back to my statement that in order to determine what has no apparent under- or overclustering, that would require me to have done an affirmative analysis.

Q. So the answer is you did not undertake that exercise? You didn't look for ones that appeared to have no overclustering or underclustering issues?

MR. RONA: Objection.

A. I looked and, I think, explored the data. If your question is did I manually review all 15 million user accounts and

1.5 million clusters and then do a quantification of which, on their face and assuming the data was accurate, which I don't think it is, which looked on its face to be potentially okay, that's not an analysis I performed.

Q. Okay. Did you perform that analysis on any of the clusters looking for clusters that appeared to be okay, to use your term?

MR. RONA: Objection.

A. I certainly looked at clusters, and the goal was to look at how well or to what extent or find indicia of the methodology, right? It's not just about whether this one cluster or that one cluster appeared to have problems, but does the methodology appear to have reliable results, not just in the aggregate, before a given participant, and what we found was that there was significant deviation for some participants more than others and in both directions.

Q. Okay. And you were able to quantify that for all of the clusters?

A. Again --

MR. RONA: Objection.

A. -- I couldn't have done that for all the clusters without either manually reviewing all two million [inaudible].

THE STENOGRAPHER: "Without either"?

A. Manually reviewing all 1.5 million clusters -- excuse me -- 1.5 in Dr. Freer, or doing my own affirmative analysis.

Q. Okay. Well, did you manual -- how many clusters did you manually review?

A. Oh, gosh. I don't know. Over the course of four years, I guess in clusters, in Freer's terminology over the last, however long it's been, maybe eight months, nine months, lots.

Q. Give me sort of a ballpark. "Lots." How many? Over a thousand?

A. Probably less than a thousand, but maybe -- maybe close. Maybe close.

Q. Okay. So something close to a thousand clusters you reviewed manually?

A. You know, that's -- again, it's a guesstimate. It's hard to say "reviewed"

because there was certainly queries we performed. I didn't always review each cluster for a singular purpose.

Q. No. No. I understand, but I'm just asking as to Cameron Freer's -- Dr. Freer's cluster set, you think you reviewed something close to or around a thousand clusters? Is that a fair estimation?

A. Honestly, I don't -- I don't know that it is. It's hard to think about every time I performed a query or looked at something over the course of the last however months.

Q. Okay. Well, you previously said you thought a good guesstimate was around a thousand?

A. Potentially. I'm thinking about it now and thinking about if I feel that's an accurate depiction, and the answer is -- I think a better answer is I just don't know.

Q. But it's more than the clusters that you included in your report?

A. Oh, certainly.

Q. By a substantial number; right?

A. Absolutely.

Q. Okay. And that's just with respect to Dr. Freer's clusters, not the Huron clusters; right?

A. In terms of how many I reviewed?

Q. Yes.

A. Yes. I mean, obviously, I reviewed more if you look at it in a broader swath of time.

Q. Okay. And can you guesstimate roughly how many queries you made against the dataset over time for Dr. Freer's clusters?

MR. RONA: Objection.

A. I guess I have a hard time -- what do you mean by "queried"?

Q. Running a query. Running a search.

A. How many times I hit execute in SQL?

Q. Yeah.

A. I couldn't begin to estimate that.

Q. "Lots" would be a --

A. Yes.

Q. -- fair statement?

A. I think "lots" is a fair statement.

Q. Okay. Just so I understand, what

you did is you wrote some SQL code, and you put it on top of the Excel database and then ran queries against that database or --

A. No. No. That's not right.

Q. Which database did you run it against?

A. So the database that I looked at was the same source database that Dr. Freer used.

Q. Okay.

A. Over -- I'm sorry -- overlaid with his conclusion as to what the clustering should be. So I wanted to look at the source data, and there were multiple tables, but the accounts table being one of the primary ones, overlaid with Dr. Freer's determination to see how those things meshed up and what his actual aggregations were.

Q. But your searches were against the actual clusters themselves; right? Find me all the clusters that show these characteristics?

A. That's right. But his clustering output, the Excel files, didn't necessarily have all of the underlying detail, so that

needed to be joined with back to the source data in order to see a fulsome picture.

Q. All right. When you say it lacked -- it required joining other source tables, what source tables were you joining?

A. So primarily I utilized the accounts tables, which is the same table that Dr. Freer -- and really the only table that Dr. Freer utilized -- that's not entirely true -- that's what he utilized for his aggregation. He also utilized Mr. Martin's conclusion as to net equity for each of the user accounts, I believe, within the rep_id summary table, and that was a table, amongst others, that Huron generated and Mr. Martin generated.

Q. So the tables you joined were the accounts and the net equity table?

A. Those are two of the primary ones. There are other tables in there that I may have, at times, pulled information from.

Q. Do you remember which ones?

A. They're listed in my report.

Q. But the ones you actually joined, is it listed in your report?

A. It is, yes.

Q. Can you show me?

A. Of course. All right. So we are going to page seven.

Q. Okay.

A. So if we're looking at the second solid bullet, there are four tables listed there. I've defined them as the "Huron Data Tables." Each of those are data tables that were generated by Mr. Martin and included in his initial expert report. Those were provided to us and ultimately Dr. Freer as well. Those all have information that are based upon resulting analysis of the tables that are listed below under the fourth solid bullet defined "Freer Data Tables," so that would include the "representante" which is the account table I previously referenced, as well as the (inaudible).

THE STENOGRAPHER: Matura?

MR. LYNE: Fatura,

F-A-T-U-R-A.

A. Thank you. And "transferencia," the transfer table, and the bonus table -- that

one is an easy one -- the bonus table. So those all contain different components of the TelexFree transactions and accounts and invoices that all come together. And so I used different components of each of those to understand the full scope and breadth of what information we had about user accounts and how they interacted with TelexFree with other potential user accounts and participants.

Q. Okay. Thank you. Now, let's go to Exhibit 7. And this is just a tree diagram showing at the top the entire TelexFree database of 17-plus million user accounts, and it goes down to a list of 1,566,383 participants or separate clusters, and then broken down between net winners and net losers. Do these numbers appear to be correct?

MR. RONA: Objection.

A. Do you mind if I just --

Q. Yeah. Sure. Go ahead.

A. And by "correct," I'm not saying that they are accurate or I agree with them, simply that they are consistent with

145

Dr. Freer's calculations. This is right after we go through the steps.

Q. Yup.

A. And here we are. Thank you. On page 20. This won't have a complete picture, but it will give me the right frame of reference.

Q. Yup.

A. Six, one, seven -- I think there's a missing comma, but fine. All right. And then the net winner and net loser -- sorry -- I was reading the numbers quietly to myself. Yes. They're consistent with what I have in my report.

Q. And with respect to doing accuracy checks, as you've defined it, talking to people, looking at the EPOC, looking at bank records, or the electronic wallet, the E-wallet records, do you have any opinion with respect to how many, for example, individuals you would have to talk to to have a statistically significant sampling that would have some meaning for purposes of determining accuracy of information?

146

MR. RONA: Objection.

A. Well, it would certainly be a nonzero number, which is currently the number of real-world datapoints he's brought in.

Q. One? One person?

A. I would probably think more than one. One is more than zero. I don't think one would be sufficient, if that's what you're asking.

Q. Well, what, in your opinion, would have been sufficient?

MR. RONA: Objection.

A. I didn't do that analysis, and it's not clear to me that's even the right way to look at it. Again, I think, you know, as I said earlier, I think error rate is not just a function of, you know, how many of his clusters are right or wrong, but there's an element of impact on individual people, right? At the end of the day, we're asking an individual person to be a net winner and net loser, and if you tell me that their net winnings are $800,000, but you look at the data and you say, gee, there's a lot of

147

uncertainty here. You know, based on just a cursory look, without talking to someone, it would seem reasonable based on Dr. Freer's own approach that it should have been half of it, that's an overstatement of 50 percent, so I think that it's important to think about things in that context.

Q. So let's focus on what I was asking, which is the accuracy check. Okay?

A. Correct.

Q. All right. And not the impact of any errors. I'm just asking with respect to sampling the data to determine its accuracy, I just want to confirm that you have no opinion with respect to how many people you would have to talk to or how many records you would have to look at to come to a statistically appropriate sampling size that could be validated under basic statistical principles?

MR. RONA: Objection.

A. That's not something I've endeavored to calculate, other than it would be a nonzero number.

Q. Okay. But a nonzero number is as

148

much as you can opine to at this point in time?

A. That's correct. I didn't endeavor to calculate, again, 'cause I'm not sure it's relevant as the impact of individuals, what the overall error rate would be, how that would -- how you'd calculate it truly in terms of what that would even mean or whether it's on a number of accounts or a dollar value. There's a lot of ways you could potentially look at it.

Q. And you have no opinion as to which way to do it; correct?

A. I think I'd want to look at it multiple different ways.

Q. Well, name all the ways that you say you would look at it.

A. I think I'd want to -- all the ways I'd look at it. So let's break that out a little bit. So I think I'd want to look at the dataset for accuracy with as much data as I had available to me in the first instance, and then determine whether or not I thought that data was sufficiently representative and

149

informative as to come to a conclusion, in this case, about how and to what degree I can rely on the data overall.

Q. Well, whether it's statistically representative or not is really a function of whether the data -- the sample size is large enough; right?

MR. RONA: Objection.

A. Could be. And that number may need to be a large number or a small number.

Q. Okay. But you're not in a position to say whether it's large or small; correct?

A. I didn't do the analysis.

Q. Okay.

A. I wasn't asked to.

Q. Okay.

A. I'm not sure it's the appropriate analysis.

MR. LYNE: Well, I'll move to strike the last part.

Q. Do you have 8 in front of you?

A. Exhibit 8? Are we on 8 yet?

MR. LYNE: Did you mark 8?

THE STENOGRAPHER: I did.

150

MR. LYNE: What did I do with it?

A. Thank you.

Q. Mr. Freer, I'll represent to you that this is Exhibit 2 to your -- excuse me -- report which is documents considered, and I just want to ask a couple of questions here.

Is this a complete list of the documents and other materials considered by you in connection with forming your opinions as contained in your report?

A. I think you referred to me as Dr. Freer --

Q. Oh, geez --

A. -- which I --

Q. -- let me try it again.

A. No. I can answer your question assuming that you meant me and not his documents considered --

Q. I did.

A. -- to save you the time. But, yes, this looks consistent with my documents considered list.

Q. Okay. And with respect to the

151

affidavits that you referred to in the last portion of legal filing or in the middle of the legal filings, the Affidavit of Lusette Balan, Frantz Balan, and Sandro Paulo Freitas, F-R-E-I-T-A-S, did you talk to any of those three people?

A. I did not.

Q. Did you direct anyone under your direction or control to talk to any of those people?

A. No, I did not.

Q. Did you ask your counsel whether or not those people were available for interviews?

MR. RONA: Objection.

A. I don't believe so.

Q. Did you take any steps to determine the accuracy and truthfulness of the information contained in any of those three affidavits?

A. I certainly -- at the very least for the Balans, I certainly looked at the data that was aggregated to them and data that was not aggregated to them in and around their

152

accounts.

Q. And did you form any opinion as to whether or not the accounts attributed to them were overaggregated, underaggregated, or correct?

MR. RONA: Objection.

A. I'd have to go back and look specifically. I certainly recall that at least with respect to the Frantz -- Frantz Balan's affidavit, he had indicated in his affidavit there were certain accounts with certain e-mail addresses that he, in his sworn testimony, said weren't his that were aggregated to his account, which would imply overaggregation.

Q. Okay. But did you take any steps to determine if he was accurate in his statements?

MR. RONA: Objection.

A. Other than confirming that the data, as he described it, what was aggregated to him or aggregated to his wife was consistent with what is shown in the data and the description of those was consistent with the data.

153

Q. When you say the description was consistent with the data, you mean he referenced a particular account with a particular e-mail address, and there was an account with the e-mail address?

A. That's certainly one thing.

Q. Well, what are you referring to when you said you confirmed? What were you confirming?

A. He describes certain situations, and by "situations," I mean user accounts, based -- sometimes based on, I think, e-mail, sometimes based on the names, and I went to the data and confirmed his description of those accounts was accurate.

Q. When you say I went to the documents, do you mean you went to the individual e-mails?

A. No. If I said "documents," I meant data.

Q. Data. And "data," you mean what was contained in the various tables?

A. Correct. The tables we walked through earlier.

154

Q. Not that you went and looked at secondary materials, like e-mails that he discusses and the like?

A. I don't think he discussed --

MR. RONA: Objection.

A. -- any -- I don't think he discussed any e-mails. I think he described an e-mail address.

Q. Okay. Can you go back to Exhibit 5, please, which is the ten-step Freer methodology.

A. Yes, sir.

Q. Am I correct in understanding that the net equity calculation, to the extent used by Dr. Freer, was used only at step ten?

A. I think that when -- as part of step ten, I think that's -- the net equity portion is true. What I'm trying to recall is whether he also used that table to isolate or drill down to those users accounts that had one paid invoice. So going from the 17,000 -- excuse me -- 17 million user accounts down to the 15 on the paid invoice, I don't remember if he relied on the rep_id summary table for that

155

step as well or -- I'm sorry -- it's actually right here in the -- in the diagram. So he utilized the rep_id summary table in both the filtering to the TelexFree members as well as to the -- it's in step ten, so he uses it at the start of the analysis and at the end of his analysis.

Q. Okay. The rep_id summary table, did you review that Huron table?

A. Yes, I did.

Q. All right. And did you determine whether or not the rep_id summary table appeared to be correct?

A. I did.

MR. RONA: Objection.

Q. And what was your conclusion?

A. That the rep_id summary table was not correct, that the net equity contained therein was inaccurate.

Q. Okay. And that's for the reasons you set forth in the last part of your report; correct?

MR. RONA: Objection.

A. Yes. And all the reasons I stated

156

in my initial Martin report that I submitted.

Q. The table marked as Exhibit 5 shows that the entity information was exported from rep_id summary. Do you see that?

A. (Witness nods head).

THE STENOGRAPHER: Is that yes?

A. I'm sorry. I didn't know his question was done. I apologize.

Q. Let me ask it again. Do you see that the rep_id summary table at step one, the only information that was exported was entity information?

A. I do see that, yes.

Q. And not net equity?

A. That's what it indicates here.

Q. All right. And is that consistent with your exploration of Dr. Freer's methodology?

A. I think that's right, yes.

Q. Okay. So with respect to steps one through nine, none of those steps relied upon Huron's net equity calculation; correct?

MR. RONA: Objection.

157

A. There are -- I don't believe so. I don't believe so. I believe the net equity calculation was brought on or attached at the end. They relied on Huron and Mr. Martin's analysis in other respects, such as the entity on this part, step one -- actually, before step one. Step zero.

Q. Did you take any steps to determine whether or not the entity information in the rep_id summary table was correct or incorrect?

A. My recollection is I reperformed the analysis. I didn't form an opinion one way or another about whether that analysis was accurate or reliable for purposes of distinguishing the Ympactus accounts from the hybrid accounts from the TelexFree accounts.

Q. So you just formed no opinion?

A. I was not asked to form an opinion on that -- on that topic.

Q. Okay. I want to go back to the Maria Neves cluster, which is attached, I think, as Exhibit 3.

A. 3 to Dr. Freer's report?

Q. 3 to Dr. Freer's report.

158

A. Yes. Do I have that?

MR. RONA: Exhibit 6.

Q. You do have it. It is Exhibit 6.

A. Thank you so much.

MS. PAPAS: We also printed Freer's report.

MR. LYNE: If this is easier, use that one? Okay.

A. I'm more familiar with my own in terms of the layout if that's available, but if --

Q. Okay.

A. Yeah. It's just a little bit easier for me to read wide body like that.

(Deposition Exhibit No. 9, Document Entitled Exhibit 16, marked for identification)

Q. Okay. If you need a (gesturing).

A. Thank you. I very well may.

Q. I usually need them. So the question here, I think, is of the 471, I think, rows in Cluster 132785, did you attempt to determine or can you tell me how many of these accounts you believe are overaggregated

159

for this account?

MR. RONA: Objection.

A. My goal was not to determine how many in totality were overaggregated but demonstrate the process that Mr. Freer utilized contains issues, instances where the individual user accounts that he has placed in here, on their face, don't appear to be the same participant. And in doing that, there's at least two different groups, the first five listed here on Exhibit 9, that while they share -- in fact, it's more than that if you look below -- the first five, at a minimum, you know, may or may not have similar names, some do; some don't, but don't share -- or share very little, if nothing else, in terms of other potential identifiers.

Q. Okay. But out of 471 or whatever the number is, account entries, only five of them showed evidence of overaggregation. Would you consider that to be an overclustering problem?

A. Yes, I would.

Q. So from your perspective, an

160

acceptable cluster has to have a zero overclustering or underclustering error rate?

MR. RONA: Objection.

A. I haven't been asked, nor have I defined "acceptable."

Q. Well, can you define "acceptable" to me now?

MR. RONA: Objection.

A. I don't think I can define it in the context of how wrong someone can be specifically in terms of number of accounts or dollar value. I can tell you that the methodology, right, the methodology extends not just in this individual example, which is the one example Dr. Freer provided, contains errors --

Q. Yes.

A. -- or apparent errors.

Q. My question is am I correct in understanding that you have no opinion, as you sit here today, how much of an error rate makes a clustering methodology unacceptable versus acceptable?

MR. RONA: Objection.

A. I have not attempted to put a number or value on the error rate across the board. What we see is the methodology contains zeroes. Some people it impacts significantly; other people less so.

Q. Well, did you attempt to quantify how many people had significant impacts and how many people had less impact and how many people had no impact?

MR. RONA: Objection.

A. No. Because, as I said earlier, in doing so, that would require me to come up with an affirmative opinion about what I believed the proper clustering and aggregation would actually be.

Q. And that opinion you do not have currently?

A. I was not asked to do an affirmative analysis in this case.

Q. Okay. I'll ask you this question. I think you may have already answered the underlying questions, but let me ask it anyway.

Suppose an aggregation methodology

has a precision of 0.980, a recall of 0.996, and an F-measure of 0.988, how often would you expect to see false positives in that population?

MR. RONA: Objection.

A. I think it depends. I think that those measures are inherently estimates because they lack real-world information such that it's simply a self-fulfilling prophecy inherent in the data itself. If you utilize unreliable and inaccurate data, then you can't truly know how wrong or right your model is.

Q. I wasn't asking about the underlying data. I was asking given the aggregation methodology that's used that shows a precision of 0.980, a recall of 0.996, and an F-measure of 0.988, how often would you expect to see false positives under that circumstance?

MR. RONA: Objection.

A. I think you can't answer that without knowing truly what is accurate, and those measures don't tell you, because you never looked, about how accurate the data on which it's based, so to calculate false

positive in those measures, I don't think, yield reliable answers.

Q. If I gave you a piece of paper and a pencil and just asked you to calculate the expected false positives under those parameters, would you be able to do it?

MR. RONA: Objection.

A. I don't know. That's never something I had endeavored to do with a piece of paper and a pencil.

Q. Well, let me give you a piece of paper and a pencil and give it your best shot. Do you know how to do it?

MR. RONA: Objection.

A. Honestly -- honestly, I'd probably go back and look at the definitions of how Dr. Freer specifically calculated them, look at the formula, and with that, I'm pretty confident I could come up with that answer.

Q. Do you know what the formulas are?

A. I have not memorized them, no.

Q. Have you ever calculated false positives off of precision, recall, and F-measure?

A. No, I have not.

Q. Okay. Let me give you a hypothetical. In a given aggregation methodology, in your opinion, would a single missed link, either overaggregation or underaggregation, would that invalidate the methodology?

A. Potentially.

Q. Under what circumstances would a single missed link invalidate the methodology?

A. I think it would depend on the methodology and what you're trying to show and the size of the data and the case-specific facts.

Q. Okay. Well, how would the size of the data impact that?

A. I think that it would be one factor, among many, to think about.

Q. Well, what would be the important factors?

A. It would depend on the case.

Q. Well, generally speaking, what are the important factors?

A. I think those all could be

potentially important factors, depending on what you're trying to show.

Q. What's the most important factor in the Telex database for determining whether a single missed link would invalidate the aggregation methodology?

A. I think I've lost track of how we were defining "factor." I apologize.

Q. Well, you said that a single missed link in an aggregation methodology could potentially invalidate the entire methodology. So you said -- you said it depended on the circumstance. So I said let's assume it's the Telex database --

A. Uh-huh.

Q. -- what are the factors which you consider to have primacy in determining that question?

A. I think I'd want to look at the impact of the missed link. I think I'd want to look at the purpose and cause of the missed link. I think those would be -- and by "purpose and cause," I mean what is the underlying reason. I think I'd want to go

back and understand whether it's derived by the source data itself, whether it was a function of the way the program worked or the methodology worked. I'd want to really seek to understand, you know, whether it was a reasonable and reliable methodology that was applied in a way that was consistent with the case facts and based upon reliable and accurate data.

Q. Generally speaking, would the number of missed links be generally proportional to the false negative rate?

A. I'm not sure I could make that generality.

Q. What would you look to to determine what the factors are to determine the linkage between the number of missed links and the false negative rate?

MR. RONA: Objection.

A. I'm sorry. I don't understand your question.

Q. Well, let's assume a given number N of missed links in a dataset of some size. Is that number of missed links proportional to

the false negative rate or to something else?

MR. RONA: Objection.

A. I guess I'm not totally clear on what you're asking me. Is it proportional to the false negative rate? I'm sorry. I don't understand the question.

Q. Let me try it differently. To validate, in your opinion, any aggregation methodology, would you be required to look for every missed link and then determine its impact, its cause, its source, and how it relates to the methodology used?

MR. RONA: Objection.

A. I guess it would depend on the case.

Q. Well, how about this case?

MR. RONA: Objection.

A. I don't know. I haven't endeavored to think about the number of missed links one would need.

Q. Well, regardless of the number of missed links, are you suggesting that every missed link would then have to be subject to an analysis of the impact of that missed link, the cause of that missed link -- missed link,

the source of that missed link, and the methodology and how it relates to the missed link?

MR. RONA: Objection.

A. I don't know that it's a one-by-one basis. I think I'd want to -- if I saw errors in my analysis, clear errors, I'd certainly want to go and understand why. I'm not sure how many times that needs to occur.

Q. Well, this goes back to the issue of error rate. If you have a .99 or a quadruple nine, nine, nine, nine, nine accuracy rate, and a point zero, zero, zero, zero one error rate, generally speaking, people tend to say that's a pretty good number, quadruple nines, I think up to six nines, six sigma. My question, I guess, is am I correct that you have no opinion as to what an acceptable error rate is in this aggregation of data?

MR. RONA: Objection.

A. Again, I think it depends on how you're determining the error rate. I appreciate the math that Dr. Freer performed. He did so on a methodology and using only data

that's contained within a database where he didn't quantify or analyze the accuracy of it, so any determination of error rate that utilizes that inaccurate data is going to be equally inaccurate. So I think it depends on what kind of math you're doing, understanding the case facts and the data itself, and also not just thinking about the totality across all users, but thinking about the error rate in the context of how it impacts individual people.

Q. So but individual people, of course, can bring that to the attention of the trustee in this case through the E-portal; right?

MR. RONA: Objection.

Q. The claims portal?

A. I think it depends on what process you're discussing.

Q. Okay. Well, let's assume that it's a process that's similar to what was used in the net loser process. So there's an E-portal, and a net winner can go in, and they can either validate the cluster or they can claim an overaggregation or an

underaggregation. Does that provide, in your opinion, the necessary accuracy check?

A. If you're --

MR. RONA: Objection.

A. If you're asking similar to the way the claims process worked, if you put the onus on the individual potential net winner to go in and do Dr. Freer's work for him and confirm whether it's right or wrong, they came in and said, yes, I agree those are my accounts in each and every case, then you're relying on real-world data. I agree with that.

Q. Okay. And that's the real-world check that you're looking for; correct?

MR. RONA: Objection.

A. No. It's not correct.

Q. How is that incorrect?

A. If you asked someone to go in and demonstrate that those are their accounts, that would certainly be helpful information.

Q. Well, besides being helpful information, what else do you say has to happen before, in your opinion, the information is sufficiently validated that it

can be used?

A. I think it would depend on the case. I think I'd want to, in the first instance, use all the data that was available to me.

Q. Well, I'm giving you a hypothetical. And the hypothetical is Dr. Freer creates a whole series of clusters. All the clusters get posted in the E-portal. Every cluster owner has the opportunity to come in and say, I validate this or you've missed something or I want to add something. And you said that's one thing that would help validate it?

A. That can mean if you went and asked every net winner to come in and demonstrate and show, as we've seen, where there are under- and overaggregations in Dr. Freer's analysis, if you got Oscar Sosa to come in and say, yeah, gee, these two clusters are mine and, in fact, that does lower my net equity by 80 percent, yeah, that would certainly be something that would be helpful to consider.

Q. Well, I want to go past helpful to consider --

A. No -- I'm sorry.

Q. -- and I want to know are you saying that's insufficient, and if so, why is that insufficient?

MR. RONA: Objection.

A. I think they're somewhat circular, that if a net winner comes in and says these are my accounts, I agree that that would be sufficient for that particular participant.

Q. Okay. Now, suppose the net winner comes in and says you underaggregated. I have some other accounts I want to put in?

A. Yes.

Q. Okay? Is that sufficient?

MR. RONA: Objection.

A. Potentially, I think you'd want to go and look at it, but I think that would certainly be -- certainly something you want to consider if that data was available.

Q. Okay. And when you say you'd want to consider, who wants to consider? The trustee? The court? You?

MR. RONA: Objection.

A. I would say all of the above. Whenever we do these types of analysis, you

173

want to consider all the information that's available to us, and if I have more information that was available to me, I'd certainly want to consider it.

Q. That's what I want to try to track down here. When you say all of the information available to us, what are you talking about exactly?

A. I think we've gone through some of these examples. We talked about information that may have been available through potential conversations, through EPOC, through the discovery process. Those were all things, I think, would be helpful to have considered. Now, whether they're sufficient, that's a different question.

Q. Well, in what way is that a different question?

A. Because you don't know if data is sufficient until you've analyzed it.

Q. Well, at what point in time do you, Joshua Dennis, agree that the data is sufficient? What has to happen?

MR. RONA: Objection.

174

A. I haven't -- that's not an analysis I've done. My goal was to look at Dr. Freer's analysis and determine whether it was a process that was reasonably applied based on the case facts and whether the data that was used was reliable and accurate. And looking at his analysis, I found that the methodology was not accurate, and that the accuracy of the data was not tested.

Q. So, as you sit here today, you do not have an opinion as to what is a sufficient quantum of proof?

MR. RONA: Objection.

A. That is not something I was asked to analyze.

Q. But, beyond that, you don't have an opinion on that subject, as you sit here today; correct?

MR. RONA: Same objection.

A. I do not have an opinion -- excuse me -- a position or opinion on the quantum of proof. Again, I -- let me back up. I'm a little unclear about how you're defining "quantum of proof."

175

Q. It's whatever is sufficient for you to sign off and say, yes, that is sufficiently accurate. I accept that cluster.

MR. RONA: Objection.

A. I haven't thought about all the possibilities of the universe of what someone could come out at to try to think about what that would look like. What I've done is look at Dr. Freer's analysis and no clear issues in aggregations, issues with methodology, and issues with underlying data.

Q. Okay. Could you turn to page -- excuse me -- to page 51 of your report?

A. Certainly.

Q. Actually, let me start with you turning to page 41. I tried to open Cluster 119427, which you show here, and my question is simple. Is this the -- is this -- what you show on page 41, is this the entire set of entries for Cluster 119427?

A. It's a summary of it.

Q. And when you say "it's a summary of it," what do you mean?

A. It is a pivot table in --

176

Q. Oh --

A. -- Excel perspective, so it's taking the data and aggregating it on certain common factors and then in summing it up, if you will, in terms of the net equity, and then on the far right side, the number of rep_id's, so user accounts that have that common, in this case, name, e-mail, address, city, and country, and phone number.

Q. So is this Dr. Freer's cluster, 119427, or is this your pivot table based on that cluster?

MR. RONA: Objection.

A. I'm not clear on the distinction. It is a -- it is a summary of Dr. Clear's -- excuse me -- Dr. Freer's Cluster ID 119427.

Q. Which would have 140 user accounts associated with it?

A. That's correct.

Q. And then on that side, you see -- on the far right-hand column, those numbers would be the actual account numbers for that individual showing under rep_nome?

A. I'm sorry. Can you ask that

177

question again?
Q. Sure. So if you look on the far right-hand column, the count of rep_id?
A. Yes.
Q. Rep, underscore, ID. The first one for Alex Lopes -- Lopes, account one, that means there's one account associated with that name?
MR. RONA: Objection.
A. That's true, in that name, but also in combination, that name, e-mail address, rep_end, R-E-P, E-N-D, which is the address, the city, the country, and that rep phone number. It's a combination of those factors.
Q. Okay. So if they match all of the columns, then you get a number associated with it?
A. Correct. It's a number of user accounts and the sum of -- the net equity, according to Mr. Martin, associated with those user accounts that have that combination of criteria.
Q. Okay. And then, for example, when you go down to Elaine Carreiro,

178

C-A-R-R-E-I-R-O, you have 53 individual accounts that match for the phone number, the country, the city, and the address, Maplewood Ave., plus Elaine at the e-mail address?
A. The e-mail addressee elainevit40@gmail.com.
Q. Correct. Okay.
A. Yes. It's that combination of -- the unique combination of each of those different variables.
Q. Okay.
A. The underlying data is on a line-by-line basis. It's shown as a footnote and demonstrates in Exhibit 13 to my report.
MR. LYNE: Can I see Exhibit 13? Mark that as the next exhibit.
(Deposition Exhibit No. 10, Document Entitled Exhibit 13, marked for identification)
Q. Okay. Showing you Exhibit 13. Is this the exploded table showing all of the entries?
A. I'll take your characterization of

179

"exploded." That's accurate.
Q. Yeah. Okay.
A. This is the underlying data.
Q. Okay. And if I go to the third page, there is a ratio calculated for each of the names and an aggregate ratio calculation. Do you see that on the --
A. Yes. On the right side, that is a -- that is the table that also appears in my report.
Q. Okay. And that appears on page 43?
A. Yes, sir.
Q. Okay. And that is the calculation you performed or somebody at your instruction?
A. I performed it, yes.
Q. You performed it. And where it refers to ratio and W ratio, what are you referring to?
A. So those are similarity scores, fuzzy matching scores using those two -- those two functions.
Q. Okay. And you were using the same methodology that Dr. Freer used for fuzzy matching?

180

A. Correct.
Q. Okay. Okay. If you could turn to page 51, paragraph 159. You are comparing two clusters, one for Shal, S-H-A-L, Group, Cluster 12650741, with Dolarex, D-O-L-A-R-E-X, Cluster 10992563; is that correct?
A. Yes.
Q. Okay. And are you suggesting that those two clusters should be aggregated?
A. Potentially.
Q. Well, did you form an opinion as to whether they, in fact, should be aggregated?
MR. RONA: Objection.
A. No. The purpose was to demonstrate the commonality and similarities between them to show that there is a great deal of uncertainty about whether they should be or not that no one has resolved.
Q. Shal Group is a corporation; is that right?
MR. RONA: Objection.
A. I don't know that to be a fact one way or the other.
Q. And the same is true for Dolarex?

181

You don't know whether that's a corporation either?
MR. RONA: Objection.
A. I don't know if it's a corporation one way or the other that exists out there or is the same corporation that is listed here.
Q. Okay. Well, assume, for purposes of my question, that those are two organized and existing corporations. Is it your position that two corporations should be joined in a single cluster under any circumstances?
MR. RONA: Objection.
A. I don't have a legal opinion one way or another. I don't know for a fact to what extent someone types in a given name, just like Mickey Mouse or anything else, if that has legal import or reflects liability or benefit of a corporation.
Q. Well, I'm asking you a general question, which is, in your opinion, in an appropriate aggregation methodology, is there any circumstance in which two corporations with their own clusters should be joined together as a single cluster?

182

MR. RONA: Objection.
A. I think that's a legal determination.
Q. I'm asking you as a data science question.
A. I don't think --
MR. RONA: Objection.
A. -- whether or not two corporations should be joined as a data science question is, in fact, a data science question. I think that's a legal question. What it looks like is that its potential likelihood is the same economic actor, right, the same beneficiary based on the commonalities in the data and the way they're structured may have received the benefits of both of those clusters.
Q. Did you take any steps to determine whether or not those are two separately organized corporations?
A. No, I did not.
Q. I think you say on 159 -- yeah, you say in the second sentence of paragraph 159, "In fact, this linkage," that's Shal Group and Dolarex, "was confirmed by the claims

183

submitted through the ePOC portal, as I can see that a single user." Can you tell me what you're referring to there?
A. Sure. So this goes back to what we were talking about a little bit earlier. Another example where I looked at the EPOC portal as potentially indicia of over- and underaggregation. In this particular instance, we have one individual, Saryas Jaff, submitted two claims, one for the Shal Group and one for Dolarex listed -- with each claim ID listed therein.
Q. So when you say this linkage was confirmed, it was confirmed in your mind because one person submitted claims for each of the two companies in the EPOC portal?
MR. RONA: Objection.
A. The purpose was simply to show, in this case, that there is a linkage to the extent that the same economic actor, in this case, Saryas Jaff appeared to own the user accounts for both clusters.
Q. So Dolarex was a net winner, according to you, in the amount of $189,701.

184

Do you see that?
A. I do.
Q. Is it your understanding that the net winners were using the E-portal?
MR. RONA: Objection.
A. Again, I think the determination of who was a net winner and a net loser is still very much a determination that's not been concluded. Under Mr. Martin's aggregation, I can tell you that both net winners and net losers certainly did attempt to go into the portal and identify accounts.
Q. Did you go into the EPOC portal and link other accounts besides this account -- these two accounts between these two companies?
MR. RONA: Objection.
A. I'm sorry. I don't -- I'm not sure I understand the question.
Q. Sure. You went into the E-portal -- the EPOC portal, and you affirmatively, you say, linked these two accounts in some fashion because a single user, Saryas Jaff, S-A-R-Y-A-S, J-A-F-F, submitted two claims;

185

correct?
A. That is what I say here, yes.
Q. Okay. And you did that; right?
A. One clarification. Are we talking about the EPOC portal? There is a portal, I understand, exists where people can go and look at their claims and submit information. I did not have direct access to the portal itself. What I received was a point-in-time extract of the data -- underlying data that was contained as a result of that portal at a point in time.
Q. All right. But are you saying that these two accounts between these two companies should affirmatively be aggregated or not --
MR. RONA: Objection.
Q. -- as a result of the fact that Saryas Jaff issued or submitted claims for both companies through the EPOC?
MR. RONA: Objection.
A. I don't have an opinion one way or the other --
Q. Okay. That's fine.
A. -- that's an economic actor.

186

Q. You don't know if it was a single economic actor?
A. Well, it was the same person that submitted -- that claimed both clusters.
Q. Well, that's not the same. The economic actors are Shal Group and Dolarex; correct?
MR. RONA: Objection.
A. I don't know that's correct.
Q. Well, what steps did you take to determine who the actual economic actor was?
A. It depends how you define "economic actor." Here, I'm saying that they're both tied together by a single individual.
Q. Well, a single individual submitted claims for two companies. I'm trying to determine if you say that's an underaggregation in Dr. Freer's methodology?
A. I think that's a legal determination. My point is simply that there is commonalities between the two, and at some point in time, I'm not sure if there's legal significance. Someone will have to determine whether what I typed into one of many fields

187

in a database when I signed up a user account has legal significance.
Q. Well, I'm trying to bring it back to an aggregation methodology. And Dr. Freer -- his aggregation methodology did not aggregate those two companies' accounts. He has separate clusters for each of them identified in paragraph 159.
Are you saying that Dr. Freer's aggregation methodology was flawed because it didn't aggregate the Shal Group account with the Dolarex account, just as a matter of methodology?
MR. RONA: Objection.
A. The underlie of your question is the assumption that because someone put Dolarex in the name field, that the extent that is a valid corporation somewhere in the world, that that -- somehow the liability or benefit goes to them. I don't know that that's the case or there's legal import.
Q. Well, you're speculating, aren't you, that maybe it is and maybe it isn't?
MR. RONA: Objection.

188

Q. You have no basis to opine one way or the other whether it should be aggregated?
MR. RONA: Objection.
A. That would be a legal determination.
Q. Okay. Was DL1 Inc. versus Linda Hackett, where do you refer to DL1 Inc.?
A. That should be right --
Q. Paragraph 156.
A. Yeah. I was going to say it should be in a similar -- similar area.
Q. Okay. Is that right? No. That's not right.
A. No. I don't think it is. I certainly have an exhibit that speaks to DL1 and Linda Hackett.
Q. Well, let's see if we can find it in your report.
A. I'm not sure I spoke to that one specifically.
Q. I thought you had.
A. It's certainly one I reviewed in detail. I don't know if I used that particular example.
Q. Well, it's Exhibit 9.5.

A. Yes.
Q. So if we pull that out. Oh, you found it?
A. Oh, wonderful.
Q. Okay. Just turn to page 50. Hold on one second. The last sentence of paragraph 156, "This evidence, which is summarized in Exhibit 6 and detailed in Exhibits 9.1 to 9.5 includes instances where the same or similar values for Potential Identifiers." Do you see that?
A. I do. And the sentence continues, but, yes.
Q. And it continues, but if we turn to Exhibit 9.5.
MR. LYNE: (To Attorney Papas). Do you have that?
(Deposition Exhibit No. 11, Document Entitled Exhibit 9.5, marked for identification)
Q. Do you recognize Exhibit 11 as your Exhibit 9.5 in your report?
A. I do.
Q. And it includes two clusters, 118449

and 707775, which --
A. I think -- I apologize. I think you may have misread the cluster ID for the first one.
Q. My eyes are bad. Let me try this again. I'm referring to Exhibit 9.5. It appears to contain two clusters, Cluster 11849449 and Cluster 70775. The first cluster is a rep_nome DL1 Inc., the company; correct?
MR. RONA: Objection.
A. That's what is shown here.
Q. That's what's shown?
A. Again, just to clarify, this particular cluster contains 928 user accounts. I'm showing a selection of them here.
Q. Okay. And you are suggesting that it should be joined with Linda Hackett --
MR. RONA: Objection.
Q. -- is that right?
A. I'm not suggesting one way or the other. My point is there are significant overlap and commonality between the potential identifiers. I actually think the trustee is the one who suggested and actually explicitly

stated that these two should be joined. The trustee stated as part of the settlement agreement that -- and I can read it exactly if we had the document -- that the Hacketts were net winners of more than $580,000. $580,000 is the aggregate total between the net winnings of 1.2 million for the Linda Hackett accounts listed here and less the 622,000 from DL1. That was the trustee's determination that these two accounts should be aggregated for purposes of settlement.
Q. But I'm asking you questions with respect to Dr. Freer's aggregation methodology.
Are you suggesting that Dr. Freer's aggregation methodology is invalid because, as a methodology, it did not aggregate both of these accounts, regardless of what the trustee ultimately determined?
MR. RONA: Objection.
A. I think it speaks to one of the issues with the entire methodology that Dr. Freer employs. His entire methodology is based on an unfounded, unsupported

understanding given to him by the trustee that people always used their own name and the contact information of other people, right, and that's something that is not in the record. That's something that's provided by the trustee. Dr. Freer pointed to my report, the judge's court order eliminating Mr. Martin, in Mr. Martin's report, as other potential sources for that understanding, none of which, other than Mr. Martin, who's now been excluded, actually stand for that proposition. So, yes, I do think that a methodology that relies on what someone happened to type into the name field as a basis for aggregating a single participant, a simple beneficiary or loser or economic actor, is not necessarily a reliable approach.
Q. So what you're doing essentially is collapsing Dr. Freer's aggregation methodology with any subsequent validation or analysis and that Dr. Freer's methodology is somehow inaccurate or incomplete, and, therefore, objectionable because it doesn't go through the secondary step of aggregating based on

193

external data?
MR. RONA: Objection.
A. No. That's not. I think what I just described was the very foundation of his methodology was flawed. It's based entirely on an assumption that name has a significance or relevance or accuracy that is unproven. His entire methodology requires two names to be fairly similar or substantially similar, and that core underlying assumption, the same one that Mr. Martin made and the Court excluded him for that very reason, here, Dr. Freer has done the exact the same thing.
Q. So it's your understanding that Dr. Freer's report is primarily based on name?
A. His methodology very much relies very heavily on the requirement that's baked into the coding, that's baked into the process. He states explicitly multiple times in his report in order for two user accounts to be aggregated, they had to have at least fairly similar name or have substantial agreement on name as well as other common identifiers. He said that very explicitly.

194

Yes.
Q. Is it your understanding that Dr. Freer's FESM [sic] methodology relies primarily on name?
MR. RONA: Objection.
A. I think name is one of the core -- and when I say relies on name, his methodology is implicit, the entire methodology, including the FS-EM portion is the assumption that participants consistently entered the name accurately.
Q. Where in the FS-EM methodology do you understand Dr. Freer gave primacy to name?
A. Sure. I'm happy to go through his report and show you.
Q. Sure.
A. So I would point you to the section marked "Special Role of the Name Field in the TelexFree Dataset."
Q. Okay. But I'm asking you a specific question, which is how in the FESM [sic] methodology, where is name given primacy?
MR. RONA: Objection.
A. Sure. Of course. So I think the

195

first place it's given primacy is part of the blocking. I'm happy to go to that and discuss it.
Q. Yup. Okay. Do you want to look at the ten steps? Would that help you?
A. Oh, no. I have his -- I have his step four right here in his report. That's helpful.
Q. Okay. Okay. So blocking?
A. That's correct. And I'd like to just --
Q. Take your time.
A. Thank you. Here we are. Paragraph 148 on page 51.
Q. Okay.
A. Paragraph 148 on page 51 reads, "Because of the special role that the name field plays in the TelexFree dataset," and it refers to the prior section, Section III.B.1, "it is important that no pairs of" accounts -- excuse me -- "account groupings that substantially disagree on the name end up matched together. I use blocking as a mechanism to accomplish this goal, by ensuring

196

that the blocking set contains only pairs of account groupings with substantial agreement on name. In this way, only such pairs, i.e., those with substantial agreement on name, are even considered for potential linkage."
Q. Okay. That's step four, blocking. Is there anywhere in steps five through nine where name is given greater weight in calculating match probabilities?
A. I believe -- I believe there is. And, again, you know, this step four is not a one in ten thing. This, by definition, the goal of this singular step is to limit the linkage to those which have names, so I think this is --
Q. Which have names that are similar --
A. Thank you.
Q. -- or the same?
A. I misspoke. Thank you. Names that are similar or the same. I agree.
Q. Okay.
A. I believe there's also relevance, obviously, with every subsequent step which relies on blocking, but I'll also note that

the step five calculates gamma values between zero -- zero and five.

Q. What's the gamma value for name?

A. It looks like thresholds. Where, for example, if the nome ratio, which is his fuzzy matching ratio is, for example, between 70, 80, it gives it a value of one. If it's 100, it gives it a value of five.

Q. Okay.

A. In terms of subsequent steps, I'm trying to -- so I guess I'd say that every subsequent step is utilizing information in the output of the blocking step, so to the extent it's implicit in the blocking step, it's going to be the same issue throughout.

Q. Are you familiar with any blocking methodology, would you say, that would be more appropriate than using the same or similar name that would give, in your opinion, more valid results?

A. Again --

MR. RONA: Objection.

A. -- I haven't determined what the right or wrong blocking approaches are. Well,

that's not entirely true. I disagree with the blocking approaches taken. I have not been asked, nor have I attempted to determine how blocking should have been done or even if doing this approach all together that includes blocking is the best and most reliable approach.

Q. You have no opinion on that?

A. That's correct. I have not been asked to do an affirmative model in this case. Simply review the reliability of Dr. Freer's model, which I do not find reliable.

Q. So --

THE WITNESS: And, I apologize, my mouth is starting to get dry. Just give me even five minutes to catch my breath.

MR. RONA: We've been going for over an hour.

MR. LYNE: Sure. Five minutes?

(Recess was taken)

Q. Sir, is it your understanding that if accounts are not joined in the blocking

step, that those accounts can never be joined in later steps?

A. No. It's not my understanding.

Q. So what is your understanding on that?

A. So as part of that step, that's where they create the pairwise functions, and they only do that if the name, based on his definition, is substantially -- in substantial agreement or don't substantially disagree, however you phrase it. They have -- accounts have the ability to come back together as part of the chaining or clustering that happens down the line.

Q. Transitive closure?

A. Whatever terminology you want. I think he describes the step as clustering, if we can pull out his exhibit. I think it's called a dynamic clustering, but things do come back together as part of that process, which is where you lead to some of those issues, quite frankly, with transitive closing where you have chaining, where A may be similar to B and B is similar to C and C is

similar to D, but you look at the outer edges, and A and D have nothing in common, so there's a potential for bringing things together. There is also potential for that creating -- it does create additional error.

Q. So let's take it in little steps. So regardless of the similarity of name, later steps can bring pairwise combinations that were not joined in the blocking phase together in the later phases, in the dynamic clustering phase; correct?

MR. RONA: Objection.

A. There is still an element of name that is considered as part of that process, right, 'cause it looks at -- for every two pairwise matches made through the blocking, it goes to compare those -- the comparison factors across each of those to other pairwise matches and bring them together. So there's very much an element in similarity and comparison of name that goes on in that, and you can see that in some of the examples we looked at before with the Carreiro example where there's a similarity in name across

certain two pairs within there, but in the aggregate, when you compare the beginning -- I'll call it the beginning of the chain to the end of chain, however you want to order it, there's substantial disagreement.

Q. So have you -- strike that.

You're describing a chaining issue that A and D may be radically different, but B and C join them together in the dynamic clustering process?

A. That is one issue that we saw, and I believe that's due to the chaining, yes.

Q. All right. And isn't overaggregation and underaggregation -- you're calling it a byproduct of the dynamic clustering process, the chaining problem?

A. I don't think that's the only source of under- and overaggregation. That is a potential source of under- and overaggregation.

Q. Have you seen any published studies that suggest, in your mind, that an appropriate clustering process, dynamic or otherwise, should result in zero

overaggregation or underaggregation?

MR. RONA: Objection.

A. Can you ask that question again?

Q. Sure. Are you aware of any studies -- published studies or research papers that suggest, in your mind, that a properly designed aggregation methodology should result in zero overaggregation and zero underaggregation?

MR. RONA: Objection.

A. I didn't endeavor to answer that question through looking at academic studies.

Q. Well, in the studies you do cite in your materials considered paragraph -- excuse me -- Exhibit 8, you cite nine other -- "Other Materials," some of which appear to be published studies. Do you see those?

A. I do. Those are studies that I believe were cited by either Dr. Freer or, I think, more likely in this case, Mr. Martin.

Q. Okay.

A. With the exception of the Peter Christen book. That was simply cited by Dr. Freer.

Q. Okay. In any of these studies which you considered, including the Christen studies, do any of those studies, including the Christen book or otherwise, suggest in your mind that an over -- excuse me -- that a properly designed aggregation methodology should inevitably result in zero overaggregation and zero underaggregation errors?

MR. RONA: Objection.

A. I'm not aware that it addressed that question either way in terms of what a perfectly designed study should or should not conclude.

Q. Let's not limit it to a perfectly designed study. I'm saying an acceptable aggregation methodology.

A. I think what's acceptable --

MR. RONA: Objection.

A. -- is a fact-specific determination, case-specific determination.

Q. Well, in all of your review of either Christen's book or any of the other materials you cite in Exhibit 8, do any of

those materials state, in your mind, that an acceptable aggregation methodology should inevitably result in zero aggregation -- overaggregation errors and zero underaggregation errors?

MR. RONA: Objection.

A. I don't think it says it one way or another. My recollection of things that were said in those kinds of text or studies was that's important to consider context in which the analysis and methodology is being performed, and that certain areas, such as healthcare or litigation may have different tolerances and different acceptances for false positives and false negatives, and that should be something that is considered when performing an analysis.

Q. Well, that consideration is done after the analysis is complete and the output is done; correct?

A. No.

MR. RONA: Objection.

A. Not necessarily.

Q. Well, do you have any support in any

materials which you have ever reviewed in the field of data science that stand, in your mind, for the proposition that an acceptable data aggregation methodology must inevitably result in zero errors in either overaggregation or underaggregation?

MR. RONA: Objection.

A. I don't know that I've seen that statement written in that way in any academic studies.

Q. Do any of those academic studies talk about the tension between overaggregation and underaggregation?

A. I believe some of the studies described what could cause over- and underaggregation.

Q. Well, what causes overaggregation and underaggregation are by the joining rules; right? If they're liberal, you overaggregate. If they're overly strict, you underaggregate? That's a general principle; right?

MR. RONA: Objection.

A. I think it would depend on the data and what you'd -- what your purpose is and

what your methodology is.

Q. The purpose of data aggregation is to aggregate data. I mean, that's what -- regardless of, you know, end use, the methodology is intended to create data clusters out of a large dataset. That's what data aggregation does; right?

MR. RONA: Objection.

A. It depends. It doesn't need to be what data aggregation does.

MR. LYNE: Okay. All right. I'm good.

THE WITNESS: Thank you.

(Whereupon, the deposition concluded at 3:06 p.m.)

COMMONWEALTH OF MASSACHUSETTS
MIDDLESEX, SS.

I, Jacqueline Curran, Registered Merit Reporter and Notary Public in and for the Commonwealth of Massachusetts, do hereby certify that pursuant to appropriate notice of taking deposition, there came before me the following-named person, to wit: JOSHUA W. DENNIS, who was by me duly sworn; that he was thereupon examined upon his oath and his examination reduced to writing by me; and that the deposition is a true record of the testimony given by the witness.

I further certify that I am not a relative or employee or counsel or attorney for any of the parties, or a relative or employee of such counsel or attorney, nor am I financially or otherwise interested in the outcome of the action.

IN WITNESS WHEREOF, I have hereunto set my hand and seal this 10th day of May, 2023.

My Commission Expires
August 18, 2028

Jacqueline M. Curran
Notary Public

The foregoing certification of this transcript does not apply to any reproduction of the same in any respect unless under the direct control and/or direction of the certifying reporter.

Excerpt from Rule 30(e):
Submission to Witness; Changes; Signing.

When the testimony is fully transcribed, the deposition shall be submitted to the witness for examination and shall be read to or by him/her, unless such examination and reading are waived by the witness and by the parties. Any changes in form or substance which the witness desires to make shall be entered upon the deposition by the officer with a statement of the reasons given by the witness for making them.

****

I, JOSHUA W. DENNIS, have examined the above transcript of my testimony and it is true and correct to the best of my knowledge, information and belief. Any corrections are noted on the attached errata sheet.

Signed under the pains and penalties of perjury this day of, 2023.

_____
Deponent's Signature

ERRATA SHEET

NAME:_____ DATE:_____

CASE NAME:_____

DATE OF DEPOSITION:_____

PAGE LINE CORRECTION

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

CURRAN COURT REPORTING

#

#134293 [1] - 2:7

$

$189,701 [1] - 183:24
$580,000 [2] - 191:5
$800,000 [1] - 146:23

'

'19 [2] - 107:22, 109:8
'20 [2] - 107:22, 109:8
'22 [1] - 14:23

0

0.980 [2] - 162:1, 162:16
0.988 [2] - 162:2, 162:17
0.996 [2] - 162:1, 162:16
003 [1] - 81:17
02109 [2] - 3:7, 3:15
02180 [1] - 2:22

1

1 [6] - 1:1, 1:1, 1:2, 4:9, 5:1, 8:2
1,566,383 [1] - 144:14
1.2 [1] - 191:7
1.5 [3] - 137:1, 138:7, 138:8
10 [2] - 4:21, 178:18
100 [1] - 197:8
105 [1] - 81:22
10992563 [1] - 180:6
10:00 [1] - 2:13
10th [1] - 207:13
11 [15] - 1:2, 1:4, 4:22, 43:22, 44:2, 47:13, 47:14, 48:3, 49:9, 49:24, 68:18, 68:20, 69:5, 189:18, 189:21
118449 [1] - 189:24
11849449 [1] - 190:8
119427 [4] - 175:17, 175:20, 176:11, 176:16
12650741 [1] - 180:5
12:30 [2] - 131:15, 131:22
13 [5] - 4:21, 178:14, 178:16, 178:19, 178:21
131 [1] - 4:18
132 [1] - 4:19
132785 [3] - 94:21, 95:14, 158:22
14-40987-EDK [1] - 1:5
14-40988-EDK [1] - 1:5
14-40989-EDK [1] - 1:6
140 [1] - 176:17
141 [1] - 86:8
148 [2] - 195:14, 195:16
15 [5] - 4:12, 38:17, 43:17, 136:24, 154:22
156 [2] - 188:8, 189:7
158 [1] - 4:20
159 [4] - 180:3, 182:21, 182:22, 187:8
16 [4] - 4:20, 88:3, 88:4, 158:16
16-4006 [2] - 1:7, 3:11
16-4007 [2] - 1:8, 3:12
17 [5] - 43:16, 68:17, 68:20, 69:5, 154:22
17,000 [1] - 154:21
17-plus [1] - 144:13
178 [1] - 4:21
18 [1] - 207:17
189 [1] - 4:22
1:26 [1] - 131:23

2

2 [10] - 4:10, 4:19, 5:3, 14:15, 15:21, 32:19, 81:24, 82:1, 132:5, 150:5
20 [4] - 38:18, 76:2, 123:5, 145:5
2018 [2] - 107:21, 109:8
2020 [5] - 9:6, 9:11, 9:19, 12:17, 12:20
2022 [4] - 10:15, 14:1, 14:2, 14:20
2023 [5] - 2:13, 14:21, 14:22, 207:13, 208:15
2028 [1] - 207:17
209 [1] - 1:1
21 [1] - 2:22
279-8400 [1] - 2:23
28 [3] - 2:11, 3:6, 3:14

3

3 [13] - 4:12, 15:6, 15:10, 15:14, 15:17, 20:1, 20:3, 20:4, 20:21, 94:19, 157:22, 157:23, 157:24
30 [1] - 7:4
30(e [1] - 208:1
300 [1] - 82:3
31 [1] - 81:20
3101 [1] - 3:6
31st [1] - 2:12
33 [1] - 4:13
34 [1] - 81:23
37 [1] - 83:20
395-9570 [1] - 3:15
3:06 [1] - 206:15

4

4 [3] - 4:13, 33:1, 33:5
41 [2] - 175:16, 175:19
423-0400 [1] - 3:7
43 [4] - 4:14, 67:14, 67:20, 179:11
430-ish [1] - 87:9
430-line [1] - 18:4
45-minute [1] - 131:16
466 [2] - 87:12, 87:13
47 [2] - 85:16, 85:18
471 [3] - 95:13, 158:21, 159:18

5

5 [8] - 4:6, 4:9, 4:10, 4:14, 67:14, 67:19, 154:9, 156:2
50 [2] - 147:5, 189:5
500 [1] - 56:2
51 [4] - 175:13, 180:3, 195:14, 195:16
53 [1] - 178:1
55 [4] - 110:10, 110:19, 110:21, 110:22

6

6 [6] - 4:16, 94:10, 94:14, 158:2, 158:3, 189:8
617 [2] - 3:7, 3:15
622,000 [1] - 191:8
67 [1] - 4:14

7

7 [3] - 4:18, 131:24, 144:11
70 [1] - 197:7
70775 [1] - 190:8
707775 [1] - 190:1
781 [1] - 2:23

8

8 [8] - 4:19, 132:4, 149:21, 149:22, 149:23, 202:15, 203:24
80 [2] - 171:20, 197:7
802 [1] - 3:14

9

9 [4] - 2:13, 4:20, 158:15, 159:11
9.1 [1] - 189:8
9.5 [7] - 4:22, 188:24, 189:8, 189:15, 189:19, 189:22, 190:6
928 [1] - 190:14
94 [1] - 4:16
98 [1] - 103:19
99 [1] - 168:11

A

A-R-G-U-E-T-A [1] - 109:22
a.m [1] - 2:13
ability [1] - 199:12
able [2] - 137:22, 163:6
absence [1] - 104:7
absent [2] - 14:20, 101:21
absolutely [3] - 122:9, 125:15, 140:1
abstract [3] - 27:10, 55:18, 89:20
academic [3] - 202:12, 205:9, 205:11
accept [1] - 175:3
acceptable [13] - 100:8, 103:10, 103:24, 125:21, 160:1, 160:5, 160:6, 160:23, 168:18, 203:16, 203:18, 204:2, 205:3
acceptances [1] - 204:14
accepted [2] - 103:11, 108:15
access [1] - 185:8
Access [1] - 35:13
accomplish [1] - 195:24
according [2] - 177:20, 183:24
account [27] - 29:6, 31:6, 31:7, 50:3, 93:3, 104:10, 108:14, 113:3, 119:23, 120:13, 121:21, 124:21, 143:18, 152:14, 153:3, 153:5, 159:1, 159:19, 176:22, 177:6, 177:7, 184:14, 187:1, 187:11, 187:12, 195:21, 196:2
accounted [1] - 31:8
accounting [1] - 34:1
accounts [111] - 18:8, 19:20, 26:21, 43:13, 43:20, 43:21, 43:22, 44:2, 47:8, 47:14, 47:17, 48:4, 48:7, 48:18, 48:19, 49:9, 49:10, 49:11, 49:13, 49:24, 50:1, 50:2, 68:23, 69:5, 69:6, 72:10, 81:6, 81:10, 87:5, 87:9, 87:12, 87:13, 91:13, 92:23, 93:3, 95:3, 95:7, 95:13, 95:20, 95:21, 95:22, 95:24, 96:3, 96:13, 96:16, 99:10, 100:16, 104:18, 110:8, 110:15, 110:19, 110:21, 110:22, 111:13, 112:1, 112:6, 112:14, 118:13, 121:10, 121:16, 128:24, 129:12, 129:13, 136:24, 141:14, 142:6, 142:12, 142:17, 144:3, 144:7, 144:9, 144:13, 148:9, 152:1, 152:3, 152:11, 153:11, 153:15, 154:20, 154:22, 157:15, 157:16, 158:24, 159:7, 160:11, 170:10, 170:19, 172:7, 172:11, 176:7, 176:17, 177:19, 177:21, 178:2, 183:22, 184:12, 184:14, 184:15, 184:22, 185:14, 187:6, 190:14, 191:8, 191:10, 191:18,

193:20, 195:20, 198:24, 199:1, 199:11
accuracy [20] - 101:19, 103:19, 104:11, 126:10, 126:11, 126:13, 134:13, 134:23, 136:5, 145:15, 145:24, 147:9, 147:13, 148:21, 151:18, 168:12, 169:2, 170:2, 174:8, 193:7
accurate [22] - 25:21, 50:12, 105:20, 128:14, 132:12, 134:9, 134:21, 134:22, 136:6, 137:3, 139:19, 144:23, 152:17, 153:15, 157:14, 162:21, 162:23, 166:9, 174:6, 174:8, 175:3, 179:1
accurately [1] - 194:11
acknowledged [3] - 130:3, 130:9, 131:1
action [4] - 3:11, 3:12, 42:9, 207:11
actor [10] - 99:12, 104:17, 121:18, 182:13, 183:20, 185:24, 186:2, 186:11, 186:13, 192:16
actors [1] - 186:6
actual [7] - 25:4, 97:8, 99:14, 141:16, 141:19, 176:22, 186:11
ad [4] - 75:8, 75:10, 100:4, 100:5
add [3] - 19:7, 44:23, 171:11
added [1] - 17:5
addition [1] - 5:24
additional [2] - 17:4, 200:5
address [10] - 30:4, 87:5, 153:4, 153:5, 154:8, 176:8, 177:11, 177:12, 178:3, 178:4
addressed [2] - 132:17, 203:11
addressee [1] - 178:5
addresses [5] - 51:4, 51:8, 52:9, 95:8, 152:12
adequacy [1] - 83:22
adequate [2] - 82:6, 82:13
Administered [1] - 1:7
Adversary [2] - 1:7, 1:8
Affidavit [1] - 151:3
affidavit [3] - 106:17, 152:10, 152:11
affidavits [3] - 106:9, 151:1, 151:20
affirmatively [4] - 97:13, 105:16, 184:21, 185:15
aggregate [19] - 38:3, 47:8, 54:17, 55:9, 66:5, 79:15, 80:23, 81:3, 120:12, 123:14, 133:24, 137:18, 179:6, 187:5, 187:11, 191:6, 191:17, 201:2, 206:3
aggregated [15] - 18:9, 49:14, 49:19, 52:11, 151:23, 151:24, 152:14, 152:21, 152:22, 180:9, 180:12, 185:15, 188:2, 191:10, 193:21
aggregates [1] - 113:22
aggregating [4] - 35:5, 176:3, 192:15, 192:24
Aggregation [5] - 4:12, 4:15, 15:7, 20:10, 67:16
aggregation [61] - 35:1, 41:3, 46:24, 47:2, 53:23, 54:2, 54:6, 59:7, 63:2, 84:4, 84:8, 87:18, 87:19, 96:22, 97:5, 99:24, 100:21, 101:1, 107:17, 108:4, 108:7, 109:10, 109:14, 113:3, 113:4, 113:6, 113:15, 114:8, 115:10, 116:3, 119:4, 119:14, 130:4, 131:2, 131:8, 142:10, 161:14, 161:24, 162:14, 164:3, 165:6, 165:10, 167:8, 168:19, 181:21, 184:9, 187:4, 187:5, 187:10, 191:13, 191:16, 192:19, 202:7, 203:6, 203:17, 204:2, 204:3, 205:4, 206:2, 206:7, 206:10
aggregations [5] - 84:18, 84:22, 108:24, 141:17, 175:10
ago [4] - 9:1, 35:17, 37:6, 52:16
agree [20] - 49:6, 50:6, 51:20, 62:24, 68:4, 116:1, 116:5, 116:9, 117:17, 119:5, 120:2, 120:22, 125:15, 129:18, 144:23, 170:10, 170:12, 172:7, 173:22, 196:20
agreed [4] - 6:10, 6:11, 89:10, 108:15
agreement [6] - 65:2, 191:3, 193:23, 196:2, 196:4, 199:10
ahead [1] - 144:21
akin [3] - 42:24, 44:15, 53:20
Alex [1] - 177:6
Alexandra [1] - 3:6
algorithm [3] - 73:18, 73:20, 73:23
allow [1] - 6:16
alternative [5] - 80:5, 92:13, 117:5, 118:24, 119:9
alternatives [3] - 116:19, 116:20, 117:22
ambiguity [1] - 72:2
amount [4] - 30:11, 30:12, 76:9, 183:24
amounts [1] - 80:2
analyses [1] - 45:9
analysis [68] - 13:12, 31:11, 38:19, 38:20, 40:16, 41:8, 42:21, 45:22, 47:3, 47:6, 48:14, 48:16, 50:17, 56:13, 59:7, 74:3, 75:13, 82:17, 88:10, 98:4, 98:13, 100:6, 114:2, 114:15, 114:17, 115:13, 117:17, 117:19, 117:23, 118:1, 118:4, 118:6, 120:6, 120:8, 121:3, 125:16, 128:13, 128:14, 130:23, 135:6, 135:18, 135:24, 136:16, 137:5, 137:7, 138:9, 143:14, 146:13, 149:13, 149:18, 155:6, 155:7, 157:5, 157:12, 157:13, 161:19, 167:23, 168:7, 171:17, 172:24, 174:1, 174:3, 174:7, 175:9, 192:20, 204:11, 204:17, 204:19
analytical [13] - 8:8, 12:16, 18:24, 22:1, 32:9, 50:19, 51:22, 84:10, 86:9, 96:7, 108:20, 120:11, 122:17
analytics [2] - 38:24, 39:23
analyze [2] - 169:2, 174:15
analyzed [9] - 41:20, 43:3, 44:22, 52:4, 107:24, 112:16, 112:20, 116:18, 173:20
AND [1] - 1:22
answer [23] - 6:7, 6:13, 6:17, 6:19, 11:2, 20:5, 35:7, 80:5, 81:16, 97:14, 98:12, 100:12, 127:21, 132:10, 134:7, 136:17, 139:19, 139:20, 150:17, 162:20, 163:19, 202:11
answered [3] - 6:9, 120:17, 161:21
answering [2] - 49:4, 59:18
answers [2] - 132:9, 163:2
anti [4] - 50:24, 51:24, 53:18, 56:9
anti-price [4] - 50:24, 51:24, 53:18, 56:9
antitrust [1] - 56:10
anyway [1] - 161:23
apapas@murphyking.com [1] - 3:8
apologies [1] - 14:23
apologize [6] - 25:8, 76:23, 156:9, 165:8, 190:2, 198:15
apostrophe [1] - 80:10
apparent [6] - 135:2, 135:19, 136:9, 136:10, 136:14, 160:18
appear [11] - 25:19, 27:16, 29:8, 41:24, 87:6, 95:20, 98:22, 137:16, 144:17, 159:8, 202:16
appeared [9] - 29:15, 96:13, 110:5, 110:13, 136:19, 137:9, 137:15, 155:13, 183:21
appearing [3] - 8:2, 29:9, 93:3
applicability [1] - 130:8
application [1] - 130:6
applied [7] - 62:20, 65:19, 92:20, 92:22, 130:1, 166:7, 174:4
applies [1] - 91:17
apply [4] - 102:20, 124:11, 124:17, 207:21
applying [1] - 124:20
appreciate [1] - 168:23
approach [17] - 39:15, 57:15, 59:14, 59:16, 66:8, 74:16, 92:6, 100:2, 102:12, 104:8, 119:6, 119:11, 120:23, 147:4, 192:17, 198:5, 198:7
approaches [11] - 38:23, 53:6, 53:7, 53:9, 53:10, 92:8, 116:23, 118:16, 120:20, 197:24, 198:2
appropriate [17] - 66:3, 67:2, 72:24, 79:11, 79:17, 89:22, 90:2, 90:6, 124:11, 132:21, 136:1, 147:18, 149:17, 181:21, 197:18, 201:23, 207:4
appropriately [3] - 65:20, 77:24, 78:5
area [10] - 10:2, 12:6, 34:23, 38:20, 39:10, 39:17, 39:18, 40:18, 85:19, 188:10
areas [3] - 11:23,

33:24, 204:12
argue [2] - 60:17, 121:1
Argueta [9] - 109:22, 109:23, 110:1, 110:4, 111:8, 112:15, 114:10, 114:24, 115:21
articles [2] - 131:6, 131:7
ascribed [1] - 66:1
aside [4] - 38:8, 69:2, 73:13, 73:17
aspect [6] - 40:7, 41:5, 51:6, 60:6, 61:1, 61:6
aspects [3] - 21:9, 61:23
assess [2] - 8:16, 39:11
assessing [1] - 38:19
assessment [1] - 129:20
assign [2] - 50:1, 124:21
assigned [4] - 47:9, 47:18, 48:7, 50:3
assigning [2] - 118:13, 119:23
assignment [1] - 73:4
assisted [2] - 8:21, 11:15
assisting [1] - 9:21
associated [7] - 18:20, 22:24, 44:11, 176:18, 177:7, 177:16, 177:20
assume [7] - 5:17, 6:8, 136:5, 165:13, 166:22, 169:19, 181:7
assuming [3] - 134:20, 137:3, 150:18
assumption [8] - 71:2, 71:5, 71:6, 108:5, 187:16, 193:6, 193:10, 194:9
assumptions [6] - 21:19, 66:8, 69:18, 69:23, 70:1, 116:24
attach [1] - 18:13
attached [7] - 16:1, 17:24, 123:17, 123:18, 157:3, 157:21, 208:12
attaching [1] - 18:3
attempt [22] - 22:22, 23:3, 71:10, 90:5, 95:2, 95:12, 96:24, 103:17, 104:5, 111:11, 114:13, 115:3, 127:24, 133:23, 134:12, 134:13, 135:1, 135:15, 136:7, 158:22, 161:6, 184:11
attempted [13] - 16:8, 97:4, 105:2, 120:21, 122:4, 122:12, 128:21, 129:4, 129:10, 129:17, 161:1, 198:3
attempting [1] - 47:8
attention [7] - 8:1, 14:14, 20:3, 33:4, 67:18, 94:14, 169:13
attorney [2] - 207:9, 207:10
Attorney [6] - 4:24, 14:16, 15:3, 67:11, 88:5, 189:16
attributed [1] - 152:3
August [1] - 207:17
auto [1] - 54:23
available [13] - 16:18, 107:5, 112:2, 112:6, 148:22, 151:13, 158:10, 171:4, 172:18, 173:2, 173:3, 173:7, 173:11
Ave [1] - 178:4
avoid [1] - 31:11
aware [5] - 105:3, 105:5, 202:4, 203:11

B

background [3] - 10:9, 11:1, 12:11
bad [1] - 190:5
baked [2] - 193:17, 193:18
BALAN [1] - 1:16
Balan [8] - 3:11, 106:16, 106:19, 107:1, 107:4, 151:4
Balan's [1] - 152:10
balance [1] - 103:8
Balans [1] - 151:22
ballpark [1] - 138:17
bank [12] - 51:1, 56:2, 105:12, 111:21, 111:22, 112:1, 112:4, 126:14, 129:11, 129:12, 129:13, 145:17
bankruptcy [1] - 5:15
BANKRUPTCY [1] - 1:3
Bankruptcy [1] - 2:5
base [3] - 45:14, 45:15, 63:19
based [25] - 13:12, 16:17, 31:2, 53:7, 72:4, 73:12, 73:14, 76:5, 134:6, 143:14, 147:1, 147:3, 153:11, 153:12, 153:13, 162:24, 166:8, 174:4, 176:11, 182:14, 191:24, 192:24, 193:5, 193:15, 199:8
basic [2] - 45:22, 147:19
basis [8] - 39:1, 50:9, 71:6, 132:22, 168:6, 178:13, 188:1, 192:15
becoming [1] - 38:17
begin [1] - 140:19
beginning [2] - 201:2, 201:3
behalf [2] - 2:3, 8:18
behind [1] - 83:14
belief [1] - 208:12
belong [2] - 48:19, 96:4
below [2] - 143:15, 159:13
Ben [2] - 110:18, 110:24
beneficiary [2] - 182:13, 192:16
benefit [2] - 181:18, 187:19
benefits [1] - 182:16
benefitted [1] - 104:17
Benjamin [2] - 109:21, 110:16
best [6] - 16:11, 36:20, 131:21, 163:12, 198:6, 208:11
better [2] - 119:10, 139:20
between [21] - 24:6, 28:20, 34:13, 49:1, 64:15, 80:13, 102:15, 103:8, 114:16, 118:5, 144:16, 166:17, 180:15, 184:15, 185:14, 186:21, 190:22, 191:6, 197:1, 197:6, 205:12
beyond [9] - 44:19, 60:15, 65:17, 90:13, 92:8, 114:18, 114:22, 115:7, 174:16
big [5] - 39:9, 39:18, 40:12, 40:18, 58:11
biggest [1] - 100:13
bit [7] - 11:19, 46:15, 102:3, 148:20, 158:13, 183:5
bits [1] - 77:7
blank [3] - 25:12, 29:18, 70:19
blocking [39] - 21:9, 22:5, 23:8, 23:18, 23:19, 24:10, 24:13, 25:24, 26:4, 26:12, 31:21, 68:10, 74:9, 74:13, 74:14, 74:17, 74:18, 74:20, 75:7, 75:11, 75:20, 76:4, 195:2, 195:9, 195:23, 196:1, 196:6, 196:24, 197:13, 197:14, 197:16, 197:24, 198:2, 198:4, 198:6, 198:24, 200:9, 200:16
blocks [1] - 22:23
blue [1] - 33:15
board [1] - 161:2
body [3] - 81:18, 81:19, 158:14
bonus [2] - 143:24, 144:1
book [3] - 202:23, 203:4, 203:23
Borelian [3] - 4:12, 15:6, 20:10
Boston [3] - 2:12, 3:7, 3:15
Brazilian [1] - 43:20
breadth [1] - 144:6
break [4] - 76:16, 131:17, 132:8, 148:19
breath [1] - 198:17
bring [7] - 54:18, 92:23, 92:24, 169:13, 187:3, 200:8, 200:19
bringing [3] - 118:17, 134:6, 200:3
broader [2] - 34:23, 140:8
broadly [1] - 77:15
broken [1] - 144:16
brought [3] - 52:10, 146:4, 157:3
BS [1] - 33:17
bucket [4] - 77:3, 77:5, 91:11, 91:12
bucketed [1] - 77:2
bucketizing [1] - 77:7
bullet [4] - 56:8, 56:12, 143:7, 143:16
business [8] - 12:4, 33:18, 33:20, 33:24, 36:5, 36:8, 38:16, 56:17
Business [1] - 57:3
business's [1] - 40:3
businesses [1] - 51:4
byproduct [1] - 201:15

C

calculate [13] - 86:21, 89:15, 89:21, 90:5, 96:24, 97:8, 115:4, 129:4, 147:22, 148:4, 148:7, 162:24, 163:4
calculated [5] - 30:15, 80:3, 163:17, 163:22, 179:5
calculates [1] - 197:1
calculating [5] - 79:23, 88:18, 89:4, 90:24, 196:9
calculation [10] - 30:10, 98:7, 98:17, 99:2, 103:19, 154:14, 156:23, 157:3, 179:6, 179:13
calculations [3] - 8:17, 99:14, 145:1
Cameron [4] - 4:16, 94:11, 94:17, 139:5
carefully [2] - 109:7, 115:19
Carreiro [2] - 177:24, 200:23
CARREIRO [1] - 178:1
CASE [1] - 209:3
case [58] - 8:9, 8:13, 25:10, 41:2, 41:6, 41:14, 42:15, 42:16, 45:10, 47:5, 47:15, 48:21, 49:7, 49:24, 51:16, 52:6, 53:19, 54:24, 62:19, 62:23, 66:12, 89:18, 89:20, 90:2, 100:13, 101:4, 101:8, 101:11, 103:15, 104:2, 104:3, 104:22, 107:19, 110:13, 115:7, 124:6, 129:24, 130:7, 130:8, 149:2,

161:19, 164:13, 164:21, 166:8, 167:14, 167:15, 169:7, 169:14, 170:11, 171:2, 174:5, 176:8, 183:19, 183:21, 187:20, 198:10, 202:20, 203:21
case-specific [3] - 66:12, 164:13, 203:21
cases [3] - 44:21, 53:19, 76:1
Cases [1] - 1:4
cat [1] - 92:2
catch [1] - 198:17
categorical [1] - 117:2
categorically [1] - 116:24
caused [3] - 16:21, 16:22, 109:6
causes [1] - 205:17
causing [2] - 29:20, 30:2
caveat [1] - 50:8
centers [1] - 56:21
certain [35] - 14:3, 14:5, 21:8, 23:16, 24:15, 25:2, 25:14, 29:19, 29:20, 45:9, 51:2, 72:15, 79:14, 80:1, 84:1, 93:5, 93:9, 93:11, 95:6, 95:19, 95:22, 96:12, 102:21, 105:16, 108:11, 108:23, 109:7, 111:10, 126:16, 152:11, 152:12, 153:10, 176:3, 201:1, 204:12
certainly [90] - 16:8, 18:7, 19:2, 20:5, 20:13, 22:7, 22:9, 31:5, 33:16, 35:11, 36:6, 36:23, 40:17, 40:24, 41:5, 42:7, 42:8, 45:8, 45:17, 50:10, 55:13, 56:11, 60:1, 63:22, 64:18, 68:9, 69:17, 71:23, 72:5, 72:13, 72:14, 75:3, 76:24, 77:20, 79:14, 80:1, 81:13, 82:8, 82:15, 85:11, 86:17, 86:18, 87:5, 88:14, 89:10, 93:9, 93:18, 95:6, 98:21, 101:10, 105:2, 108:22, 109:6, 109:12, 109:21, 111:9, 116:22, 117:10, 118:2, 119:5, 119:7, 119:11, 120:20, 120:23, 123:12, 125:15, 126:10, 128:12, 129:23, 130:11, 130:23, 133:14, 137:11, 139:1, 139:23, 146:2, 151:21, 151:22, 152:8, 153:6, 168:7, 170:20, 171:20, 172:17, 173:4, 175:14, 184:11, 188:14, 188:21
certainty [1] - 134:9
certification [1] - 207:21
certify [2] - 207:4, 207:9
certifying [1] - 207:22
chain [2] - 201:3, 201:4
chaining [5] - 199:13, 199:23, 201:7, 201:12, 201:16
chance [2] - 75:15, 132:9
change [2] - 16:22, 132:11
Changes [1] - 208:1
changes [2] - 17:6, 208:5
Chapter [1] - 1:4
character [5] - 25:12, 70:21, 71:7, 72:11, 72:12
characteristics [3] - 83:5, 95:23, 141:21
characterization [2] - 63:23, 178:24
characterize [1] - 71:3
characters [11] - 70:24, 71:8, 71:11, 71:12, 72:5, 72:11, 72:15, 72:19, 73:1
Charles [3] - 10:18, 11:5, 58:4
chart [1] - 67:12
check [3] - 147:9, 170:2, 170:14
checks [2] - 22:23, 145:16
choose [2] - 77:24, 78:6
chose [3] - 74:17, 74:18, 104:8
chosen [2] - 84:12, 133:7
Christen [3] - 202:23, 203:2, 203:4
Christen's [1] - 203:23
circular [1] - 172:5
circumstance [4] - 101:9, 162:18, 165:13, 181:22
circumstances [4] - 125:9, 127:14, 164:9, 181:11
cite [5] - 131:7, 131:13, 202:13, 202:15, 203:24
cited [2] - 202:19, 202:23
cites [1] - 131:6
city [3] - 176:8, 177:13, 178:3
Civil [1] - 2:6
claim [3] - 108:16, 169:24, 183:11
claimed [8] - 108:13, 109:1, 110:9, 110:19, 110:22, 113:21, 113:23, 186:4
claims [9] - 169:16, 170:6, 182:24, 183:10, 183:15, 184:24, 185:7, 185:18, 186:16
clarification [1] - 185:4
clarify [2] - 106:23, 190:13
Class [2] - 3:11, 3:12
class [1] - 42:9
CLASS [2] - 1:17, 1:23
clean [2] - 6:22, 70:13
cleaned [1] - 25:4
cleaning [12] - 21:10, 24:18, 24:20, 24:23, 26:12, 62:4, 68:3, 68:6, 68:13, 68:15, 70:12, 72:24
cleansing [5] - 24:16, 26:3, 31:14, 31:22, 31:23
clear [13] - 9:2, 10:24, 21:18, 28:13, 41:15, 78:16, 125:19, 135:8, 146:14, 167:3, 168:7, 175:9, 176:14
clear's [1] - 176:15
clearly [3] - 27:18, 95:20, 96:1
client [1] - 6:24
close [5] - 76:14, 138:20, 138:21, 139:7
closer [1] - 38:18
closing [2] - 102:22, 199:22
closure [50] - 76:22, 77:6, 77:10, 91:7, 91:8, 91:9, 91:20, 92:1, 92:17, 102:1, 102:14, 103:2, 103:8, 116:2, 116:6, 116:16, 117:4, 117:6, 117:8, 117:11, 117:14, 117:16, 117:18, 118:11, 119:1, 119:8, 119:12, 119:17, 120:6, 120:15, 120:24, 121:2, 121:6, 121:13, 122:1, 122:10, 122:13, 122:18, 122:23, 123:16, 124:10, 124:20, 124:22, 125:9, 125:11, 129:21, 130:14, 130:19, 130:21, 199:15
Cluster [10] - 94:21, 95:14, 158:22, 175:16, 175:20, 176:16, 180:5, 180:6, 190:7, 190:8
cluster [71] - 18:1, 18:5, 29:1, 47:3, 47:6, 47:18, 47:23, 47:24, 48:1, 48:16, 49:1, 49:14, 49:22, 50:16, 77:11, 77:12, 84:4, 84:11, 84:24, 85:22, 86:11, 86:18, 87:3, 87:4, 87:9, 89:23, 91:13, 91:14, 91:18, 93:4, 93:7, 94:4, 94:5, 94:20, 94:23, 97:2, 97:10, 97:20, 108:15, 110:2, 110:3, 110:9, 114:17, 119:17, 119:23, 119:24, 121:22, 121:23, 124:21, 124:22, 132:22, 132:24, 133:12, 135:22, 137:15, 139:3, 139:6, 157:21, 160:1, 169:23, 171:8, 175:3, 176:10, 176:12, 181:11, 181:24, 190:3, 190:8, 190:14
clustering [29] - 28:23, 68:11, 92:23, 98:19, 99:23, 114:1, 117:19, 118:11, 118:12, 119:1, 119:4, 121:9, 121:20, 134:24, 135:3, 135:7, 135:18, 136:2, 141:11, 141:22, 160:22, 161:14, 199:13, 199:17, 199:19, 200:10, 201:10, 201:16, 201:23
clusterings [1] - 98:24
clusters [69] - 19:16, 19:19, 28:24, 47:10, 86:19, 93:2, 93:5, 93:8, 93:10, 93:12, 93:19, 97:1, 97:18, 97:19, 99:7, 100:7, 109:17, 112:7, 112:12, 119:18, 119:21, 119:22, 121:9, 133:4, 133:7, 133:15, 133:18, 133:24, 134:5, 134:11, 134:16, 135:2, 135:12, 135:17, 136:8, 137:1, 137:8, 137:11, 137:23, 138:3, 138:8, 138:11, 138:13, 138:22, 139:7, 139:21, 140:3, 140:4, 140:12, 141:19, 141:20, 144:15, 146:18, 171:7, 171:18, 180:4, 180:9, 181:23, 182:16, 183:22, 186:4, 187:7, 189:24, 190:7, 206:6
COCHRAN [1] - 80:10
Cochran's [1] - 80:9
code [56] - 13:5, 13:13, 14:4, 14:6, 14:8, 14:10, 14:12, 15:3, 20:14, 20:20, 21:3, 21:4, 21:7, 21:12, 21:16, 21:20, 22:5, 22:11, 22:20, 22:24, 23:3, 23:9, 23:15, 23:17, 23:19,

24:5, 24:10, 24:12, 24:13, 24:23, 25:19, 25:20, 26:5, 27:7, 31:13, 31:14, 31:20, 32:8, 32:12, 61:15, 65:15, 65:18, 69:16, 69:19, 69:21, 71:10, 72:9, 73:7, 73:16, 73:17, 73:23, 74:10, 74:13, 116:10, 141:1
Code [3] - 4:12, 15:7, 20:10
coding [2] - 26:2, 193:18
collapsing [1] - 192:19
college [1] - 35:20
column [3] - 71:17, 176:21, 177:3
columns [2] - 108:12, 177:16
combination [9] - 52:7, 53:1, 61:22, 62:3, 177:11, 177:14, 177:21, 178:8, 178:9
combinations [1] - 200:8
combine [1] - 123:15
coming [2] - 100:11, 132:15
comma [1] - 145:10
commencing [1] - 2:13
comment [2] - 84:3, 99:5
commentary [1] - 111:10
commenting [1] - 9:16
Commission [1] - 207:17
common [5] - 52:10, 176:3, 176:7, 193:23, 200:2
commonalities [2] - 182:14, 186:21
commonality [2] - 180:15, 190:22
Commonwealth [2] - 2:10, 207:4
COMMONWEALTH [1] - 207:1
companies [5] - 183:16, 184:16, 185:14, 185:19, 186:16
companies' [1] - 187:6
company [4] - 39:12, 51:9, 54:17, 190:9
compare [4] - 108:6, 109:16, 200:17, 201:2
compared [4] - 40:21, 79:14, 80:22, 81:5
comparing [2] - 33:22, 180:3
comparison [4] - 45:2, 65:2, 200:17, 200:21
comparisons [1] - 76:9
complete [5] - 105:20, 132:12, 145:5, 150:8, 204:19
completely [2] - 6:9, 6:17
Complex [1] - 57:3
complex [4] - 38:2, 40:19, 56:13, 56:17
compliance [6] - 38:21, 39:3, 39:9, 39:12, 40:1, 40:4
component [2] - 60:11, 117:18
components [5] - 90:16, 90:24, 91:15, 144:2, 144:5
comport [3] - 22:21, 96:20, 98:24
composed [1] - 93:2
comprised [1] - 70:22
compute [1] - 128:22
computer [8] - 12:4, 34:4, 34:7, 34:11, 34:18, 34:22, 75:21, 75:23
computing [8] - 59:11, 59:21, 60:4, 60:15, 61:2, 61:9, 61:17, 61:18
concentrated [2] - 33:23, 35:4
concept [8] - 22:17, 50:13, 60:17, 78:18, 102:11, 124:5, 124:6, 124:11
conceptual [4] - 54:1, 118:6, 118:8, 118:10
conceptually [1] - 124:18
conclude [1] - 203:14
concluded [3] - 107:19, 184:9, 206:14
conclusion [6] - 108:5, 109:13, 141:11, 142:11, 149:1, 155:16
conclusions [2] - 17:7, 109:4
concrete [3] - 27:11, 55:18, 89:20
conducting [1] - 42:21
confident [1] - 163:19
confirm [9] - 13:6, 13:7, 25:13, 74:14, 81:10, 99:15, 111:2, 147:14, 170:8
confirmed [5] - 153:8, 153:14, 182:24, 183:14
confirming [3] - 11:16, 152:20, 153:9
confusion [1] - 56:24
connected [4] - 90:15, 90:24, 91:15, 91:16
connecting [1] - 91:14
connection [7] - 8:9, 8:21, 9:6, 21:17, 71:14, 108:2, 150:10
conservative [1] - 102:5
consider [17] - 34:22, 42:24, 46:5, 61:16, 92:7, 92:13, 136:5, 159:21, 165:17, 171:21, 171:23, 172:18, 172:20, 173:1, 173:4, 204:10
consideration [1] - 204:18
considered [25] - 25:15, 32:15, 34:20, 44:15, 53:20, 58:16, 62:11, 66:16, 95:4, 97:9, 98:18, 114:4, 123:24, 124:10, 135:13, 150:6, 150:9, 150:19, 150:23, 173:14, 196:5, 200:14, 202:14, 203:2, 204:16
consistency [4] - 70:7, 101:20, 104:11, 126:12
consistent [11] - 20:20, 41:4, 123:16, 144:24, 145:13, 150:22, 152:22, 152:24, 153:2, 156:17, 166:7
consistently [1] - 194:10
consolidating [1] - 31:9
constitute [2] - 15:14, 15:22
contact [2] - 95:9, 192:3
contain [3] - 54:14, 144:2, 190:7
contained [8] - 16:23, 42:19, 150:11, 151:19, 153:22, 155:18, 169:1, 185:11
contains [6] - 94:24, 159:6, 160:15, 161:3, 190:14, 196:1
contended [1] - 107:16
contends [1] - 86:4
contention [1] - 108:3
context [13] - 22:17, 35:9, 58:23, 81:8, 82:4, 84:2, 123:11, 125:7, 127:21, 147:7, 160:10, 169:10, 204:10
contexts [1] - 84:1
continue [2] - 36:15, 77:22
continues [2] - 189:12, 189:14
continuing [8] - 11:7, 11:10, 12:5, 36:23, 37:1, 37:2, 37:8, 38:8
contracts [1] - 39:20
control [2] - 151:9, 207:22
conversant [2] - 45:12, 46:6
conversations [1] - 173:12
converting [1] - 25:10
copy [4] - 4:24, 13:2, 18:23, 33:6
core [2] - 193:10, 194:6
corners [1] - 17:15
corporation [6] - 180:19, 181:1, 181:4, 181:6, 181:18, 187:18
corporations [5] - 181:9, 181:10, 181:22, 182:8, 182:19
correct [70] - 5:22, 8:6, 11:18, 15:23, 17:16, 21:4, 33:18, 34:5, 34:6, 48:8, 51:13, 63:2, 63:3, 63:10, 67:4, 69:7, 74:5, 77:14, 80:5, 80:8, 82:1, 82:14, 93:6, 103:19, 103:22, 106:18, 113:13, 113:16, 114:22, 124:1, 126:2, 128:14, 132:13, 133:8, 133:13, 133:17, 133:19, 144:18, 144:22, 147:10, 148:3, 148:13, 149:12, 152:5, 153:23, 154:13, 155:13, 155:18, 155:22, 156:23, 157:10, 160:19, 168:17, 170:14, 170:16, 174:18, 176:19, 177:18, 178:7, 180:1, 180:6, 185:1, 186:7, 186:9, 190:9, 195:10, 198:9, 200:11, 204:20, 208:11
CORRECTION [1] - 209:6
corrections [1] - 208:12
correctly [4] - 13:8, 52:7, 65:12, 85:3
corresponding [1] - 112:1
counsel [4] - 4:24, 151:12, 207:9, 207:10
count [1] - 177:3
country [3] - 176:9, 177:13, 178:3
couple [3] - 37:15, 37:17, 150:7
course [23] - 18:14, 18:19, 36:3, 36:9, 36:11, 36:13, 37:9, 37:11, 37:12, 37:14, 37:23, 38:6, 39:7, 55:23, 76:19, 133:12, 138:13, 139:12, 143:3, 169:12, 194:24
courses [10] - 34:2, 34:3, 34:11, 34:17, 35:3, 35:10, 35:21, 35:23
coursework [5] - 11:10, 12:6, 35:11, 36:20, 37:4
Court [1] - 193:11
court [3] - 4:23, 172:21, 192:7
COURT [2] - 1:3, 2:21
courtesy [1] - 6:18
covered [1] - 126:4
create [4] - 19:6,

199:7, 200:5, 206:5
created [7] - 18:22, 21:2, 26:15, 96:4, 96:5, 99:10, 135:17
creates [1] - 171:6
creating [7] - 27:1, 27:4, 27:8, 40:15, 40:18, 121:9, 200:4
credits [1] - 34:8
criteria [1] - 177:22
critiques [1] - 17:4
CROSS [1] - 4:5
CSR [1] - 2:7
culmination [1] - 66:11
curious [1] - 110:11
Curran [1] - 207:3
curran [1] - 2:7
CURRAN [1] - 2:21
currancourtrep@yahoo.com [1] - 2:23
current [6] - 9:8, 10:15, 14:20, 15:1, 33:6, 33:8
Curriculum [2] - 4:13, 33:2
cursory [1] - 147:2
CV [4] - 32:24, 33:6, 42:24, 44:5

D

daily [1] - 38:24
damages [5] - 10:11, 11:14, 39:19, 40:16, 40:23
Dan [1] - 5:14
Daniel [1] - 3:5
DARR [2] - 1:12, 1:18
Darr [2] - 2:4, 3:2
dashboard [1] - 54:21
Data [2] - 143:8, 143:16
data [182] - 10:2, 10:5, 10:7, 10:23, 11:2, 11:21, 11:23, 12:6, 12:10, 13:11, 13:12, 13:22, 14:11, 19:2, 19:5, 21:11, 24:16, 34:16, 34:23, 34:24, 35:1, 35:3, 36:22, 38:3, 38:10, 38:19, 38:20, 38:24, 39:23, 41:2, 41:8, 41:19, 42:13, 42:19, 43:10, 44:19, 44:21, 44:24, 46:2, 46:24, 47:2, 51:3, 52:4, 54:14, 54:21, 56:13, 59:6, 59:7, 62:4, 63:1, 68:3, 68:6, 70:4, 70:7, 70:8, 70:14, 72:16, 72:18, 73:9, 73:12, 73:14, 74:2, 75:4, 80:23, 83:10, 84:7, 84:13, 85:2, 86:1, 86:13, 93:19, 100:15, 100:20, 101:1, 101:18, 101:20, 101:22, 103:11, 104:6, 104:10, 104:14, 105:4, 105:5, 105:11, 105:12, 105:18, 107:7, 107:9, 107:10, 107:14, 107:17, 108:4, 108:7, 108:17, 109:1, 109:9, 109:16, 110:1, 110:12, 111:5, 111:19, 111:21, 111:22, 112:1, 112:4, 113:1, 121:9, 126:8, 126:10, 126:15, 129:11, 129:19, 130:3, 130:4, 130:10, 131:2, 131:8, 134:4, 134:6, 134:20, 134:21, 136:6, 136:23, 137:3, 141:13, 142:2, 143:9, 146:24, 147:13, 148:21, 148:24, 149:3, 149:6, 151:22, 151:23, 152:20, 152:23, 152:24, 153:2, 153:14, 153:20, 153:21, 162:10, 162:11, 162:14, 162:23, 164:13, 164:16, 166:2, 166:9, 168:19, 168:24, 169:4, 169:7, 170:12, 171:4, 172:18, 173:19, 173:22, 174:5, 174:9, 175:11, 176:3, 178:12, 179:3, 182:4, 182:9, 182:10, 182:14, 185:10, 193:1, 205:2, 205:4, 205:23, 206:2, 206:3, 206:5, 206:7, 206:10
database [20] - 13:3, 13:4, 13:8, 19:3, 19:13, 19:15, 20:17, 20:19, 44:16, 47:9, 141:2, 141:3, 141:5, 141:7, 141:8, 144:13, 165:4, 165:14, 169:1, 187:1
databases [1] - 13:1
datapoints [2] - 121:12, 146:4
Dataset [1] - 194:19
dataset [46] - 41:3, 41:9, 41:12, 41:16, 41:17, 43:1, 43:2, 53:14, 55:7, 65:20, 78:2, 78:8, 78:13, 78:14, 79:5, 79:7, 79:16, 80:24, 81:3, 83:1, 83:3, 83:5, 83:7, 83:11, 86:6, 86:11, 86:13, 86:23, 94:7, 102:19, 102:22, 112:17, 112:18, 112:23, 113:12, 117:19, 118:12, 119:15, 127:10, 128:2, 140:12, 148:21, 166:23, 195:18, 206:6
datasets [11] - 35:2, 35:5, 40:20, 41:14, 44:5, 54:12, 54:22, 74:21, 75:20, 102:21, 105:21
DATE [2] - 209:2, 209:4
days [2] - 7:2, 7:4
daytime [1] - 37:10
de [1] - 30:11
deal [1] - 180:16
dealing [2] - 47:4, 90:23
deals [1] - 128:1
Debtors [1] - 1:11
decision [2] - 68:19, 71:2
deduping [3] - 54:24, 55:2, 60:5
deduplicate [4] - 51:6, 51:8, 51:23, 123:14
deduplicating [3] - 51:12, 51:17, 54:22
deduplication [7] - 53:23, 54:10, 54:11, 54:13, 55:5, 55:8, 55:12
deemed [5] - 16:6, 16:15, 17:9, 17:13, 86:22
deeper [1] - 66:14
DEFENDANT [2] - 1:17, 1:23
Defendants [3] - 1:17, 1:24, 3:11
define [8] - 43:7, 87:1, 91:7, 127:11, 128:3, 160:6, 160:9, 186:12
defined [4] - 143:8, 143:16, 145:16, 160:5
defines [1] - 91:16
defining [2] - 165:8, 174:23
definitely [2] - 27:14, 30:19
definition [2] - 196:12, 199:9
definitions [3] - 88:23, 89:14, 163:16
degree [11] - 11:6, 11:9, 12:3, 39:1, 88:11, 88:12, 99:7, 121:13, 124:23, 124:24, 149:2
demonstrate [7] - 97:5, 97:24, 99:22, 159:5, 170:19, 171:14, 180:14
demonstrated [2] - 126:11, 134:22
demonstrates [1] - 178:14
DENNIS [4] - 2:2, 5:7, 207:6, 208:10
Dennis [9] - 4:3, 4:9, 4:11, 4:13, 5:2, 5:4, 5:14, 33:2, 173:22
Dennis's [1] - 32:24
denominator [1] - 81:8
depended [1] - 165:12
depiction [1] - 139:19
Deponent's [1] - 208:19
DEPOSITION [2] - 2:2, 209:4
deposition [12] - 5:17, 5:19, 5:23, 6:1, 8:3, 17:2, 115:16, 206:14, 207:5, 207:7, 208:3, 208:6
Deposition [12] - 4:9, 5:1, 5:2, 15:6, 33:1, 67:14, 94:10, 131:24, 132:4, 158:15, 178:18, 189:18
depositions [1] - 106:10
derived [1] - 166:1
describe [8] - 8:11, 39:24, 50:21, 62:14, 64:14, 66:15, 80:15, 91:7
described [11] - 64:19, 68:10, 88:15, 88:17, 90:20, 90:21, 113:21, 152:21, 154:7, 193:4, 205:15
describes [2] - 153:10, 199:17
describing [2] - 77:4, 201:7
Description [1] - 4:8
description [5] - 25:20, 42:1, 152:23, 153:1, 153:14
descriptions [1] - 59:20
designed [5] - 39:13, 202:7, 203:6, 203:13, 203:16
desires [1] - 208:6
detail [6] - 28:1, 87:18, 112:3, 113:3, 141:24, 188:22
detailed [3] - 26:17, 27:19, 189:8
detect [1] - 39:14
determination [18] - 8:19, 68:22, 72:3, 75:8, 75:11, 79:17, 92:10, 92:13, 141:15, 169:3, 182:3, 184:6, 184:8, 186:20, 188:4, 191:9, 203:20, 203:21
determinations [1] - 92:14
determine [53] - 21:6, 24:11, 47:18, 48:4, 48:5, 59:16, 64:24, 66:23, 71:19, 72:8, 75:14, 75:17, 79:11, 82:18, 84:11, 86:10, 87:14, 93:13, 93:21, 95:3, 95:12, 98:2, 103:18, 106:2, 107:4, 107:18, 117:14, 119:13, 127:24, 134:8, 134:12, 134:14, 135:1, 135:16, 136:1, 136:13, 147:13, 148:23, 151:17, 152:17, 155:11, 157:8, 158:23, 159:3,

166:15, 166:16, 167:10, 174:3, 182:17, 186:11, 186:17, 186:23, 198:3
determined [9] - 28:12, 30:23, 59:15, 65:4, 80:17, 108:1, 114:6, 191:19, 197:23
determining [9] - 48:19, 49:7, 49:10, 97:13, 98:3, 145:23, 165:4, 165:17, 168:22
deterministic [7] - 61:18, 61:23, 73:6, 73:8, 73:9, 73:21, 74:4
deviation [1] - 137:20
deviations [1] - 79:4
diagram [2] - 144:11, 155:2
differed [1] - 26:23
difference [5] - 28:20, 29:12, 64:15, 80:12, 118:5
differences [3] - 27:2, 27:5, 74:15
different [36] - 17:20, 19:22, 20:8, 28:17, 41:14, 42:19, 45:1, 52:5, 56:23, 60:21, 62:20, 72:3, 83:6, 84:17, 84:22, 95:8, 95:9, 95:23, 105:22, 110:18, 111:16, 111:17, 116:22, 119:10, 144:2, 144:5, 148:15, 159:10, 173:16, 173:18, 178:10, 201:8, 204:13, 204:14
differentiate [2] - 95:21, 96:19
differentiated [2] - 96:1, 96:14
differently [3] - 59:4, 72:6, 167:7
differs [3] - 64:2, 64:9, 127:17
digitally [1] - 18:23
direct [4] - 94:14, 151:8, 185:8, 207:22
DIRECT [2] - 4:5, 5:12
directed [1] - 40:2
directing [5] - 8:1, 14:14, 20:3, 33:4, 67:18
direction [3] - 20:19, 151:9, 207:22
directions [1] - 137:21
directly [2] - 106:5, 106:6
dirty [1] - 54:22
disaggregating [2] - 69:6, 118:7
disagree [10] - 66:7, 69:4, 69:17, 71:2, 74:16, 75:10, 120:7, 195:22, 198:1, 199:10
disagreed [1] - 69:22
disagreement [2] - 69:24, 201:5
disassociating [1] - 35:5
disassociation [1] - 35:1
discern [1] - 134:1
disclosed [1] - 17:15
discount [1] - 72:4
discovery [1] - 173:13
discrete [2] - 23:1, 23:4
discuss [1] - 195:2
discussed [6] - 24:2, 56:9, 101:17, 135:11, 154:4, 154:6
discusses [1] - 154:3
discussing [1] - 169:18
discussion [2] - 95:24, 132:15
discussions [1] - 21:8
disparate [1] - 79:7
disposal [1] - 52:6
disputes [2] - 39:20, 56:17
Disputes [1] - 57:3
distance [2] - 62:2, 63:5
distill [1] - 75:4
distilled [1] - 40:21
distinct [1] - 34:18
distinction [4] - 28:20, 42:12, 49:1, 176:14
distinguishing [2] - 73:22, 157:15
DISTRICT [1] - 1:3
diverse [1] - 79:7
DL1 [5] - 188:5, 188:6, 188:14, 190:9, 191:9
dlyne@murphyking.com [1] - 3:8
document [4] - 20:4, 20:7, 94:15, 191:4
Document [10] - 4:18, 4:19, 4:20, 4:21, 4:22, 132:1, 132:4, 158:15, 178:18, 189:18
documents [7] - 58:16, 150:6, 150:9, 150:19, 150:22, 153:17, 153:19
Dolarex [8] - 180:5, 180:24, 182:24, 183:11, 183:23, 186:6, 187:12, 187:16
DOLAREX [1] - 180:5
dollar [2] - 148:9, 160:12
domestic [5] - 43:21, 43:23, 43:24, 47:15
done [30] - 5:24, 37:5, 50:11, 51:8, 65:12, 71:9, 75:22, 83:19, 101:19, 104:23, 122:15, 123:20, 123:24, 124:8, 126:1, 126:17, 126:20, 126:21, 127:2, 128:16, 135:24, 136:15, 138:2, 156:9, 174:2, 175:8, 193:13, 198:4, 204:18, 204:20
doubt [1] - 55:4
down [17] - 16:11, 33:8, 42:8, 47:23, 54:18, 68:18, 68:20, 69:5, 75:4, 116:1, 144:14, 144:16, 154:20, 154:22, 173:6, 177:24, 199:14
Dr [132] - 4:14, 4:16, 14:4, 14:8, 18:8, 18:20, 20:7, 21:7, 21:16, 24:19, 25:20, 26:5, 26:16, 31:6, 32:11, 42:20, 43:3, 58:14, 63:23, 64:5, 64:7, 65:10, 65:19, 67:2, 67:15, 67:20, 68:19, 69:15, 73:8, 74:14, 75:14, 76:4, 77:10, 80:3, 80:23, 81:7, 83:20, 84:3, 84:8, 84:12, 85:1, 85:15, 85:23, 86:3, 88:14, 88:16, 89:23, 90:14, 90:20, 90:21, 91:12, 91:22, 92:8, 92:20, 93:6, 94:5, 94:11, 94:17, 94:19, 96:21, 97:18, 98:8, 98:19, 100:1, 100:7, 102:11, 103:18, 104:23, 108:7, 109:16, 110:1, 110:14, 113:5, 114:1, 114:17, 125:8, 125:11, 125:16, 126:11, 128:16, 130:24, 131:6, 133:24, 135:1, 135:17, 136:8, 138:8, 139:5, 140:3, 140:12, 141:8, 141:15, 142:7, 142:8, 143:12, 145:1, 147:3, 150:13, 154:15, 156:18, 157:23, 157:24, 160:15, 163:17, 168:23, 170:8, 171:6, 171:16, 174:2, 175:9, 176:10, 176:15, 176:16, 179:23, 186:18, 187:4, 187:9, 191:13, 191:15, 191:23, 192:6, 192:19, 192:21, 193:13, 193:15, 194:3, 194:13, 198:11, 202:19, 202:24
drill [1] - 154:19
driver's [1] - 5:9
drug [1] - 57:5
dry [2] - 76:13, 198:16
due [2] - 42:10, 201:12
duly [2] - 5:9, 207:6
duplicate [1] - 28:22
duplicated [3] - 26:22, 28:15, 28:18
duplication [19] - 26:21, 27:1, 27:4, 27:8, 27:10, 27:12, 27:13, 27:14, 27:17, 27:19, 27:20, 28:1, 28:12, 29:20, 29:24, 30:3, 30:8, 31:2, 31:11
DURAN [1] - 3:13
during [1] - 132:8
dynamic [5] - 199:19, 200:10, 201:9, 201:15, 201:23

E

e-mail [10] - 2:23, 152:12, 153:4, 153:5, 153:12, 154:7, 176:8, 177:11, 178:4, 178:5
e-mailed [1] - 4:23
e-mails [3] - 153:18, 154:2, 154:7
E-portal [5] - 169:14, 169:22, 171:8, 184:4, 184:20
E-wallet [7] - 105:5, 105:11, 111:18, 112:6, 126:14, 129:6, 145:19
easier [2] - 158:7, 158:13
easy [1] - 144:1
economic [16] - 10:11, 11:14, 39:19, 40:15, 40:23, 99:11, 104:16, 121:18, 182:13, 183:20, 185:24, 186:2, 186:6, 186:11, 186:12, 192:16
economics [1] - 34:2
edges [1] - 200:1
education [14] - 11:3, 11:7, 36:15, 36:24, 37:1, 37:2, 37:8, 38:8, 38:9, 38:14, 38:15, 39:1, 40:2, 40:14
effect [1] - 28:3
efficacy [2] - 40:7, 40:9
eight [2] - 42:18, 138:15
either [18] - 10:1, 18:23, 35:4, 36:16, 70:20, 73:7, 102:2, 103:3, 132:12, 138:3, 138:6, 164:5, 169:23, 181:2, 202:19, 203:12, 203:23, 205:5
Elaine [2] - 177:24, 178:4
elainevit40@gmail.com [1] - 178:6
elected [2] - 71:24, 94:9
electronic [1] - 145:18
elects [1] - 70:18
element [3] - 146:19, 200:13, 200:20
elements [3] - 68:11, 70:15, 70:17
eliminated [2] - 25:14, 70:22

eliminating [1] - 192:7
EM [16] - 64:1, 64:9, 65:4, 65:11, 65:19, 65:24, 66:17, 66:20, 67:2, 127:9, 127:12, 127:17, 128:1, 128:6, 194:9, 194:12
employee [2] - 207:9, 207:10
employs [1] - 191:23
encompasses [1] - 42:14
END [1] - 177:12
end [11] - 6:21, 51:15, 85:14, 91:11, 91:17, 146:20, 155:6, 157:4, 195:22, 201:4, 206:4
endeavor [2] - 148:3, 202:11
endeavored [4] - 120:18, 147:21, 163:9, 167:17
engage [1] - 75:19
engaged [5] - 8:7, 8:24, 36:21, 91:22, 122:16
engagement [25] - 8:12, 8:15, 16:7, 16:16, 17:10, 17:14, 44:20, 51:24, 52:16, 57:24, 59:10, 59:15, 59:16, 60:10, 71:15, 74:19, 76:4, 79:10, 79:21, 82:11, 83:23, 103:18, 124:9, 124:12, 133:13
engagements [20] - 41:24, 42:23, 44:4, 44:12, 50:14, 50:18, 50:22, 53:20, 55:1, 55:16, 57:19, 59:5, 59:9, 59:13, 61:8, 75:19, 76:21, 77:1, 122:24
engages [1] - 70:11
ensuring [1] - 195:24
entered [4] - 70:7, 70:8, 194:10, 208:6
entire [15] - 15:15, 15:22, 23:5, 28:14, 87:18, 112:16, 112:18, 113:11, 144:12, 165:11, 175:19, 191:22, 191:23, 193:8, 194:8
entirely [6] - 25:15, 70:23, 110:6, 142:9, 193:5, 198:1
entirety [1] - 58:15
Entitled [10] - 4:15, 4:19, 4:20, 4:21, 4:22, 67:15, 132:5, 158:16, 178:19, 189:19
entity [6] - 37:5, 54:17, 156:3, 156:12, 157:5, 157:9
entries [3] - 159:19, 175:20, 178:23
ePOC [1] - 183:1
EPOC [36] - 105:4, 105:10, 107:7, 107:9, 107:10, 107:14, 107:16, 108:4, 108:7, 108:16, 109:9, 109:16, 110:1, 110:12, 110:19, 111:5, 112:10, 112:14, 112:16, 112:18, 113:1, 113:12, 113:14, 113:23, 114:2, 114:16, 126:14, 128:23, 145:17, 173:12, 183:6, 183:16, 184:13, 184:21, 185:5, 185:19
equal [1] - 127:8
equally [1] - 169:5
equals [3] - 77:14, 77:15
equity [12] - 133:3, 142:12, 142:17, 154:14, 154:17, 155:18, 156:15, 156:23, 157:2, 171:19, 176:5, 177:19
equivalent [1] - 41:9
errata [2] - 7:5, 208:12
ERRATA [1] - 209:1
error [42] - 86:22, 87:1, 89:5, 89:15, 89:21, 89:22, 90:2, 90:6, 96:24, 97:8, 97:20, 98:2, 98:6, 98:17, 99:14, 99:17, 99:18, 100:8, 100:19, 102:2, 103:10, 115:5, 125:21, 132:13, 132:15, 132:17, 132:19, 132:23, 135:2, 146:16, 148:6, 160:2, 160:21, 161:2, 168:11, 168:13, 168:18, 168:22, 169:3, 169:9, 200:5
errors [13] - 10:14, 11:17, 26:1, 73:7, 147:12, 160:16, 160:18, 168:6, 168:7, 203:9, 204:4, 204:5, 205:5
Esq [3] - 3:5, 3:6, 3:14
essentially [3] - 51:6, 51:7, 192:18
Estates [1] - 3:3
ESTATES [2] - 1:13, 1:19
estimate [4] - 46:12, 78:1, 78:6, 140:19
estimates [1] - 162:7
estimation [1] - 139:8
evidence [3] - 114:3, 159:20, 189:7
evidentiary [2] - 5:20, 7:20
exact [4] - 18:6, 30:21, 84:5, 193:13
exactly [6] - 33:21, 72:21, 85:4, 100:15, 173:8, 191:3
Examination [1] - 4:5
examination [4] - 134:17, 207:7, 208:4, 208:4
EXAMINATION [1] - 5:12
examine [3] - 112:7, 112:13, 112:15
examined [5] - 5:10, 112:8, 133:12, 207:6, 208:10
example [22] - 12:16, 17:24, 18:4, 23:19, 29:4, 29:21, 29:24, 62:24, 70:18, 100:6, 111:4, 114:10, 123:21, 145:20, 160:14, 160:15, 177:23, 183:6, 188:23, 197:5, 197:6, 200:23
examples [15] - 56:16, 96:12, 98:22, 99:6, 100:4, 100:5, 109:19, 109:20, 115:8, 123:13, 133:1, 133:9, 133:16, 173:10, 200:22
Excel [6] - 53:4, 53:5, 60:23, 141:2, 141:23, 176:2
except [2] - 7:12, 7:18
exception [1] - 202:22
Excerpt [1] - 208:1
exclude [1] - 43:20
excluded [3] - 29:17, 192:11, 193:12
exclusion [1] - 50:10
excuse [12] - 6:13, 17:2, 131:4, 138:8, 150:5, 154:21, 174:20, 175:13, 176:16, 195:21, 202:14, 203:5
execute [1] - 140:17
execution [1] - 20:20
Execution [3] - 4:12, 15:7, 20:11
exercise [10] - 6:21, 8:22, 48:17, 49:7, 49:23, 50:7, 114:7, 115:3, 134:24, 136:18
Exhibit [58] - 4:19, 4:20, 4:21, 4:22, 5:1, 5:3, 8:2, 14:15, 15:6, 15:10, 15:14, 15:17, 15:21, 20:1, 20:3, 20:4, 20:21, 32:19, 33:1, 33:5, 67:14, 67:19, 81:24, 82:1, 88:3, 88:4, 94:10, 94:14, 94:19, 131:24, 132:4, 132:5, 144:11, 149:22, 150:5, 154:9, 156:2, 157:22, 158:2, 158:3, 158:15, 158:16, 159:11, 178:14, 178:16, 178:18, 178:19, 178:21, 188:24, 189:8, 189:15, 189:18, 189:19, 189:21, 189:22, 190:6, 202:15, 203:24
exhibit [8] - 17:1, 87:19, 87:23, 95:5, 132:3, 178:17, 188:14, 199:18
exhibits [16] - 1:2, 4:23, 10:13, 16:1, 17:16, 17:23, 86:20, 87:24, 93:10, 108:11, 108:18, 108:20, 113:20, 114:19, 114:21, 115:1
Exhibits [1] - 189:8
exist [2] - 117:2, 122:5
existed [1] - 27:17
existing [1] - 181:9
exists [2] - 181:5, 185:6
expect [2] - 162:3, 162:17
expectation [4] - 63:18, 64:16, 64:23
expected [1] - 163:5
experience [11] - 41:8, 45:3, 45:6, 58:6, 60:14, 61:3, 88:18, 89:4, 89:9, 90:23
Experience [2] - 42:1, 59:20
experiential [6] - 37:3, 38:9, 38:14, 38:15, 40:1, 40:14
Expert [5] - 4:10, 4:16, 5:4, 94:10, 94:16
expert [4] - 8:15, 16:3, 21:17, 143:11
experts [1] - 8:18
Expires [1] - 207:17
explain [1] - 127:17
explicitly [3] - 190:24, 193:19, 193:24
exploded [2] - 178:22, 179:1
exploration [5] - 19:3, 71:14, 72:16, 72:18, 156:18
exploratory [1] - 114:23
explored [5] - 108:23, 109:3, 110:12, 117:1, 136:22
exported [2] - 156:3, 156:12
expose [1] - 74:15
exposed [2] - 58:1, 74:2
extended [1] - 30:3
extends [1] - 160:13
extensive [1] - 126:19
extent [16] - 6:7, 17:12, 29:23, 30:22, 33:23, 44:18, 108:8, 108:10, 123:17, 137:13, 154:14, 181:15, 183:20, 187:17, 197:14
external [2] - 127:9, 193:1
extract [2] - 93:18, 185:10
eyes [1] - 190:5

F

F-A-T-U-R-A [1] - 143:22
F-measure [4] - 90:12, 162:2, 162:16, 163:24
F1 [1] - 90:12
face [3] - 137:2, 137:4, 159:8
fact [28] - 50:12, 62:19, 62:23, 84:9, 84:11, 84:12, 84:21, 86:10, 86:11, 99:10, 101:24, 103:15, 104:10, 105:17, 116:10, 122:11, 125:16, 130:7, 130:8, 159:12, 171:19, 180:12, 180:22, 181:14, 182:10, 182:23, 185:17, 203:20
fact-specific [1] - 203:20
factor [3] - 164:17, 165:3, 165:8
factors [10] - 79:3, 79:5, 164:20, 164:23, 165:1, 165:16, 166:16, 176:4, 177:14, 200:18
facts [10] - 66:12, 75:5, 101:5, 101:9, 124:7, 129:24, 164:14, 166:8, 169:7, 174:5
factually [1] - 50:9
fair [24] - 8:10, 12:9, 16:16, 16:17, 18:24, 40:10, 49:15, 57:23, 89:8, 91:4, 97:18, 99:21, 102:18, 117:24, 118:19, 124:14, 124:16, 124:19, 125:2, 125:5, 125:10, 139:8, 140:22, 140:23
fairly [3] - 13:1, 193:9, 193:22
false [15] - 80:13, 80:16, 80:18, 162:3, 162:18, 162:24, 163:5, 163:22, 166:12, 166:18, 167:1, 167:5, 204:14, 204:15
familiar [10] - 20:16, 57:8, 57:10, 78:18, 80:9, 88:8, 101:24, 102:8, 158:9, 197:16
far [5] - 81:21, 82:9, 176:6, 176:21, 177:2
fashion [4] - 16:9, 54:15, 83:10, 184:22
fast [1] - 101:6
Fatura [1] - 143:21
faulty [1] - 109:14
February [1] - 115:17
Federal [2] - 2:4, 2:5
federal [1] - 57:4
fell [1] - 103:24
FELLEGI [1] - 57:17
Fellegi [18] - 57:11, 57:14, 57:20, 58:1, 58:6, 58:9, 58:18, 58:19, 58:24, 63:1, 63:18, 63:20, 64:15, 64:16, 92:18, 116:15, 127:7, 127:18
Fellegi-Sunter [18] - 57:11, 57:14, 57:20, 58:1, 58:6, 58:9, 58:18, 58:19, 58:24, 63:1, 63:18, 63:20, 64:15, 64:16, 92:18, 116:15, 127:7, 127:18
Fer [1] - 29:21
Fer-Do [1] - 29:21
Fernando [1] - 29:21
FESM [2] - 194:3, 194:21
few [1] - 51:2
fewer [2] - 71:18, 71:20
Field [1] - 194:18
field [22] - 25:2, 27:22, 28:6, 28:11, 28:15, 28:18, 28:21, 29:13, 31:3, 59:6, 68:13, 68:15, 69:12, 70:11, 72:4, 72:24, 82:3, 131:5, 187:17, 192:14, 195:18, 205:2
fields [11] - 27:21, 28:11, 30:4, 52:5, 65:1, 65:4, 65:11, 66:1, 66:15, 70:14, 186:24
fight [1] - 69:3
file [1] - 43:9
files [4] - 20:8, 22:19, 22:20, 141:23
filing [1] - 151:2
filings [2] - 106:9, 151:3
filter [2] - 38:3, 68:20
filtering [3] - 68:17, 69:10, 155:4
final [1] - 26:14
finance [1] - 34:2
financial [1] - 39:22
Financial [1] - 3:3
FINANCIAL [3] - 1:11, 1:14, 1:20
financially [1] - 207:11
fine [2] - 145:10, 185:23
finish [1] - 6:16
first [20] - 24:21, 39:2, 39:8, 55:24, 56:4, 57:24, 97:14, 98:3, 98:15, 100:11, 130:13, 130:19, 148:22, 159:10, 159:13, 171:3, 177:5, 190:3, 190:8, 195:1
fit [1] - 92:18
five [10] - 42:3, 159:10, 159:13, 159:19, 196:7, 197:1, 197:2, 197:8, 198:16, 198:20
fixing [7] - 50:24, 51:24, 53:19, 54:24, 56:9, 56:10, 60:2
flawed [2] - 187:10, 193:5
Floor [1] - 2:12
focus [8] - 31:19, 37:1, 43:21, 54:9, 89:1, 89:3, 147:8
focused [5] - 26:7, 26:11, 32:4, 95:24, 115:12
focusing [6] - 43:2, 54:21, 55:15, 82:10, 84:23, 99:13
follow [1] - 130:20
followed [1] - 129:20
following [1] - 207:5
following-named [1] - 207:5
follows [1] - 5:11
footnote [4] - 28:3, 81:18, 81:22, 178:13
footnotes [1] - 11:16
foregoing [1] - 207:21
forgotten [1] - 94:8
form [24] - 7:12, 7:18, 8:8, 18:24, 27:3, 40:23, 65:9, 65:18, 66:1, 66:24, 72:22, 82:5, 82:12, 93:1, 96:15, 109:24, 113:24, 121:2, 121:17, 152:2, 157:12, 157:18, 180:11, 208:5
formal [5] - 11:20, 11:22, 11:24, 37:2, 38:7
formally [1] - 36:16
formed [5] - 16:6, 17:13, 79:20, 109:9, 157:17
forming [2] - 14:8, 150:10
Formula [1] - 80:10
formula [1] - 163:18
formulas [1] - 163:20
forth [3] - 8:18, 20:21, 155:21
foundation [1] - 193:4
four [15] - 17:15, 42:3, 42:17, 52:17, 52:19, 52:21, 71:8, 71:12, 72:1, 128:15, 138:13, 143:7, 195:7, 196:6, 196:11
fourth [3] - 42:8, 56:11, 143:15
frame [2] - 115:17, 145:6
frames [1] - 46:2
franchisor [1] - 56:21
frankly [3] - 105:6, 126:7, 199:22
Frantz [8] - 3:11, 106:16, 106:19, 107:1, 107:4, 151:4, 152:9
FRANTZ [1] - 1:16
fraud [1] - 39:14
Freer [75] - 4:17, 14:4, 14:8, 18:8, 19:13, 19:14, 19:16, 20:7, 24:12, 26:16, 42:20, 43:3, 58:14, 65:19, 67:2, 69:15, 75:14, 76:4, 77:10, 80:3, 81:7, 83:20, 84:3, 84:8, 84:12, 85:1, 85:15, 85:23, 86:3, 90:20, 90:21, 91:22, 92:8, 92:20, 94:5, 94:11, 94:17, 104:23, 110:14, 112:7, 114:7, 117:7, 117:10, 121:6, 121:15, 125:8, 126:11, 128:16, 130:24, 131:6, 135:17, 138:8, 141:8, 142:7, 142:8, 143:12, 143:16, 150:4, 150:13, 154:10, 154:15, 159:5, 160:15, 163:17, 168:23, 171:6, 179:23, 187:4, 191:23, 192:6, 193:13, 194:13, 202:19, 202:24
freer's [1] - 175:9
Freer's [74] - 4:14, 18:20, 21:7, 21:16, 24:19, 25:20, 26:5, 31:6, 32:11, 63:23, 64:5, 64:7, 65:10, 67:15, 67:20, 68:19, 73:8, 74:14, 80:23, 88:14, 88:16, 89:23, 90:14, 91:12, 93:6, 94:19, 96:21, 97:18, 98:8, 98:19, 100:1, 100:7, 102:11, 103:18, 108:7, 109:16, 110:1, 113:5, 114:1, 114:17, 116:5, 125:8, 125:11, 125:16, 133:24, 135:1, 136:8, 138:14, 139:5, 140:3, 140:12, 141:15, 145:1, 147:3, 156:18, 157:23, 157:24, 158:6, 170:8, 171:16, 174:2, 176:10, 176:16, 186:18, 187:9, 191:13, 191:15, 192:19, 192:21, 193:15, 194:3, 198:11
FREITAS [2] - 1:23, 151:5
Freitas [2] - 3:12, 151:4
frequently [5] - 39:22, 46:8, 78:1, 78:7, 110:5
front [6] - 43:10, 86:2, 88:1, 92:21, 121:7, 149:21
FS [18] - 64:1, 64:2, 64:9, 65:4, 65:11, 65:19, 65:24, 66:17, 66:20, 67:2, 127:9, 127:12, 127:17, 128:1, 128:6, 194:9,

194:12
FS-EM [16] - 64:1, 64:9, 65:4, 65:11, 65:19, 65:24, 66:17, 66:20, 67:2, 127:9, 127:12, 127:17, 128:1, 128:6, 194:9, 194:12
fulfilling [2] - 104:13, 162:9
full [9] - 20:22, 21:15, 36:12, 118:21, 119:17, 119:21, 119:22, 144:6
fully [3] - 6:9, 46:6, 208:3
fulsome [4] - 95:16, 95:17, 135:6, 142:2
function [5] - 55:8, 55:13, 146:17, 149:5, 166:3
functions [2] - 179:21, 199:7
fuzzy [21] - 52:8, 52:12, 52:24, 53:5, 60:4, 60:6, 60:10, 60:13, 60:16, 60:17, 61:4, 61:24, 62:15, 63:14, 68:3, 68:5, 68:8, 68:12, 179:20, 179:23, 197:6

G

gamma [2] - 197:1, 197:3
garbage [3] - 25:15, 70:24, 73:2
gee [2] - 146:24, 171:18
geez [1] - 150:14
general [13] - 53:23, 64:20, 64:22, 84:6, 85:2, 86:5, 86:12, 94:6, 123:7, 123:10, 130:2, 181:19, 205:21
generality [1] - 166:14
generally [34] - 7:14, 16:17, 37:18, 38:20, 45:20, 45:23, 46:18, 54:7, 55:6, 58:10, 58:17, 61:20, 79:2, 83:8, 84:12, 85:24, 88:22, 89:6, 89:7, 91:3, 98:7, 102:9, 102:13, 103:11, 121:8, 127:18, 130:6, 130:11, 130:17, 131:1, 164:22, 166:10, 166:11, 168:14
generate [2] - 19:10, 19:12
generated [11] - 18:12, 18:14, 18:19, 18:21, 18:22, 19:17, 19:19, 19:20, 142:15, 143:10
gestures [1] - 6:20
gesturing) [1] - 158:18
gist [1] - 85:5
given [19] - 78:7, 78:12, 78:17, 101:9, 106:1, 128:15, 132:24, 137:18, 162:14, 164:3, 166:22, 181:15, 192:1, 194:22, 195:1, 196:8, 207:8, 208:7
Gladys [3] - 9:22, 10:8, 11:13
go-to [1] - 46:17
goal [13] - 55:10, 59:14, 96:21, 97:24, 121:15, 121:21, 122:12, 124:20, 137:12, 159:3, 174:2, 195:24, 196:13
goals [2] - 99:1, 100:2
gold [2] - 128:2, 128:6
golden [2] - 100:14, 100:21
gosh [2] - 110:16, 138:12
graduate [1] - 11:10
graduated [2] - 36:15, 36:16
graph [1] - 90:17
great [2] - 39:1, 180:16
greater [1] - 196:8
greatly [1] - 96:13
ground [3] - 6:3, 127:10, 128:2
ground-true [1] - 127:10
ground-truth [1] - 128:2
Group [6] - 180:4, 180:19, 182:23, 183:10, 186:6, 187:11
group [1] - 123:13
grouped [1] - 123:19
grouping [2] - 54:3, 54:7
groupings [2] - 195:21, 196:2
groups [2] - 19:20, 159:10
grow [3] - 82:21, 83:9, 83:13
guess [14] - 21:22, 43:9, 51:18, 62:21, 71:16, 78:16, 114:22, 119:20, 138:13, 140:14, 167:3, 167:14, 168:17, 197:11
guesstimate [3] - 138:24, 139:15, 140:10
guide [2] - 48:3, 88:1

H

Hackett [4] - 188:6, 188:15, 190:17, 191:7
Hacketts [1] - 191:4
half [5] - 22:14, 22:15, 36:11, 76:14, 147:4
half-year [1] - 36:11
hand [4] - 14:17, 176:21, 177:3, 207:13
handful [4] - 46:10, 46:11, 46:16, 135:12
happy [2] - 194:14, 195:2
hard [13] - 4:24, 18:23, 43:4, 45:2, 49:4, 62:21, 66:5, 66:9, 101:6, 118:7, 138:24, 139:10, 140:14
hard-and-fast [1] - 101:6
hardcopy [1] - 56:2
head [1] - 16:11
head) [1] - 156:5
healthcare [1] - 204:13
hear [4] - 25:7, 54:5, 76:8, 85:3
heard [1] - 82:11
hearing [2] - 7:20, 103:1
hearings [1] - 5:21
heavily [1] - 193:17
help [6] - 13:6, 64:24, 88:1, 134:7, 171:12, 195:5
helped [7] - 10:10, 10:12, 13:3, 13:5, 13:11, 14:3, 114:9
helpful [6] - 170:20, 170:21, 171:21, 171:22, 173:14, 195:8
helps [1] - 59:16
hence [1] - 56:24
hereby [1] - 207:4
hereunto [1] - 207:12
high [4] - 64:20, 64:22, 99:7, 101:18
high-level [2] - 64:20, 64:22
Hill [1] - 2:22
him/her [1] - 208:4
hit [1] - 140:17
hoc [4] - 75:8, 75:10, 100:4, 100:5
hold [1] - 189:5
holds [3] - 78:2, 78:7, 78:10
honestly [3] - 139:9, 163:15
hour [2] - 76:14, 198:19
house [2] - 10:12, 11:14
Hubbs [2] - 10:19, 11:8
HUBBS [1] - 10:19
huhs [1] - 6:20
hundred [2] - 30:19, 30:20
hundreds [1] - 30:17
Huron [11] - 41:20, 107:11, 109:10, 109:13, 115:9, 121:3, 140:3, 142:14, 143:8, 155:9, 157:4
Huron's [1] - 156:23
hybrid [1] - 157:16
hyphen [1] - 57:17
hypothetical [4] - 134:19, 164:3, 171:5, 171:6

I

i.e [1] - 196:3
ID [5] - 94:21, 176:16, 177:5, 183:12, 190:3
identification [10] - 5:5, 15:9, 33:3, 67:17, 94:12, 132:1, 132:6, 158:17, 178:20, 189:20
identified [6] - 5:8, 62:8, 93:5, 112:15, 135:11, 187:7
identifier [1] - 113:1
identifiers [3] - 159:17, 190:23, 193:24
Identifiers [1] - 189:10
identify [4] - 94:15, 95:3, 99:17, 184:12
IDF [4] - 62:1, 62:8, 63:9, 63:11
III.B.1 [1] - 195:19
ijr@mrdklaw.com [1] - 3:16
illustration [2] - 86:5, 93:24
illustrations [2] - 94:2, 97:5
Ilyas [2] - 3:14, 6:23
imagine [1] - 42:4
impact [13] - 71:21, 132:23, 133:2, 133:3, 146:19, 147:11, 148:5, 161:8, 161:9, 164:16, 165:20, 167:11, 167:23
impacts [3] - 161:4, 161:7, 169:10
imparting [1] - 126:9
imperfect [2] - 102:12, 103:4
imperfectly [1] - 94:8
implemented [1] - 39:13
implicated [1] - 30:7
implicit [2] - 194:8, 197:14
imply [1] - 152:14
import [2] - 181:17, 187:21
important [15] - 6:12, 6:15, 16:15, 42:12, 66:16, 66:19, 66:22, 124:17, 147:6, 164:19, 164:23, 165:1, 165:3, 195:20, 204:10
improves [2] - 63:19, 64:1
IN [1] - 207:12
inaccuracies [1] - 25:22
inaccurate [5] - 155:19, 162:11, 169:4, 169:5, 192:22
inappropriate [2] - 66:3, 67:4
inaudible) [4] - 25:5, 54:3, 76:6, 143:19
inaudible] [1] - 138:4
Inc [5] - 3:3, 3:4, 188:5, 188:6, 190:9

INC [6] - 1:10, 1:11, 1:13, 1:14, 1:19, 1:20
include [10] - 16:5, 16:14, 17:24, 18:13, 39:15, 108:11, 127:22, 133:7, 133:16, 143:17
included [14] - 18:8, 28:1, 30:1, 68:23, 87:2, 87:3, 87:6, 93:9, 108:18, 109:1, 110:15, 111:10, 139:22, 143:10
includes [6] - 40:14, 40:19, 68:8, 189:9, 189:24, 198:5
including [8] - 29:20, 33:12, 94:24, 97:1, 108:19, 194:8, 203:2, 203:3
inclusion [1] - 17:10
incomplete [1] - 192:22
inconsistent [1] - 100:1
incorrect [7] - 67:5, 89:12, 103:20, 103:23, 108:5, 157:10, 170:17
incorrectly [2] - 65:12, 74:4
increases [1] - 82:22
independent [7] - 65:9, 72:23, 79:16, 88:18, 90:23, 117:22, 136:4
independently [2] - 64:11, 64:13
index [1] - 87:24
indicated [1] - 152:10
indicates [1] - 156:16
indication [1] - 84:7
indicia [5] - 100:3, 105:6, 110:20, 137:13, 183:7
individual [34] - 47:9, 48:6, 49:9, 50:2, 62:22, 70:14, 77:7, 93:1, 97:1, 99:9, 109:18, 109:20, 118:13, 120:13, 120:14, 121:10, 121:11, 121:16, 121:18, 124:21, 146:19, 146:21, 153:18, 159:7, 160:14, 169:10, 169:12, 170:7, 176:23, 178:1, 183:9, 186:14, 186:15
individualized [1] - 54:19
individuals [3] - 10:22, 145:20, 148:5
inevitably [3] - 203:7, 204:3, 205:4
inform [1] - 114:10
informally [1] - 36:17
information [32] - 16:18, 16:21, 19:21, 77:8, 95:10, 100:17, 104:6, 104:7, 126:9, 134:7, 134:13, 142:20, 143:13, 144:7, 145:24, 151:19, 156:3, 156:12, 156:13, 157:9, 162:8, 170:20, 170:22, 170:24, 173:1, 173:3, 173:7, 173:10, 185:7, 192:3, 197:12, 208:12
informative [1] - 149:1
infringement [1] - 39:21
inherent [4] - 116:2, 125:1, 125:10, 162:10
inherently [7] - 68:6, 101:2, 102:12, 102:24, 117:11, 134:3, 162:7
initial [2] - 143:11, 156:1
input [1] - 77:2
inputs [10] - 21:19, 58:8, 58:15, 58:19, 62:16, 66:16, 66:19, 67:3, 67:4, 70:3
inserted [1] - 28:3
instance [8] - 55:10, 82:14, 98:15, 122:11, 125:23, 148:22, 171:3, 183:9
instances [9] - 77:6, 98:22, 98:23, 99:6, 99:23, 108:13, 108:23, 159:6, 189:9
instruction [1] - 179:14
Instructions [3] - 4:12, 15:8, 20:11
instructions [2] - 15:4, 20:21
insufficient [2] - 172:2, 172:3
integrated [1] - 40:20
intellectual [3] - 10:10, 39:21, 56:17
intended [1] - 206:5
intention [3] - 16:4, 16:13, 21:14
interacted [1] - 144:8
interest [2] - 44:10, 47:14
interested [2] - 60:9, 207:11
interesting [2] - 6:2, 44:7
internal [1] - 40:4
internally [1] - 8:22
international [2] - 43:24, 47:16
interviews [1] - 151:14
intro [1] - 36:2
invalid [1] - 191:16
invalidate [4] - 164:6, 164:10, 165:5, 165:11
investigation [4] - 18:15, 18:19, 38:21, 39:4
invoice [3] - 43:18, 154:21, 154:23
invoices [1] - 144:4
involve [2] - 44:5, 68:2
involved [6] - 12:24, 34:1, 47:2, 59:10, 75:3, 76:22
involves [3] - 38:19, 39:22, 41:2
irrespective [2] - 134:12, 134:23
isolate [2] - 93:12, 154:19
isolation [1] - 66:7
issue [12] - 31:3, 51:3, 68:19, 69:9, 73:19, 85:9, 100:13, 104:8, 168:10, 197:15, 201:7, 201:11
issued [1] - 185:18
issues [18] - 65:14, 69:14, 69:20, 70:12, 73:15, 98:1, 101:16, 134:1, 134:17, 135:14, 136:20, 159:6, 175:9, 175:10, 175:11, 191:22, 199:22
itself [20] - 20:17, 24:18, 26:2, 65:15, 69:16, 97:6, 101:18, 101:22, 105:19, 113:2, 126:8, 126:10, 134:4, 135:3, 135:22, 136:6, 162:10, 166:2, 169:7, 185:9

J

J-A-C-C-A-R-D [1] - 63:8
Jaccard [4] - 62:2, 62:8, 63:7, 63:8
Jacqueline [2] - 2:6, 207:3
Jaff [4] - 183:9, 183:21, 184:23, 185:18
JAFF [1] - 184:24
James [2] - 10:19, 11:8
Jaro [1] - 53:10
Jaro-Winkler [1] - 53:10
join [2] - 38:3, 201:9
joined [12] - 40:20, 142:1, 142:16, 142:23, 181:10, 181:23, 182:9, 190:17, 191:1, 198:24, 199:1, 200:9
joining [3] - 142:4, 142:5, 205:18
Jointly [1] - 1:7
Joshua [8] - 4:3, 4:9, 4:11, 4:13, 5:2, 5:4, 33:1, 173:22
JOSHUA [4] - 2:2, 5:7, 207:5, 208:10
judge's [1] - 192:7
jumped [1] - 74:2

K

keep [1] - 56:15
kind [7] - 8:23, 13:23, 21:10, 26:17, 102:2, 116:3, 169:6
kinds [1] - 204:9
KING [3] - 2:11, 3:5, 3:13
knowing [3] - 100:11, 126:8, 162:21
knowledge [9] - 24:4, 45:14, 45:16, 98:14, 101:21, 104:16, 104:19, 104:24, 208:11
known [2] - 98:12, 100:14

L

L-E-U-N-G [1] - 9:23
lack [3] - 70:4, 128:1, 162:8
lacked [1] - 142:3
lacks [2] - 99:24, 126:10
laid [1] - 8:15
language [1] - 86:3
large [16] - 13:1, 13:2, 38:18, 40:19, 41:2, 41:3, 44:5, 47:9, 74:21, 75:3, 75:20, 108:24, 149:6, 149:10, 149:12, 206:6
large-scale [1] - 41:2
largely [1] - 11:15
larger [1] - 26:19
last [9] - 56:12, 76:1, 123:5, 138:14, 139:12, 149:20, 151:1, 155:21, 189:6
law [1] - 2:10
layout [1] - 158:10
lays [1] - 67:23
lead [1] - 199:21
leading [1] - 25:9
learning [1] - 38:1
least [19] - 13:3, 27:22, 29:12, 35:11, 36:1, 43:17, 66:22, 68:10, 68:15, 75:13, 108:14, 109:21, 113:19, 115:13, 135:2, 151:21, 152:9, 159:10, 193:21
leaving [4] - 38:8, 69:2, 73:13, 73:17
led [2] - 39:1, 66:11
legal [14] - 68:21, 69:2, 106:9, 151:2, 151:3, 181:13, 181:17, 182:2, 182:11, 186:19, 186:22, 187:2, 187:21, 188:4
legally [1] - 50:9
length [1] - 70:21
lengths [1] - 72:15
less [5] - 22:15, 138:19, 161:5, 161:8, 191:8
Leung [2] - 9:22, 10:8
level [11] - 54:19, 55:11, 61:16, 64:20, 64:22, 121:20, 125:12, 125:13,

125:16, 125:17, 125:21
Levenshtein [7] - 53:7, 53:9, 60:19, 60:24, 62:2, 62:9, 63:5
Levenshtein-based [1] - 53:7
liability [2] - 181:17, 187:19
liberal [2] - 102:7, 205:19
license [1] - 5:9
likelihood [1] - 182:12
likely [8] - 28:19, 40:11, 52:9, 53:4, 110:14, 123:3, 123:6, 202:20
limit [2] - 196:13, 203:15
limited [1] - 114:20
Linda [4] - 188:5, 188:15, 190:17, 191:7
line [3] - 178:13, 199:14
LINE [1] - 209:6
line-by-line [1] - 178:13
linear [1] - 83:9
linearly [1] - 83:12
link [14] - 164:5, 164:10, 165:5, 165:10, 165:20, 165:22, 167:10, 167:22, 167:23, 167:24, 168:1, 168:3, 184:14
linkage [10] - 57:16, 59:2, 61:21, 75:16, 166:16, 182:23, 183:13, 183:19, 196:5, 196:14
linked [1] - 184:22
links [6] - 166:11, 166:17, 166:23, 166:24, 167:18, 167:21
list [6] - 32:17, 32:18, 55:21, 144:14, 150:8, 150:23
listed [12] - 56:4, 87:9, 95:13, 142:22, 142:24, 143:7, 143:15, 159:11, 181:6, 183:11, 183:12, 191:8
literacy [1] - 45:21
literature [1] - 103:7
litigation [4] - 38:21, 39:17, 40:13, 204:13
live [1] - 106:14
lives [1] - 19:7
LLC [5] - 1:10, 1:13, 1:19, 3:3, 3:13
load [2] - 13:3, 13:11
loaded [2] - 13:8, 113:11
loading [1] - 112:23
located [1] - 29:24
logarithmic [1] - 83:16
login [3] - 113:2, 113:14
logins [1] - 113:22
look [92] - 20:16, 21:13, 22:16, 22:18, 23:21, 23:22, 24:24, 26:24, 28:4, 30:12, 33:10, 34:14, 40:7, 42:18, 44:3, 44:21, 52:22, 53:12, 55:11, 61:20, 63:12, 64:19, 65:6, 66:6, 67:7, 67:10, 75:9, 75:12, 79:2, 79:6, 79:9, 81:4, 81:12, 87:11, 87:17, 89:10, 91:10, 94:9, 104:4, 104:5, 105:10, 105:11, 107:7, 107:9, 107:13, 108:23, 109:7, 111:13, 111:19, 111:21, 112:3, 117:21, 120:19, 123:22, 126:23, 129:10, 129:17, 132:21, 133:21, 133:22, 134:5, 136:7, 136:18, 137:12, 140:8, 141:12, 146:15, 146:23, 147:2, 147:17, 148:11, 148:14, 148:17, 148:19, 148:20, 152:7, 159:13, 163:16, 163:17, 165:19, 165:21, 166:15, 167:9, 172:16, 174:2, 175:8, 177:2, 185:7, 195:4, 200:1
looked [35] - 23:16, 25:13, 26:13, 29:10, 38:2, 52:3, 66:20, 72:6, 72:14, 72:17, 75:4, 80:1, 84:17, 84:21, 86:18, 95:6, 105:13, 106:3, 106:8, 107:18, 109:18, 111:9, 123:15, 129:7, 132:20, 136:22, 137:4, 137:11, 139:11, 141:7, 151:22, 154:1, 162:23, 183:6, 200:23
looking [27] - 24:9, 26:11, 40:15, 49:8, 52:4, 65:17, 70:3, 81:5, 89:16, 91:13, 93:22, 105:4, 105:5, 105:18, 105:19, 107:13, 111:8, 116:10, 126:14, 137:8, 143:6, 145:17, 170:14, 174:6, 202:12
looks [5] - 64:24, 150:22, 182:11, 197:4, 200:15
Lopes [2] - 177:6
loser [7] - 48:5, 49:8, 145:11, 146:22, 169:21, 184:7, 192:16
losers [4] - 8:20, 47:19, 144:17, 184:11
lost [1] - 165:7
lower [2] - 25:10, 171:19
lunch [2] - 131:17, 132:10
Lunch [1] - 131:22
Lusette [1] - 151:3
Lyne [5] - 3:5, 4:6, 4:24, 5:13, 5:14
LYNE [30] - 6:23, 7:2, 7:10, 7:21, 9:13, 9:16, 14:16, 15:2, 15:18, 32:13, 32:18, 32:22, 33:12, 67:11, 76:15, 76:19, 88:5, 124:3, 131:15, 132:2, 132:7, 143:21, 149:19, 149:23, 150:1, 158:7, 178:15, 189:16, 198:20, 206:11

M

MA [2] - 3:7, 3:15
magnitude [3] - 18:18, 30:13, 30:21
mail [11] - 2:23, 51:9, 152:12, 153:4, 153:5, 153:12, 154:7, 176:8, 177:11, 178:4, 178:5
mail-ins [1] - 51:9
mailed [2] - 4:23, 4:24
mails [3] - 153:18, 154:2, 154:7
maintained [1] - 107:11
major [1] - 12:3
majority [3] - 96:5, 134:14, 134:15
management [3] - 12:3, 33:17, 33:20
manipulate [1] - 38:3
manipulating [1] - 35:4
manipulation [2] - 34:24, 41:8
manual [1] - 138:10
manually [5] - 136:23, 138:3, 138:7, 138:11, 138:22
Maplewood [1] - 178:3
March [1] - 115:17
Marco [1] - 3:12
MARCO [1] - 1:22
Maria [8] - 18:4, 18:9, 85:12, 94:2, 94:4, 94:23, 95:1, 157:21
mark [3] - 132:2, 149:23, 178:16
marked [15] - 15:8, 15:13, 20:2, 33:2, 33:5, 67:16, 67:19, 94:11, 132:1, 132:5, 156:2, 158:16, 178:19, 189:19, 194:18
marketing [1] - 34:3
Martin [20] - 9:12, 9:17, 12:24, 71:9, 71:23, 107:16, 115:14, 119:3, 120:1, 120:21, 122:11, 142:15, 143:10, 156:1, 177:20, 192:8, 192:10, 193:11, 202:20
Martin's [10] - 13:15, 72:2, 72:7, 108:3, 109:12, 113:4, 142:11, 157:4, 184:9, 192:8
MASSACHUSETTS [2] - 1:3, 207:1
Massachusetts [4] - 2:10, 2:12, 2:22, 207:4
master's [1] - 10:6
match [11] - 80:17, 101:21, 112:24, 113:8, 114:16, 129:20, 130:13, 130:20, 177:15, 178:2, 196:9
matched [2] - 113:14, 195:23
matches [5] - 80:18, 80:20, 92:24, 200:16, 200:19
matching [20] - 52:8, 52:12, 52:24, 60:6, 60:11, 60:13, 60:16, 60:18, 61:5, 62:1, 62:15, 63:14, 68:3, 68:5, 68:8, 68:12, 73:9, 179:20, 179:24, 197:6
Materials [1] - 202:16
materials [7] - 32:15, 150:9, 154:2, 202:14, 203:24, 204:1, 205:1
math [10] - 11:6, 11:9, 30:15, 36:6, 36:7, 36:10, 69:7, 83:14, 168:23, 169:6
mathematical [3] - 69:10, 69:14, 74:6
mathematics [4] - 11:6, 12:11, 65:7, 69:21
matter [8] - 37:23, 42:9, 50:23, 50:24, 54:1, 60:2, 105:18, 187:12
matters [4] - 38:22, 54:8, 54:14, 75:3
Matura [1] - 143:20
maximization [3] - 63:19, 64:16, 64:24
MBA [4] - 11:8, 33:21, 33:22, 33:23
MBA-type [1] - 33:21
mean [45] - 12:18, 13:7, 18:21, 19:14, 28:23, 29:3, 37:20, 38:14, 40:18, 41:19, 45:16, 53:10, 53:22, 58:12, 68:21, 73:9, 78:11, 78:12, 78:17, 88:13, 88:22, 91:24, 93:15, 95:18, 98:11, 100:23, 101:15, 103:22, 105:20, 106:7, 112:13, 118:12, 119:20,

119:23, 140:7, 140:15, 148:8, 153:2, 153:11, 153:17, 153:21, 165:23, 171:13, 175:23, 206:3
meaning [1] - 145:23
meaningful [1] - 44:9
means [1] - 177:7
meant [3] - 106:12, 150:18, 153:19
measure [5] - 21:23, 90:12, 162:2, 162:16, 163:24
measures [3] - 162:7, 162:22, 163:1
mechanism [1] - 195:24
medical [1] - 121:11
members [1] - 155:4
memorized [1] - 163:21
memory [1] - 9:10
mentioned [1] - 23:8
Merit [2] - 2:8, 207:3
meshed [1] - 141:16
messy [1] - 54:12
method [7] - 58:9, 58:19, 58:20, 63:2, 63:20, 65:19, 116:15
Methodology [5] - 4:12, 4:15, 15:7, 20:10, 67:16
methodology [68] - 8:17, 62:16, 89:23, 97:6, 100:1, 116:3, 116:6, 117:7, 117:11, 131:9, 137:13, 137:16, 154:11, 156:19, 160:13, 160:22, 161:3, 161:24, 162:15, 164:4, 164:7, 164:10, 164:12, 165:6, 165:10, 165:11, 166:4, 166:6, 167:9, 167:12, 168:2, 168:24, 174:7, 175:10, 179:23, 181:21, 186:18, 187:4, 187:5, 187:10, 187:13, 191:14, 191:16, 191:17, 191:22, 191:23, 192:13, 192:19, 192:21, 193:5, 193:8, 193:16, 194:3, 194:7, 194:8, 194:12, 194:22, 197:17, 202:7, 203:6, 203:17, 204:2, 204:11, 205:4, 206:1, 206:5
methods [1] - 61:4
metric [1] - 44:3
Mickey [1] - 181:16
Microsoft [1] - 35:12
middle [1] - 151:2
MIDDLESEX [1] - 207:1
might [6] - 39:11, 61:23, 71:20, 94:7, 110:11, 124:9
MILLIGAN [1] - 3:13
million [22] - 43:16, 43:17, 43:22, 44:2, 47:13, 47:14, 48:4, 49:9, 49:24, 68:18, 68:20, 69:5, 69:6, 136:24, 137:1, 138:4, 138:7, 144:13, 154:22, 191:7
mind [11] - 35:17, 48:24, 105:17, 132:15, 144:20, 183:14, 201:22, 202:6, 203:5, 204:1, 205:3
mine [1] - 171:18
minimis [1] - 30:12
minimize [4] - 51:9, 76:6, 76:8, 76:9
minimizing [1] - 75:23
minimum [9] - 60:1, 68:9, 70:19, 82:21, 128:18, 128:23, 129:5, 129:12, 159:13
minor [3] - 12:4, 34:4, 34:7
minority [1] - 134:15
minutes [2] - 198:17, 198:21
misread [1] - 190:3
missed [20] - 164:5, 164:10, 165:5, 165:9, 165:20, 165:21, 166:11, 166:17, 166:23, 166:24, 167:10, 167:18, 167:21, 167:22, 167:23, 167:24, 168:1, 168:2, 171:10
missing [1] - 145:10
misspoke [1] - 196:19
mistake [1] - 14:24
model [19] - 58:2, 58:18, 63:15, 63:18, 64:2, 65:11, 65:24, 66:17, 66:20, 67:2, 67:7, 92:18, 98:8, 125:8, 127:18, 128:1, 162:12, 198:10, 198:12
modeling [2] - 39:23, 57:8
models [5] - 40:15, 40:18, 62:13, 62:18, 102:14
momentarily [1] - 19:8
monitorships [1] - 39:10
months [3] - 138:15, 138:16, 139:13
morning [1] - 8:2
most [10] - 24:4, 56:5, 66:16, 66:19, 66:21, 75:5, 110:5, 115:15, 165:3, 198:6
motions [1] - 7:18
Mouse [1] - 181:16
mouth [2] - 76:12, 198:15
move [3] - 36:14, 91:6, 149:19
MR [208] - 6:23, 7:1, 7:2, 7:3, 7:6, 7:8, 7:10, 7:16, 7:21, 7:23, 9:11, 9:13, 9:15, 9:16, 9:18, 14:16, 15:2, 15:16, 15:18, 31:4, 32:13, 32:18, 32:22, 33:8, 33:12, 35:6, 47:20, 48:22, 49:16, 50:5, 51:19, 55:3, 59:12, 59:22, 61:11, 61:19, 63:21, 64:3, 64:17, 65:5, 65:13, 65:21, 66:4, 66:18, 67:6, 67:11, 69:8, 76:15, 76:19, 78:3, 78:9, 78:15, 79:1, 79:13, 81:24, 82:23, 83:24, 84:14, 86:14, 86:24, 87:16, 88:5, 89:24, 90:8, 90:18, 91:23, 92:3, 96:10, 96:18, 97:3, 97:11, 97:22, 98:9, 98:20, 99:4, 99:20, 100:9, 100:22, 101:3, 101:7, 101:12, 102:10, 102:23, 103:12, 103:21, 106:22, 115:6, 115:11, 116:17, 117:9, 117:20, 119:2, 119:19, 120:16, 121:14, 122:2, 124:3, 124:13, 124:15, 125:4, 125:14, 125:24, 126:3, 127:4, 127:20, 128:8, 128:10, 129:22, 130:5, 130:15, 130:22, 131:3, 131:10, 131:15, 131:18, 132:2, 132:7, 134:2, 134:18, 135:4, 135:21, 136:11, 136:21, 137:10, 138:1, 140:13, 143:21, 144:19, 146:1, 146:12, 147:20, 149:8, 149:19, 149:23, 150:1, 151:15, 152:6, 152:19, 154:5, 155:15, 155:23, 156:24, 158:2, 158:7, 159:2, 160:3, 160:8, 160:24, 161:10, 162:5, 162:19, 163:7, 163:14, 166:19, 167:2, 167:13, 167:16, 168:4, 168:20, 169:15, 170:4, 170:15, 172:4, 172:14, 172:22, 173:24, 174:13, 174:19, 175:4, 176:13, 177:9, 178:15, 180:13, 180:21, 181:3, 181:12, 182:1, 182:7, 183:17, 184:5, 184:17, 185:16, 185:20, 186:8, 187:14, 187:24, 188:3, 189:16, 190:10, 190:18, 191:20, 193:2, 194:5, 194:23, 197:22, 198:18, 198:20, 200:12, 202:2, 202:10, 203:10, 203:19, 204:6, 204:22, 205:7, 205:22, 206:8, 206:11
MS [4] - 15:5, 32:17, 32:21, 158:5
multiday [1] - 37:12
multiple [11] - 5:16, 33:15, 34:1, 44:24, 49:13, 54:16, 60:21, 113:20, 141:13, 148:15, 193:19
multiweek [1] - 37:14
MURPHY [2] - 2:11, 3:5
must [1] - 205:4

N

N-O-M-E [1] - 28:9
nail [1] - 116:1
name [51] - 5:14, 18:9, 25:2, 25:5, 26:21, 27:22, 28:6, 28:18, 28:20, 29:13, 29:24, 30:5, 31:24, 62:22, 72:4, 82:3, 83:19, 110:5, 110:15, 110:24, 148:16, 176:8, 177:8, 177:10, 177:11, 181:15, 187:17, 192:2, 192:14, 193:6, 193:15, 193:22, 193:23, 194:4, 194:6, 194:7, 194:10, 194:13, 194:22, 195:17, 195:22, 196:3, 196:4, 196:8, 197:3, 197:19, 199:8, 200:7, 200:13, 200:21, 200:24
NAME [2] - 209:2, 209:3
Name [1] - 194:18
named [1] - 207:5
names [17] - 18:11, 25:14, 29:14, 51:4, 70:20, 70:22, 72:14, 81:7, 94:24, 95:8, 153:13, 159:14, 179:6, 193:8, 196:14, 196:16, 196:19
NAN [6] - 29:17, 30:1, 31:3
national [1] - 56:20
native [1] - 61:15
nature [4] - 34:3, 52:5, 83:5, 95:9
necessarily [17] - 17:17, 19:5, 22:21, 47:21, 49:2, 59:15, 89:13, 103:2,

105:20, 116:4, 122:1, 124:23, 126:18, 126:19, 141:23, 192:17, 204:23
necessary [7] - 44:19, 45:18, 101:10, 117:18, 122:9, 122:10, 170:2
necktie [1] - 33:9
need [11] - 15:3, 32:14, 40:20, 54:14, 54:17, 132:11, 149:9, 158:18, 158:20, 167:19, 206:9
needed [4] - 48:14, 51:1, 55:12, 142:1
needs [1] - 168:9
negative [5] - 80:18, 166:12, 166:18, 167:1, 167:5
negatives [2] - 80:13, 204:15
neighborhood [1] - 43:16
net [42] - 8:19, 8:20, 47:18, 47:19, 48:4, 48:5, 49:8, 108:24, 133:3, 142:11, 142:17, 144:16, 145:11, 146:21, 146:22, 154:14, 154:17, 155:18, 156:15, 156:23, 157:2, 169:21, 169:22, 170:7, 171:14, 171:19, 172:6, 172:9, 176:5, 177:19, 183:23, 184:4, 184:7, 184:10, 191:5, 191:6
NET [2] - 1:17, 1:23
network [1] - 90:16
never [6] - 63:1, 119:7, 122:12, 162:23, 163:8, 199:1
Neves [17] - 18:4, 18:9, 85:12, 86:10, 86:22, 87:2, 87:4, 87:9, 93:7, 94:2, 94:4, 94:20, 94:22, 94:23, 95:1, 97:2, 157:21
next [2] - 74:8, 178:16
Nicole [1] - 10:20
night [1] - 37:9
nine [9] - 138:15, 156:22, 168:12, 196:7, 202:15
nines [2] - 168:15, 168:16
nome [1] - 197:5
none [2] - 156:22, 192:9
nonzero [3] - 146:3, 147:22, 147:24
norms [1] - 104:1
notarization [2] - 7:9, 7:11
Notary [4] - 2:9, 5:10, 207:3, 207:19
note [1] - 196:24
notebooks [3] - 45:4, 45:12, 45:19
noted [1] - 208:12
nothing [4] - 17:5, 17:19, 159:16, 200:2
Notice [2] - 4:9, 5:1
notice [2] - 8:3, 207:4
notify [2] - 51:1, 51:2
notwithstanding [1] - 50:11
null [3] - 25:12, 29:18, 70:19
null/one [1] - 73:1
number [60] - 18:6, 19:12, 22:19, 37:6, 43:8, 44:1, 44:8, 44:11, 44:14, 47:14, 47:16, 48:2, 48:17, 48:18, 49:23, 51:9, 53:13, 72:5, 72:14, 79:3, 79:8, 81:6, 81:9, 87:22, 87:23, 96:8, 106:17, 110:8, 113:22, 128:17, 128:18, 128:23, 129:5, 129:12, 133:1, 139:24, 146:3, 147:23, 147:24, 148:9, 149:9, 149:10, 159:19, 160:11, 161:1, 166:10, 166:17, 166:22, 166:24, 167:18, 167:20, 168:15, 176:6, 176:9, 177:14, 177:16, 177:18, 178:2
numbers [7] - 25:17, 70:23, 73:1, 144:17, 145:12, 176:21, 176:22

O

oath [1] - 207:6
objection [163] - 15:16, 31:4, 47:20, 48:22, 49:16, 50:5, 51:19, 55:3, 59:12, 59:22, 61:11, 61:19, 63:21, 64:3, 64:17, 65:5, 65:13, 65:21, 66:4, 66:18, 67:6, 69:8, 78:3, 78:9, 78:15, 79:1, 79:13, 82:23, 83:24, 84:14, 86:14, 86:24, 87:16, 89:24, 90:8, 90:18, 91:23, 92:3, 96:10, 96:18, 97:3, 97:11, 97:22, 98:9, 98:20, 99:4, 99:20, 100:9, 100:22, 101:3, 101:7, 101:12, 102:10, 102:23, 103:12, 103:21, 115:6, 115:11, 116:17, 117:9, 117:20, 119:2, 119:19, 120:16, 121:14, 122:2, 124:3, 124:13, 124:15, 125:4, 125:14, 125:24, 126:4, 127:4, 127:20, 128:8, 128:10, 129:22, 130:5, 130:15, 130:22, 131:3, 131:10, 134:2, 134:18, 135:4, 135:21, 136:11, 136:21, 137:10, 138:1, 140:13, 144:19, 146:1, 146:12, 147:20, 149:8, 151:15, 152:6, 152:19, 154:5, 155:15, 155:23, 156:24, 159:2, 160:3, 160:8, 160:24, 161:10, 162:5, 162:19, 163:7, 163:14, 166:19, 167:2, 167:13, 167:16, 168:4, 168:20, 169:15, 170:4, 170:15, 172:4, 172:14, 172:22, 173:24, 174:13, 174:19, 175:4, 176:13, 177:9, 180:13, 180:21, 181:3, 181:12, 182:1, 182:7, 183:17, 184:5, 184:17, 185:16, 185:20, 186:8, 187:14, 187:24, 188:3, 190:10, 190:18, 191:20, 193:2, 194:5, 194:23, 197:22, 200:12, 202:2, 202:10, 203:10, 203:19, 204:6, 204:22, 205:7, 205:22, 206:8
Objection [1] - 35:6
objectionable [1] - 192:23
objections [3] - 7:12, 7:13, 7:17
observant [1] - 33:11
obtained [1] - 16:21
obviously [5] - 16:10, 39:18, 76:1, 140:7, 196:23
occur [1] - 168:9
OF [12] - 1:3, 1:12, 1:13, 1:16, 1:17, 1:18, 1:19, 1:23, 1:23, 2:2, 207:1, 209:4
offered [2] - 36:4, 41:1
office [2] - 8:8, 18:13
officer [1] - 208:6
offices [1] - 2:11
often [4] - 49:3, 61:24, 162:2, 162:17
oftentimes [1] - 96:20
Oleo [2] - 110:16, 110:24
omit [1] - 17:8
once [3] - 29:8, 29:10
one [141] - 6:12, 6:15, 9:21, 20:15, 22:21, 23:9, 23:15, 24:20, 25:12, 26:17, 26:18, 29:1, 29:7, 31:10, 34:20, 35:17, 36:1, 40:6, 41:5, 42:7, 42:8, 43:10, 43:17, 47:17, 47:24, 48:1, 48:6, 50:4, 50:23, 51:6, 53:5, 55:24, 56:8, 56:11, 56:13, 56:20, 57:2, 57:15, 60:6, 61:1, 61:6, 62:19, 68:17, 69:12, 69:15, 70:20, 71:7, 72:10, 72:12, 77:2, 77:3, 77:5, 77:11, 80:22, 91:11, 91:13, 91:14, 91:18, 93:4, 94:1, 94:9, 101:16, 108:14, 111:3, 113:19, 113:21, 114:3, 118:19, 119:10, 119:24, 120:9, 120:14, 121:4, 121:22, 122:5, 124:22, 128:4, 132:18, 137:14, 137:15, 141:14, 144:1, 145:9, 146:5, 146:7, 146:8, 153:6, 154:20, 156:11, 156:21, 157:6, 157:7, 157:12, 158:8, 160:15, 164:17, 167:18, 168:5, 168:13, 171:12, 177:5, 177:6, 177:7, 180:4, 180:18, 180:22, 181:5, 181:13, 183:9, 183:10, 183:11, 183:15, 185:4, 185:21, 186:24, 188:1, 188:18, 188:21, 189:6, 190:4, 190:20, 190:24, 191:21, 193:11, 194:6, 196:12, 197:7, 201:11, 204:7
One [2] - 4:18, 131:24
one-by-one [1] - 168:5
One-Page [2] - 4:18, 131:24
ones [10] - 18:15, 18:16, 40:12, 53:3, 59:24, 136:19, 141:14, 142:18, 142:21, 142:23
onus [1] - 170:6
open [1] - 175:16
opened [2] - 49:13, 104:17
operated [1] - 25:21
opine [4] - 90:1, 127:6, 148:1, 188:1
opined [1] - 90:4
opinion [53] - 14:9, 17:6, 27:3, 40:22, 66:2, 75:15, 79:20, 82:5, 82:12, 87:14, 91:19, 95:14, 96:8, 96:15, 97:23, 98:14, 99:3, 109:6, 109:9, 114:1, 115:14, 122:7, 122:8, 128:20, 129:2, 129:15, 135:6,

145:19, 146:10, 147:14, 148:12, 152:2, 157:12, 157:17, 157:18, 160:20, 161:13, 161:16, 164:4, 167:8, 168:18, 170:2, 170:23, 174:11, 174:17, 174:20, 174:21, 180:11, 181:13, 181:20, 185:21, 197:19, 198:8
opinions [16] - 16:5, 16:9, 16:14, 16:22, 17:8, 17:13, 17:14, 17:18, 31:1, 40:24, 65:10, 65:18, 66:24, 72:23, 109:24, 150:10
opportunity [1] - 171:9
opposed [4] - 37:2, 49:22, 61:17, 72:11
options [3] - 106:2, 117:2, 122:4
order [7] - 18:18, 30:12, 136:13, 142:2, 192:7, 193:20, 201:4
organized [2] - 181:8, 182:19
original [5] - 4:23, 13:2, 13:14, 41:19, 120:3
originally [4] - 107:15, 107:23, 107:24, 113:4
originated [1] - 50:24
Oscar [1] - 171:17
otherwise [9] - 13:9, 22:6, 28:21, 54:22, 73:7, 100:24, 201:24, 203:4, 207:11
outcome [2] - 80:19, 207:11
outer [1] - 200:1
output [24] - 23:6, 26:16, 26:18, 26:19, 27:16, 27:19, 29:5, 29:8, 30:6, 31:6, 31:9, 31:10, 31:16, 31:18, 58:10, 71:16, 89:22, 101:2, 119:14, 120:2, 120:22, 141:23, 197:13, 204:19
overaggregate [1] - 205:19
overaggregated [3] - 152:4, 158:24, 159:4
overaggregation [20] - 97:10, 115:4, 124:24, 125:12, 125:17, 125:22, 152:15, 159:20, 164:5, 169:24, 201:14, 201:18, 201:20, 202:1, 202:8, 203:8, 204:4, 205:6, 205:12, 205:17
overaggregations [5] - 87:7, 87:15, 96:9, 96:17, 171:16
overall [9] - 62:16, 83:10, 85:2, 86:13, 98:17, 100:7, 127:22, 148:6, 149:3
overclustered [3] - 95:4, 95:15, 102:3
overclustering [12] - 98:23, 99:18, 103:9, 110:14, 110:21, 135:13, 135:19, 136:9, 136:14, 136:19, 159:22, 160:2
overclusters [2] - 98:7, 98:18
overinclusive [1] - 102:16
overlaid [2] - 141:10, 141:15
overlap [1] - 190:22
overly [1] - 205:20
overstatement [1] - 147:5
own [8] - 117:22, 135:6, 138:9, 147:3, 158:9, 181:23, 183:21, 192:2
owned [1] - 100:16
owner [2] - 56:20, 171:9

P

P.C [2] - 2:11, 3:5
p.m [3] - 131:23, 206:15
package [1] - 61:14
packages [1] - 46:6
page [20] - 56:24, 57:2, 67:20, 81:20, 81:23, 83:20, 85:16, 85:18, 143:4, 145:5, 175:12, 175:13, 175:16, 175:19, 179:5, 179:11, 180:3, 189:5, 195:14, 195:16
PAGE [1] - 209:6
Page [5] - 4:8, 4:14, 4:18, 67:14, 131:24
pages [4] - 1:1, 42:2, 42:3, 56:23
paid [3] - 43:17, 154:20, 154:23
pains [1] - 208:15
pairs [5] - 75:14, 195:20, 196:1, 196:3, 201:1
pairwise [8] - 92:24, 129:20, 130:13, 130:20, 199:7, 200:8, 200:16, 200:18
Pandas [4] - 45:24, 46:1, 46:2, 46:6
Papas [5] - 3:6, 14:17, 15:3, 67:12, 88:6
PAPAS [4] - 15:5, 32:17, 32:21, 158:5
Papas) [1] - 189:17
paper [4] - 16:12, 163:3, 163:10, 163:12
papers [1] - 202:5
paragraph [9] - 86:8, 180:3, 182:22, 187:8, 188:8, 189:6, 195:13, 195:16, 202:14
parameters [1] - 163:6
parsing [1] - 52:8
part [33] - 5:19, 12:2, 14:9, 32:7, 40:17, 45:18, 48:13, 51:5, 55:5, 59:1, 62:1, 72:15, 73:3, 92:22, 100:13, 108:15, 110:9, 113:5, 113:23, 115:12, 119:8, 127:12, 134:3, 149:20, 154:16, 155:21, 157:6, 191:2, 195:1, 199:6, 199:12, 199:20, 200:14
partially [1] - 46:7
participant [19] - 48:8, 48:9, 48:20, 49:2, 49:11, 49:12, 49:20, 49:22, 50:4, 99:11, 118:14, 120:13, 120:14, 121:17, 121:22, 137:18, 159:9, 172:8, 192:15
participants [10] - 48:6, 48:18, 50:3, 51:2, 106:13, 115:16, 137:20, 144:9, 144:15, 194:10
particular [11] - 20:15, 84:8, 85:8, 86:17, 110:3, 153:3, 153:4, 172:8, 183:8, 188:23, 190:14
particularly [2] - 25:1, 108:24
parties [2] - 207:10, 208:5
parts [1] - 39:5
past [2] - 46:4, 171:22
paste [1] - 19:6
patent [1] - 39:21
Patrick [2] - 9:22, 10:6
Paulo [2] - 3:12, 151:4
PAULO [1] - 1:22
penalties [1] - 208:15
pencil [3] - 163:4, 163:10, 163:12
people [26] - 9:24, 10:1, 51:16, 105:8, 111:10, 120:20, 126:13, 128:17, 128:18, 145:17, 146:19, 147:15, 151:6, 151:10, 151:13, 161:4, 161:5, 161:7, 161:8, 161:9, 168:14, 169:11, 169:12, 185:6, 192:2, 192:3
percent [3] - 81:17, 147:5, 171:20
percentage [4] - 21:21, 28:6, 30:6, 30:7
percentages [1] - 80:4
perfectly [2] - 203:13, 203:15
perform [9] - 38:23, 74:20, 75:12, 98:6, 98:17, 99:2, 122:17, 129:19, 137:7
performance [2] - 14:3, 14:5
performed [20] - 13:12, 48:14, 58:22, 65:15, 74:4, 116:6, 116:23, 117:4, 117:22, 119:3, 120:5, 130:24, 137:6, 139:2, 139:11, 168:23, 179:14, 179:15, 179:16, 204:12
performing [1] - 204:17
perjury [1] - 208:15
person [7] - 24:3, 100:16, 146:5, 146:21, 183:15, 186:3, 207:5
perspective [10] - 42:20, 66:21, 67:8, 69:11, 69:14, 73:15, 73:16, 74:7, 159:24, 176:2
pertinent [1] - 75:6
Peter [1] - 202:22
phase [3] - 73:21, 200:9, 200:11
phases [1] - 200:10
phone [3] - 176:9, 177:13, 178:2
phonetic [1] - 110:16
phrase [2] - 29:17, 199:11
pick [1] - 94:1
picture [3] - 21:18, 142:2, 145:5
piece [4] - 114:3, 163:3, 163:9, 163:11
pieces [1] - 23:16
pivot [2] - 175:24, 176:11
place [2] - 91:13, 195:1
placed [1] - 159:7
Plaintiff [3] - 1:14, 2:3, 3:2
plaintiff [1] - 1:20
plays [1] - 195:18
plenty [2] - 98:21, 99:5
plural [1] - 43:1
plus [3] - 15:14, 48:1, 178:4
point [12] - 29:16, 85:18, 135:12, 148:1, 168:13, 173:21, 185:9, 185:12, 186:20, 186:22, 190:21, 194:17
point-in-time [1] - 185:9
pointed [1] - 192:6
population [2] - 82:22, 162:4
portal [16] - 169:14, 169:16, 169:22, 171:8, 183:1, 183:7, 183:16, 184:4, 184:12, 184:13, 184:20, 184:21,

185:5, 185:8, 185:11
portion [11] - 21:20, 23:11, 23:12, 24:13, 32:15, 38:18, 39:9, 51:11, 151:2, 154:17, 194:9
portions [17] - 14:11, 21:3, 21:7, 21:12, 23:19, 24:5, 24:10, 24:12, 24:22, 25:18, 26:4, 31:13, 31:14, 31:19, 74:11, 74:12
position [4] - 99:16, 149:11, 174:21, 181:9
positive [2] - 80:16, 163:1
positives [6] - 80:13, 162:3, 162:18, 163:5, 163:23, 204:15
possibilities [1] - 175:6
possible [14] - 6:22, 33:16, 77:18, 77:20, 97:13, 98:2, 100:10, 100:18, 101:22, 103:5, 120:19, 126:5, 134:8, 135:9
posted [1] - 171:8
potential [22] - 27:8, 27:9, 27:12, 27:18, 42:10, 110:14, 114:11, 118:16, 132:23, 133:15, 134:7, 144:9, 159:17, 170:7, 173:11, 182:12, 190:22, 192:9, 196:5, 200:3, 200:4, 201:19
Potential [1] - 189:10
potentially [16] - 75:16, 96:3, 101:13, 101:14, 103:5, 105:13, 106:3, 137:5, 139:17, 148:10, 164:8, 165:1, 165:11, 172:15, 180:10, 183:7
power [2] - 37:19, 37:20
precision [7] - 88:9, 88:19, 89:4, 89:16, 162:1, 162:15, 163:23
premarked [1] - 5:5
prepared [1] - 10:16
preparing [2] - 10:12, 17:1
present [1] - 54:15
presented [1] - 114:19
presumed [1] - 106:12
pretty [3] - 125:19, 163:18, 168:15
prevent [1] - 39:14
previous [1] - 41:7
previously [5] - 20:23, 62:13, 127:16, 139:14, 143:18
price [7] - 50:24, 51:24, 53:18, 54:23, 56:9, 56:10, 60:1
primacy [4] - 165:17, 194:13, 194:22, 195:1
primarily [5] - 32:4, 40:2, 142:6, 193:15, 194:4
primary [7] - 9:24, 26:6, 26:10, 37:1, 42:16, 141:14, 142:18
principle [1] - 205:21
principles [1] - 147:19
printed [1] - 158:5
probabilistic [14] - 57:15, 59:1, 59:6, 59:11, 59:21, 60:3, 60:11, 60:15, 61:2, 61:8, 61:17, 61:21, 61:24, 62:12
probabilities [1] - 196:9
problem [3] - 134:4, 159:22, 201:16
problems [3] - 133:16, 133:19, 137:16
Procedure [2] - 2:5, 2:6
proceedings [2] - 5:16
process [72] - 22:1, 23:1, 23:5, 28:23, 50:14, 50:16, 50:19, 51:22, 52:1, 53:21, 53:22, 55:6, 62:16, 62:18, 65:16, 67:1, 67:24, 69:4, 69:10, 69:18, 69:23, 70:13, 71:21, 72:24, 73:8, 73:10, 74:14, 77:14, 91:21, 95:12, 96:22, 97:6, 98:1, 99:23, 101:1, 103:2, 112:22, 113:4, 113:6, 114:8, 116:14, 118:6, 118:8, 118:10, 119:1, 119:8, 120:4, 120:11, 121:3, 121:7, 125:2, 125:11, 127:12, 127:22, 130:3, 130:12, 130:16, 159:5, 169:17, 169:20, 169:21, 170:6, 173:13, 174:4, 193:19, 199:20, 200:14, 201:10, 201:16, 201:23
processes [2] - 68:11, 102:1
produced [1] - 26:15
production [2] - 5:8, 26:13
products [1] - 56:14
professional [1] - 42:23
Professional [3] - 2:8, 42:1, 59:19
program [3] - 39:12, 57:5, 166:3
programming [1] - 38:2
programs [2] - 36:5, 38:23
projects [3] - 46:24, 47:1, 47:3
proof [3] - 174:12, 174:22, 174:24
proper [1] - 161:14
properly [2] - 202:6, 203:6
property [9] - 10:11, 39:21, 56:18, 78:2, 78:7, 78:10, 78:12, 78:13, 78:17
prophecy [2] - 104:13, 162:9
proportional [3] - 166:11, 166:24, 167:4
proposition [2] - 192:12, 205:3
provide [5] - 8:8, 13:22, 13:24, 113:18, 170:1
provided [11] - 14:4, 26:14, 96:12, 98:21, 100:2, 107:10, 115:8, 133:9, 143:12, 160:15, 192:5
Public [4] - 2:9, 5:10, 207:3, 207:19
published [3] - 201:21, 202:5, 202:17
pull [2] - 189:2, 199:18
pulled [1] - 142:20
punctuation [1] - 25:16
purely [2] - 43:6, 74:6
purpose [14] - 58:22, 71:13, 75:11, 91:9, 91:12, 93:23, 105:22, 139:3, 165:21, 165:23, 180:14, 183:18, 205:24, 206:2
purposes [9] - 45:1, 75:23, 79:12, 94:3, 111:16, 145:23, 157:14, 181:7, 191:11
pursuant [4] - 2:4, 8:3, 57:4, 207:4
put [12] - 8:17, 16:8, 44:7, 62:21, 120:8, 121:16, 121:21, 141:2, 161:1, 170:6, 172:11, 187:16
putting [3] - 77:5, 77:11, 121:4
PUZZARINI [1] - 1:22
Puzzarini [1] - 3:12
Python [9] - 13:17, 29:18, 45:4, 45:8, 45:11, 45:13, 45:17, 45:23, 46:14

Q

QC [2] - 13:6, 13:7
quadruple [2] - 168:11, 168:15
qualify [2] - 104:24, 114:14
quality [9] - 84:6, 84:13, 85:2, 85:24, 86:5, 86:12, 94:7, 101:18, 104:10
quantification [3] - 114:24, 115:3, 137:2
quantified [2] - 72:13, 133:10
quantify [7] - 114:14, 133:2, 133:23, 136:7, 137:22, 161:6, 169:2
quantum [3] - 174:12, 174:21, 174:24
queried [1] - 140:15
queries [8] - 19:11, 19:12, 19:24, 93:18, 94:3, 139:1, 140:11, 141:3
query [7] - 19:3, 19:8, 93:17, 93:20, 93:21, 139:11, 140:16
questioned [1] - 95:7
questions [6] - 6:4, 73:11, 136:5, 150:7, 161:22, 191:12
quietly [1] - 145:12
quite [4] - 8:24, 35:16, 105:6, 199:22
quote [1] - 86:4

R

race [2] - 81:15, 85:11
radically [1] - 201:8
ran [11] - 21:3, 23:9, 24:1, 24:14, 25:19, 31:20, 31:21, 72:9, 74:11, 141:3
random [1] - 111:13
randomly [1] - 82:3
range [1] - 110:10
rarely [3] - 42:15, 44:6, 44:8
rate [39] - 86:22, 87:1, 89:5, 89:16, 89:21, 89:22, 90:2, 90:6, 97:9, 97:20, 98:2, 98:6, 98:17, 99:14, 99:17, 99:18, 100:8, 100:19, 102:2, 103:10, 103:19, 125:21, 132:24, 146:16, 148:6, 160:2, 160:21, 161:2, 166:12, 166:18, 167:1, 167:5, 168:11, 168:12, 168:14, 168:19, 168:22, 169:3, 169:9
rates [5] - 96:24, 115:5, 132:16, 132:17, 132:19
ratio [8] - 60:19, 60:24, 179:5, 179:6, 179:17, 197:5, 197:6
re [2] - 14:3, 14:5
Re [1] - 1:9
re-performance [2] - 14:3, 14:5
read [9] - 6:24, 63:22, 88:14, 90:13, 115:19, 115:21, 158:14, 191:3, 208:4
reading [8] - 10:13, 11:16, 21:17, 45:21, 56:19, 88:16, 145:12, 208:5
reads [1] - 195:16

real [11] - 101:20, 104:6, 104:7, 104:14, 126:9, 146:4, 162:8, 170:12, 170:13
real-world [8] - 104:6, 104:7, 104:14, 126:9, 146:4, 162:8, 170:12, 170:13
realistic [1] - 75:21
realized [1] - 80:21
really [17] - 44:23, 45:1, 45:9, 54:21, 55:7, 55:15, 55:17, 60:8, 62:7, 67:7, 82:10, 84:23, 98:16, 99:13, 142:8, 149:5, 166:4
reason [3] - 104:3, 165:24, 193:12
reasonable [2] - 147:3, 166:6
reasonably [2] - 39:13, 174:4
reasons [4] - 120:3, 155:20, 155:24, 208:7
rebate [1] - 57:5
Rebuttal [2] - 4:10, 5:3
received [16] - 12:23, 13:1, 14:7, 20:7, 20:14, 22:19, 22:20, 26:15, 31:18, 44:22, 44:24, 51:3, 107:23, 115:18, 182:15, 185:9
recent [2] - 56:5, 115:15
recess [1] - 131:22
Recess [2] - 76:20, 198:22
recognize [4] - 20:4, 20:6, 94:20, 189:21
recollect [1] - 55:19
recollecting [1] - 24:8
recollection [16] - 12:23, 22:13, 23:17, 24:19, 25:17, 28:5, 36:20, 43:14, 53:17, 83:17, 110:23, 123:2, 123:7, 123:10, 157:11, 204:8
record [7] - 28:22, 57:16, 59:1, 61:21, 106:22, 192:5, 207:7
records [16] - 18:7, 26:22, 28:14, 29:11, 30:7, 42:15, 44:8, 44:11, 44:14, 51:17, 51:23, 54:16, 145:18, 145:19, 147:16
RECROSS [1] - 4:5
REDIRECT [1] - 4:5
reduce [2] - 69:5, 75:20
reduced [1] - 207:7
refer [5] - 38:13, 39:3, 64:4, 106:16, 188:6
reference [2] - 85:8, 145:7
referenced [3] - 84:19, 143:18, 153:3
referred [6] - 54:23, 83:20, 84:24, 85:23, 150:12, 151:1
referring [18] - 9:4, 15:24, 19:10, 23:13, 39:4, 41:13, 41:16, 60:4, 61:12, 64:6, 85:20, 104:15, 104:20, 130:17, 153:7, 179:18, 183:3, 190:6
refers [3] - 70:11, 179:17, 195:19
reflect [1] - 132:9
reflects [1] - 181:17
regard [2] - 85:21, 85:22
regardless [6] - 44:13, 122:21, 167:20, 191:18, 200:7, 206:4
regards [2] - 70:6, 110:17
Registered [3] - 2:7, 2:8, 207:3
reimbursement [1] - 57:4
relate [1] - 54:16
related [7] - 24:15, 25:1, 35:12, 36:24, 38:10, 51:3, 56:13
relates [9] - 39:8, 39:19, 79:21, 83:10, 89:5, 108:13, 114:7, 167:12, 168:2
relating [1] - 36:21
relative [8] - 65:1, 65:3, 65:24, 66:14, 67:1, 135:12, 207:9, 207:10
relevance [2] - 193:7, 196:22
relevant [7] - 16:6, 16:15, 17:9, 17:13, 66:2, 75:6, 148:5
reliability [3] - 99:24, 127:23, 198:11
reliable [9] - 137:17, 157:14, 163:2, 166:6, 166:8, 174:6, 192:17, 198:6, 198:12
reliance [1] - 70:4
relied [8] - 14:8, 32:5, 42:20, 58:14, 85:1, 154:24, 156:22, 157:4
relies [5] - 192:13, 193:16, 194:3, 194:7, 196:24
rely [5] - 12:15, 12:18, 31:2, 119:11, 149:3
relying [2] - 32:11, 170:11
remember [20] - 18:3, 18:6, 20:11, 20:15, 22:2, 28:2, 30:5, 30:20, 32:2, 34:10, 35:10, 35:24, 37:18, 43:5, 52:15, 52:23, 53:2, 81:17, 142:21, 154:23
remind [1] - 122:14
remotely [1] - 106:15
remove [4] - 70:18, 71:7, 71:24, 72:4
removed [2] - 71:11, 72:10
removing [3] - 25:9, 71:6, 73:1
REP [1] - 177:12
rep [2] - 177:5, 177:13
rep_end [1] - 177:12
rep_id [18] - 28:7, 29:5, 29:7, 29:9, 29:11, 31:10, 55:11, 142:13, 154:24, 155:3, 155:8, 155:12, 155:17, 156:4, 156:11, 157:10, 177:3
rep_id's [1] - 176:6
rep_nome [5] - 28:8, 28:10, 29:1, 176:23, 190:9
reperformed [1] - 157:11
rephrase [1] - 6:6
Report [7] - 4:11, 4:14, 4:16, 5:4, 67:15, 94:11, 94:16
report [89] - 9:4, 9:6, 9:8, 9:12, 9:14, 9:17, 10:13, 10:15, 11:16, 12:17, 12:21, 13:14, 14:1, 14:2, 14:20, 15:1, 15:11, 15:15, 15:22, 16:3, 16:5, 16:14, 16:23, 17:9, 17:15, 17:24, 18:20, 21:18, 25:3, 28:1, 31:2, 32:16, 32:20, 43:15, 63:23, 64:5, 64:7, 67:21, 70:3, 72:2, 72:7, 81:4, 81:11, 83:21, 84:20, 85:7, 86:20, 87:12, 87:20, 88:1, 88:14, 88:17, 90:14, 90:22, 93:6, 93:14, 94:19, 101:17, 106:17, 120:4, 122:20, 132:17, 133:1, 133:8, 139:22, 142:22, 142:24, 143:11, 145:14, 150:6, 150:11, 155:21, 156:1, 157:23, 157:24, 158:6, 175:13, 178:14, 179:10, 188:17, 189:22, 192:6, 192:8, 193:15, 193:20, 194:15, 195:7
Reporter [3] - 2:8, 2:9, 207:3
reporter [2] - 4:23, 207:22
REPORTING [1] - 2:21
reports [3] - 8:16, 9:1, 12:12
represent [5] - 5:15, 49:11, 67:20, 110:21, 150:4
representante [1] - 143:17
representative [9] - 84:5, 84:9, 84:13, 85:1, 85:24, 86:12, 94:6, 148:24, 149:5
Representative [1] - 3:11
REPRESENTATIVE [1] - 1:16
Representatives [1] - 3:12
REPRESENTATIVES [1] - 1:23
representing [1] - 3:2
Representing [1] - 3:11
represents [1] - 29:6
reproduction [1] - 207:21
require [16] - 47:3, 50:18, 98:13, 119:16, 120:15, 121:13, 122:1, 127:9, 127:11, 128:6, 129:24, 130:12, 135:5, 135:23, 136:15, 161:12
required [10] - 34:8, 50:15, 75:19, 77:1, 83:9, 91:21, 116:13, 117:12, 142:4, 167:9
requirement [2] - 128:12, 193:17
requires [7] - 74:20, 100:21, 100:23, 100:24, 117:8, 122:17, 193:8
reran [1] - 14:11
research [2] - 122:4, 202:5
reserve [1] - 7:17
reside [1] - 114:24
resolved [1] - 180:18
respect [41] - 9:8, 12:17, 13:24, 15:11, 22:23, 23:4, 24:4, 25:24, 46:14, 56:9, 65:10, 65:18, 65:23, 66:14, 66:24, 69:24, 70:13, 72:23, 73:5, 73:19, 83:22, 97:17, 109:23, 109:24, 112:9, 114:1, 115:13, 134:11, 135:3, 135:16, 136:7, 140:2, 145:15, 145:20, 147:12, 147:15, 150:24, 152:9, 156:21, 191:13, 207:22
respects [1] - 157:5
response [1] - 13:14
result [12] - 19:17, 114:2, 114:6, 115:2, 124:23, 185:11, 185:17, 201:24, 202:7, 203:7, 204:3, 205:5
resulting [1] - 143:14
results [3] - 121:11, 137:17, 197:20
retain [1] - 44:19
retained [2] - 4:23, 56:20
review [12] - 8:16, 13:5, 39:14, 114:23, 115:15, 116:12, 136:24, 138:11, 139:2, 155:9,

198:11, 203:22
reviewed [17] - 13:13, 14:10, 21:2, 72:7, 80:3, 81:7, 82:9, 86:17, 115:20, 129:1, 138:22, 138:24, 139:6, 140:5, 140:7, 188:21, 205:1
reviewing [6] - 20:12, 20:15, 39:15, 82:2, 138:3, 138:7
right-hand [2] - 176:21, 177:3
rise [1] - 61:16
risk [2] - 102:5, 102:7
Road [1] - 2:22
Role [1] - 194:18
role [1] - 195:17
RONA [179] - 3:13, 7:1, 7:3, 7:6, 7:8, 7:16, 7:23, 9:11, 9:15, 9:18, 15:16, 31:4, 33:8, 35:6, 47:20, 48:22, 49:16, 50:5, 51:19, 55:3, 59:12, 59:22, 61:11, 61:19, 63:21, 64:3, 64:17, 65:5, 65:13, 65:21, 66:4, 66:18, 67:6, 69:8, 78:3, 78:9, 78:15, 79:1, 79:13, 81:24, 82:23, 83:24, 84:14, 86:14, 86:24, 87:16, 89:24, 90:8, 90:18, 91:23, 92:3, 96:10, 96:18, 97:3, 97:11, 97:22, 98:9, 98:20, 99:4, 99:20, 100:9, 100:22, 101:3, 101:7, 101:12, 102:10, 102:23, 103:12, 103:21, 106:22, 115:6, 115:11, 116:17, 117:9, 117:20, 119:2, 119:19, 120:16, 121:14, 122:2, 124:13, 124:15, 125:4, 125:14, 125:24, 126:3, 127:4, 127:20, 128:8, 128:10, 129:22, 130:5, 130:15, 130:22, 131:3, 131:10, 131:18, 134:2, 134:18, 135:4, 135:21, 136:11, 136:21, 137:10, 138:1, 140:13, 144:19, 146:1, 146:12, 147:20, 149:8, 151:15, 152:6, 152:19, 154:5, 155:15, 155:23, 156:24, 158:2, 159:2, 160:3, 160:8, 160:24, 161:10, 162:5, 162:19, 163:7, 163:14, 166:19, 167:2, 167:13, 167:16, 168:4, 168:20, 169:15, 170:4, 170:15, 172:4, 172:14, 172:22, 173:24, 174:13, 174:19, 175:4, 176:13, 177:9, 180:13, 180:21, 181:3, 181:12, 182:1, 182:7, 183:17, 184:5, 184:17, 185:16, 185:20, 186:8, 187:14, 187:24, 188:3, 190:10, 190:18, 191:20, 193:2, 194:5, 194:23, 197:22, 198:18, 200:12, 202:2, 202:10, 203:10, 203:19, 204:6, 204:22, 205:7, 205:22, 206:8
Rona [1] - 3:14
Rona's [1] - 8:7
roughly [1] - 140:11
Rowe [1] - 2:22
rows [2] - 43:8, 158:22
Rule [1] - 208:1
rule [2] - 101:6, 102:20
Rules [2] - 2:5, 2:6
rules [4] - 6:3, 102:16, 102:17, 205:18
run [22] - 21:7, 21:11, 21:21, 22:4, 22:10, 22:22, 23:3, 23:20, 24:5, 24:23, 26:4, 31:13, 31:14, 31:16, 71:10, 74:9, 74:13, 75:21, 75:23, 102:5, 102:7, 141:5
running [4] - 25:18, 94:3, 140:16
Ryan [4] - 9:22, 10:6, 12:16, 12:20

S

S-Q-L [1] - 13:20
S-T-E-D-M-A-N [1] - 10:20
S-U-N-T-E-R [1] - 57:18
sample [15] - 78:1, 78:6, 78:19, 78:23, 79:11, 79:17, 79:23, 82:7, 82:13, 82:21, 83:9, 83:22, 83:23, 111:11, 149:6
sampled [1] - 129:14
samples [1] - 21:11
sampling [4] - 83:19, 145:22, 147:13, 147:18
SANDRO [1] - 1:22
Sandro [2] - 3:12, 151:4
Saryas [4] - 183:9, 183:21, 184:23, 185:18
SARYAS [1] - 184:24
satisfactorily [1] - 5:8
save [1] - 150:21
saw [8] - 26:20, 29:9, 29:11, 84:7, 109:6, 133:15, 168:6, 201:11
scale [1] - 41:2
scanned [1] - 4:23
science [26] - 10:2, 10:7, 12:4, 12:7, 13:22, 34:5, 34:7, 34:11, 34:16, 34:18, 34:23, 34:24, 35:3, 36:22, 38:10, 103:11, 129:19, 130:3, 130:10, 131:2, 131:5, 182:4, 182:9, 182:10, 205:2
scientist [1] - 11:21
scientists [4] - 10:5, 10:23, 11:2, 12:10
scope [4] - 8:12, 8:14, 73:3, 144:6
score [1] - 90:12
scores [2] - 179:19, 179:20
screen [1] - 19:4
seal [1] - 207:13
search [4] - 93:12, 93:15, 104:5, 140:16
searches [1] - 141:18
searching [1] - 93:22
second [7] - 24:21, 42:2, 56:12, 57:2, 143:6, 182:22, 189:6
secondary [2] - 154:2, 192:24
section [3] - 85:15, 194:17, 195:19
Section [1] - 195:19
see [29] - 21:12, 24:24, 25:22, 26:1, 27:6, 44:21, 67:24, 69:9, 69:13, 69:20, 81:14, 85:4, 141:15, 142:2, 156:4, 156:10, 156:14, 161:3, 162:3, 162:17, 176:20, 178:15, 179:7, 183:2, 184:1, 188:16, 189:10, 200:22, 202:17
seeing [1] - 30:5
seek [1] - 166:4
seem [1] - 147:3
Select [2] - 41:24, 59:19
select [1] - 42:23
selected [3] - 18:16, 82:3, 84:8
selection [2] - 69:13, 190:15
selective [1] - 18:15
self [2] - 104:13, 162:9
self-fulfilling [2] - 104:13, 162:9
semester [2] - 36:11, 36:13
sense [5] - 83:2, 90:16, 90:19, 111:24, 112:5
sentence [3] - 182:22, 189:6, 189:12
separate [3] - 99:7, 144:15, 187:7
separately [3] - 13:11, 36:8, 182:18
series [3] - 6:3, 48:17, 171:7
servers [1] - 13:4
set [14] - 20:19, 20:21, 29:8, 78:23, 100:15, 100:21, 119:17, 119:21, 119:22, 139:6, 155:21, 175:19, 196:1, 207:12
sets [2] - 24:9, 43:1
setting [2] - 20:17, 25:11
settlement [2] - 191:2, 191:11
seven [2] - 143:4, 145:9
several [1] - 37:24
Shal [6] - 180:4, 180:19, 182:23, 183:10, 186:6, 187:11
SHAL [1] - 180:4
shall [3] - 208:3, 208:4, 208:6
share [4] - 99:7, 159:12, 159:15, 159:16
sheet [1] - 208:12
SHEET [1] - 209:1
shirt [1] - 33:13
shirts [1] - 33:15
short [1] - 121:20
shot [1] - 163:12
show [16] - 19:4, 20:2, 54:18, 55:10, 99:6, 99:23, 141:20, 143:2, 164:12, 165:2, 171:15, 175:17, 175:19, 180:16, 183:18, 194:15
showed [5] - 134:17, 135:2, 135:18, 136:9, 159:20
showing [5] - 144:12, 176:23, 178:21, 178:22, 190:15
shown [5] - 8:4, 152:23, 178:13, 190:11, 190:12
shows [2] - 156:2, 162:15
sic [2] - 194:3, 194:21
side [9] - 10:11, 11:14, 40:14, 120:9, 121:4, 133:3, 176:6, 176:20, 179:8
SIG [6] - 13:2, 13:8, 41:18, 43:13, 44:16
sigma [1] - 168:16
sign [2] - 6:24, 175:2
Signature [1] - 208:19
signed [2] - 12:13, 187:1
Signed [1] - 208:15
significance [3] - 186:23, 187:2, 193:6
significant [11] - 26:20, 30:11, 128:20, 129:2, 129:8, 129:15, 134:15, 137:19, 145:22, 161:7, 190:21
significantly [1] -

161:4
Signing [1] - 208:1
similar [18] - 11:13, 51:7, 83:6, 159:14, 169:20, 170:5, 188:10, 189:9, 193:9, 193:22, 196:16, 196:20, 197:18, 199:24, 200:1
similarities [1] - 180:15
similarity [5] - 99:8, 179:19, 200:7, 200:20, 200:24
simple [4] - 60:16, 60:18, 175:18, 192:16
simply [16] - 21:15, 44:9, 51:16, 71:4, 71:17, 100:17, 101:22, 105:18, 105:21, 106:8, 144:24, 162:9, 183:18, 186:20, 198:11, 202:23
single [26] - 42:14, 49:11, 49:12, 49:14, 49:20, 54:16, 66:10, 84:4, 84:11, 84:24, 85:22, 120:12, 133:12, 164:4, 164:10, 165:5, 165:9, 181:11, 181:24, 183:2, 184:23, 186:1, 186:14, 186:15, 192:15
singular [3] - 20:6, 139:3, 196:13
sit [19] - 24:11, 55:20, 64:8, 64:14, 67:9, 89:8, 99:15, 116:21, 117:24, 118:9, 118:18, 118:23, 119:15, 120:10, 126:24, 131:13, 160:21, 174:10, 174:17
sitting [8] - 17:21, 44:10, 52:15, 68:16, 72:20, 77:21, 123:4, 123:21
situation [1] - 125:6
situations [2] - 153:10, 153:11
six [4] - 34:13, 145:9, 168:16
size [32] - 41:9, 42:24, 43:1, 43:6, 43:7, 43:9, 44:15, 53:14, 71:21, 78:1, 78:6, 78:19, 78:23, 79:4, 79:15, 79:17, 79:23, 80:24, 81:3, 82:7, 82:13, 82:21, 82:22, 83:9, 83:10, 83:22, 83:23, 147:18, 149:6, 164:13, 164:15, 166:23
sizes [1] - 79:11
Skidmore [3] - 33:18, 34:12, 34:17
skin [1] - 92:2
small [5] - 14:10, 21:2, 21:3, 149:10, 149:12
software [1] - 61:14
Soha [11] - 10:18, 11:5, 13:22, 14:2, 20:24, 23:24, 24:3, 24:14, 32:6, 32:10, 58:4
SOHA [1] - 10:19
solely [1] - 58:16
solid [2] - 143:7, 143:15
solution [1] - 126:7
someone [10] - 110:18, 110:23, 147:2, 160:10, 170:18, 175:6, 181:15, 186:23, 187:16, 192:13
sometimes [4] - 26:23, 44:7, 153:12, 153:13
somewhat [1] - 172:5
somewhere [1] - 187:18
sorry [31] - 9:11, 13:10, 14:21, 14:22, 15:18, 31:17, 31:18, 41:12, 52:19, 54:4, 54:6, 56:22, 61:13, 62:5, 76:7, 76:11, 85:21, 86:7, 96:11, 106:10, 133:22, 141:10, 145:11, 155:1, 156:8, 166:20, 167:5, 171:24, 176:24, 184:18
sort [8] - 36:2, 37:18, 58:10, 58:17, 75:8, 94:8, 111:12, 138:17
Sosa [1] - 171:17
sought [4] - 112:24, 117:21, 118:15, 123:13
sounds [2] - 9:19, 54:20
source [12] - 41:19, 68:4, 141:8, 141:12, 142:1, 142:4, 142:5, 166:2, 167:11, 168:1, 201:17, 201:19
sources [1] - 192:9
spaces [1] - 25:9
spanned [1] - 37:15
speaking [10] - 58:18, 77:16, 83:8, 88:22, 102:9, 102:14, 121:8, 164:22, 166:10, 168:14
speaks [2] - 188:14, 191:21
Special [1] - 194:18
special [1] - 195:17
specific [29] - 14:6, 18:2, 24:9, 27:24, 44:14, 47:7, 53:16, 62:19, 62:23, 66:12, 79:24, 81:2, 86:3, 88:23, 89:13, 103:16, 106:11, 122:20, 122:22, 123:21, 130:7, 130:8, 132:22, 133:4, 133:6, 164:13, 194:20, 203:20, 203:21
specifically [23] - 13:24, 24:1, 30:5, 53:3, 59:18, 81:5, 81:7, 82:8, 85:19, 96:11, 96:12, 107:12, 111:20, 111:23, 114:5, 123:4, 131:11, 131:12, 133:2, 152:8, 160:11, 163:17, 188:19
speculating [1] - 187:22
Splinc [2] - 46:20, 46:21
split [1] - 52:8
SQL [16] - 13:18, 13:19, 36:24, 37:4, 37:18, 37:20, 37:23, 38:2, 38:4, 38:5, 38:6, 45:23, 46:19, 112:23, 140:17, 141:1
SS [1] - 207:1
stand [2] - 192:11, 205:2
standard [6] - 79:3, 100:14, 128:2, 128:7, 129:19, 129:24
stands [1] - 63:11
start [6] - 6:17, 39:2, 69:19, 130:19, 155:6, 175:15
starting [1] - 198:15
state [1] - 204:1
State [3] - 2:11, 3:6, 3:14
statement [14] - 16:16, 19:1, 49:15, 99:21, 102:18, 118:20, 124:14, 125:3, 125:5, 136:13, 140:22, 140:23, 205:9, 208:7
statements [3] - 25:3, 56:2, 152:18
states [2] - 116:8, 193:19
STATES [1] - 1:3
statistic [2] - 44:9, 98:11
statistical [10] - 12:11, 78:24, 88:9, 88:10, 100:6, 100:19, 104:1, 111:12, 114:15, 147:19
statistically [7] - 128:20, 129:1, 129:8, 129:14, 145:22, 147:17, 149:4
statistics [12] - 11:9, 35:21, 36:2, 36:4, 36:5, 36:6, 36:7, 36:8, 36:10, 78:21, 78:24, 113:8
Stedman [2] - 10:20, 11:12
stenographer [1] - 6:18
STENOGRAPHER [7] - 25:6, 54:4, 76:7, 138:5, 143:20, 149:24, 156:6
step [48] - 14:12, 23:1, 23:5, 23:6, 24:18, 24:21, 27:7, 31:15, 31:24, 66:6, 66:10, 68:17, 69:12, 69:15, 73:5, 73:6, 74:4, 74:8, 80:22, 91:21, 116:7, 116:13, 116:16, 117:5, 154:10, 154:15, 154:16, 155:1, 155:5, 156:11, 157:6, 157:7, 192:24, 195:7, 196:6, 196:11, 196:13, 196:23, 197:1, 197:12, 197:13, 197:14, 199:1, 199:6, 199:17
step-by-step [1] - 14:12
STEPHEN [2] - 1:12, 1:18
Stephen [2] - 2:4, 3:2
Steps [2] - 4:15, 67:15
steps [41] - 21:24, 22:4, 22:5, 22:10, 22:14, 22:15, 22:17, 22:21, 23:4, 23:15, 24:15, 24:17, 24:19, 24:20, 25:24, 26:7, 26:10, 26:11, 32:9, 67:13, 67:23, 68:2, 68:7, 70:10, 92:21, 107:3, 109:15, 145:2, 151:17, 152:16, 156:21, 156:22, 157:8, 182:17, 186:10, 195:5, 196:7, 197:10, 199:2, 200:6, 200:8
still [5] - 44:18, 47:11, 47:24, 184:7, 200:13
Stoneham [1] - 2:22
StoneTurn [2] - 8:22, 18:18
StoneTurn's [1] - 39:9
stopped [1] - 121:19
Street [3] - 2:12, 3:6, 3:14
strict [1] - 205:20
strictly [1] - 5:19
strike [13] - 7:19, 30:24, 31:15, 35:19, 57:7, 57:9, 112:13, 115:23, 122:15, 127:8, 127:15, 149:20, 201:6
string [2] - 25:4, 29:16
strings [3] - 93:17, 93:20, 93:21
strong [1] - 110:20
stronger [1] - 45:20
structured [1] - 182:15
struggling [1] - 59:17
studied [4] - 64:11, 64:13, 65:8, 91:5
studies [14] - 201:21, 202:4, 202:5, 202:12, 202:13, 202:17, 202:18,

203:1, 203:3, 204:9, 205:10, 205:11, 205:14
study [4] - 33:21, 36:21, 203:13, 203:16
stuff [1] - 37:19
subcategories [1] - 56:18
subject [5] - 37:23, 79:21, 79:22, 167:22, 174:17
sublinear [1] - 83:16
Submission [1] - 208:1
submit [1] - 185:7
submitted [9] - 156:1, 183:1, 183:10, 183:15, 184:24, 185:18, 186:4, 186:15, 208:3
subsequent [6] - 35:18, 75:13, 192:20, 196:23, 197:10, 197:12
subsequently [2] - 16:20, 113:5
subset [2] - 19:4, 42:19
subsidiary [1] - 26:10
substance [1] - 208:5
substantial [6] - 139:24, 193:22, 196:2, 196:4, 199:9, 201:5
substantially [4] - 193:9, 195:22, 199:9, 199:10
substantively [2] - 17:6, 17:19
sufficiency [3] - 40:3, 40:6, 40:9
sufficient [11] - 105:21, 126:18, 146:8, 146:11, 172:8, 172:13, 173:15, 173:20, 173:23, 174:11, 175:1
sufficiently [3] - 148:24, 170:24, 175:2
suggest [4] - 80:20, 201:22, 202:6, 203:4
suggested [2] - 74:3, 190:24
suggesting [5] - 167:21, 180:8, 190:16, 190:20, 191:15
suggestion [1] - 102:24
Suite [2] - 3:6, 3:14
sum [2] - 43:5, 177:19
summarize [1] - 55:9
summarized [2] - 86:20, 189:7
summary [18] - 17:17, 26:18, 27:15, 27:16, 54:15, 54:19, 142:13, 154:24, 155:3, 155:8, 155:12, 155:17, 156:4, 156:11, 157:10, 175:21, 175:22, 176:15
summing [3] - 54:3, 54:7, 176:4
Sunter [18] - 57:11, 57:14, 57:20, 58:1, 58:6, 58:9, 58:18, 58:19, 58:24, 63:1, 63:18, 63:20, 64:15, 64:16, 92:18, 116:15, 127:7, 127:18
Supplemental [2] - 4:10, 5:3
support [7] - 12:10, 13:14, 13:23, 109:10, 110:6, 110:7, 204:24
supported [4] - 70:16, 108:4, 109:13, 110:1
supportive [1] - 107:17
suppose [2] - 161:24, 172:9
supposed [2] - 57:1, 97:15
suspicion [1] - 81:21
swath [1] - 140:8
sworn [3] - 5:9, 152:12, 207:6
synonymously [1] - 73:24
system [2] - 61:17, 61:18
systems [1] - 40:4

T

table [39] - 18:4, 18:5, 18:7, 19:7, 26:18, 26:19, 27:15, 27:19, 27:21, 31:10, 31:16, 31:19, 42:14, 67:21, 141:14, 142:7, 142:8, 142:13, 142:14, 142:17, 143:18, 143:24, 144:1, 154:19, 154:24, 155:3, 155:8, 155:9, 155:12, 155:17, 156:2, 156:11, 157:10, 175:24, 176:11, 178:22, 179:9
Tables [2] - 143:9, 143:16
tables [29] - 14:20, 15:12, 15:13, 15:14, 15:21, 15:24, 18:1, 18:12, 18:14, 18:18, 18:23, 18:24, 19:22, 19:23, 26:14, 26:17, 42:17, 45:1, 141:13, 142:4, 142:5, 142:7, 142:16, 142:19, 143:7, 143:9, 143:14, 153:22, 153:23
targeted [2] - 111:14, 111:15
technical [1] - 12:9
technique [2] - 129:19, 129:23
techniques [3] - 60:22, 93:12, 93:16
telephone [1] - 2:23
Telex [13] - 5:16, 43:1, 43:21, 44:16, 47:4, 48:20, 51:15, 57:24, 59:10, 119:14, 120:12, 165:4, 165:14
TELEXFREE [9] - 1:10, 1:11, 1:13, 1:13, 1:19, 1:19
TelexFree [11] - 3:3, 3:3, 81:9, 144:3, 144:8, 144:12, 155:4, 157:16, 194:19, 195:18
ten [13] - 21:24, 22:4, 22:5, 24:19, 34:13, 67:13, 67:23, 154:10, 154:15, 154:17, 155:5, 195:5, 196:12
ten-step [1] - 154:10
tend [3] - 44:18, 46:18, 168:14
tension [1] - 205:12
term [6] - 35:2, 41:18, 49:21, 57:10, 118:4, 137:9
terminology [8] - 108:12, 112:24, 119:12, 120:24, 122:20, 122:22, 138:14, 199:16
terms [24] - 8:19, 11:1, 19:24, 26:16, 42:20, 43:3, 52:4, 66:21, 69:16, 73:24, 81:3, 88:9, 88:15, 88:17, 89:13, 130:1, 140:5, 148:7, 158:10, 159:16, 160:11, 176:5, 197:10, 203:12
test [1] - 9:10
tested [2] - 134:22, 174:9
testified [1] - 5:10
testimony [5] - 5:24, 152:13, 207:8, 208:3, 208:11
testing [4] - 39:16, 40:3, 101:19, 127:22
text [1] - 204:9
TF [4] - 62:1, 62:8, 63:9, 63:11
TF-IDF [4] - 62:1, 62:8, 63:9, 63:11
THE [18] - 1:12, 1:18, 7:5, 7:7, 25:6, 33:10, 33:14, 54:4, 76:7, 76:11, 76:17, 131:20, 138:5, 143:20, 149:24, 156:6, 198:14, 206:13
themselves [4] - 95:21, 96:1, 96:20, 141:19
theory [1] - 90:17
therefore [4] - 67:3, 67:5, 103:4, 192:22
therein [2] - 155:19, 183:12
thereof [1] - 62:3
thereupon [1] - 207:6
thinking [7] - 11:2, 40:5, 123:20, 139:17, 139:18, 169:8, 169:9
third [4] - 42:2, 56:8, 57:2, 179:4
thirty [1] - 7:2
thoughts [1] - 16:10
thousand [7] - 12:20, 14:19, 138:18, 138:19, 138:22, 139:7, 139:16
thousands [1] - 30:16
three [13] - 42:3, 42:7, 52:17, 52:19, 52:21, 71:8, 71:11, 72:1, 72:19, 73:5, 105:7, 151:6, 151:19
thresholds [1] - 197:4
throughout [1] - 197:15
tie [1] - 113:2
tied [1] - 186:14
tier [1] - 113:23
titled [1] - 26:18
today [27] - 17:22, 24:11, 44:10, 52:15, 55:20, 64:8, 64:14, 67:9, 68:16, 72:21, 77:21, 89:8, 99:16, 116:21, 117:24, 118:9, 118:18, 118:23, 119:15, 120:10, 123:4, 123:22, 126:24, 131:13, 160:21, 174:10, 174:18
together [20] - 15:12, 15:13, 15:21, 24:7, 52:10, 92:24, 93:1, 118:17, 121:17, 144:4, 181:24, 186:14, 195:23, 198:5, 199:12, 199:20, 200:3, 200:9, 200:19, 201:9
tolerances [1] - 204:14
took [7] - 14:10, 34:11, 36:8, 36:23, 37:24, 38:7, 38:8
tool [6] - 118:3, 118:5, 118:6, 124:1, 124:4, 130:9
tools [11] - 52:12, 52:24, 54:24, 55:2, 60:15, 60:24, 61:7, 61:10, 62:12, 62:13, 62:17
toolset [1] - 131:2
top [2] - 141:2, 144:12
topic [1] - 157:19
total [4] - 43:5, 69:5, 81:9, 191:6
totality [2] - 159:4, 169:8
totally [1] - 167:3
totals [1] - 113:22
towards [2] - 40:2, 85:14
track [2] - 165:7, 173:5
trade [1] - 102:15
trade-off [1] - 102:15
trademark [1] - 39:22
trailing [1] - 25:9

trained [1] - 10:2
training [3] - 11:20, 11:22, 12:1
transaction [1] - 39:16
transactional [1] - 41:17
transactions [4] - 80:2, 129:6, 144:3
transcribed [1] - 208:3
transcript [4] - 6:22, 34:14, 207:21, 208:11
transcripts [1] - 115:16
transfer [1] - 143:24
transferencia [1] - 143:23
transformation [1] - 69:13
transitive [53] - 76:22, 77:6, 77:10, 77:13, 91:6, 91:8, 91:9, 91:20, 92:1, 92:17, 102:1, 102:14, 102:16, 102:17, 103:2, 103:7, 116:2, 116:6, 117:3, 117:5, 117:8, 117:10, 117:14, 117:16, 117:17, 118:10, 118:24, 119:7, 119:12, 119:16, 120:6, 120:15, 120:24, 121:2, 121:6, 121:13, 122:1, 122:10, 122:13, 122:18, 122:23, 123:16, 124:10, 124:20, 124:22, 125:8, 125:11, 129:21, 130:13, 130:18, 130:21, 199:15, 199:22
transitively [1] - 102:21
travel [1] - 56:21
tree [1] - 144:11
trial [2] - 5:24, 7:19
tried [2] - 132:16, 175:16
trimming [3] - 25:5, 25:8, 31:21
true [20] - 12:14, 25:23, 46:14, 47:22, 78:13, 80:18, 101:21, 104:16, 104:19, 104:24, 124:18, 127:10, 133:20, 142:9, 154:18, 177:10, 180:24, 198:1, 207:7, 208:11
truly [6] - 5:23, 80:20, 101:20, 148:7, 162:12, 162:21
Trustee [1] - 3:3
trustee [9] - 5:15, 8:18, 169:13, 172:21, 190:23, 191:2, 191:18, 192:1, 192:6
TRUSTEE [2] - 1:12, 1:18
trustee's [1] - 191:9
truth [1] - 128:2
truthfully [1] - 6:10
truthfulness [1] - 151:18
try [15] - 6:6, 6:14, 23:24, 52:8, 59:3, 67:9, 75:12, 93:1, 116:1, 123:15, 150:16, 167:7, 173:5, 175:7, 190:5
trying [28] - 6:20, 19:9, 19:18, 20:13, 20:14, 24:11, 28:2, 30:12, 53:2, 55:9, 55:13, 55:17, 66:23, 72:8, 75:17, 76:23, 76:24, 79:9, 99:15, 117:13, 119:13, 123:12, 154:18, 164:12, 165:2, 186:16, 187:3, 197:11
Tuesday [1] - 2:13
turn [6] - 32:14, 94:18, 175:12, 180:2, 189:5, 189:14
turning [1] - 175:16
two [57] - 9:1, 9:23, 10:4, 12:20, 14:19, 26:6, 26:9, 26:16, 26:20, 29:10, 29:14, 32:3, 39:5, 39:24, 40:12, 42:3, 42:7, 50:2, 71:8, 71:11, 72:11, 72:18, 80:17, 80:19, 99:6, 138:4, 142:18, 159:10, 171:18, 179:20, 179:21, 180:3, 180:9, 181:8, 181:10, 181:22, 182:8, 182:18, 183:10, 183:16, 184:15, 184:22, 184:24, 185:14, 186:16, 186:21, 187:6, 189:24, 190:7, 191:1, 191:10, 193:8, 193:20, 200:15, 201:1
type [5] - 33:21, 36:2, 37:19, 47:4, 192:14
typed [1] - 186:24
types [2] - 172:24, 181:15
typically [2] - 88:24, 89:2
typographical [2] - 10:14, 11:17

U

Uh-huhs [1] - 6:20
Uh-Uhs [1] - 6:20
Uhs [1] - 6:20
ultimate [3] - 17:6, 99:11, 109:5
ultimately [4] - 27:6, 51:10, 143:12, 191:19
unacceptable [1] - 160:22
unbundling [1] - 56:14
uncertainty [2] - 147:1, 180:17
unclear [2] - 70:16, 174:23
uncover [2] - 73:6, 73:14
under [20] - 18:9, 20:18, 24:19, 56:12, 125:9, 136:14, 143:15, 147:19, 151:8, 162:18, 163:5, 164:9, 171:16, 176:23, 181:11, 184:9, 201:18, 201:19, 207:22, 208:15
underaggregate [2] - 103:3, 205:20
underaggregated [2] - 152:4, 172:10
underaggregation [18] - 115:5, 125:1, 125:13, 125:18, 125:23, 164:6, 170:1, 183:8, 186:18, 201:14, 202:1, 202:9, 203:8, 204:5, 205:6, 205:13, 205:16, 205:18
underaggregations [1] - 114:12
underclustered [3] - 102:4, 102:6, 102:8
underclustering [9] - 99:19, 103:9, 133:5, 135:14, 135:20, 136:2, 136:10, 136:20, 160:2
underclusters [1] - 99:3
undergrad [3] - 11:6, 11:9, 33:21
undergraduate [1] - 12:2
underinclusive [1] - 102:17
underlie [1] - 187:15
underlying [11] - 14:7, 134:21, 141:24, 161:22, 162:13, 165:24, 175:11, 178:12, 179:3, 185:10, 193:10
underscore [1] - 177:5
understandings [1] - 70:5
understood [4] - 6:8, 9:9, 14:13, 62:17
undertake [1] - 136:18
unfounded [1] - 191:24
unique [4] - 29:5, 29:6, 113:1, 178:9
UNITED [1] - 1:3
universe [6] - 106:2, 117:2, 118:16, 120:19, 122:5, 175:6
unless [4] - 123:24, 124:10, 207:22, 208:4
unproven [1] - 193:7
unquestioned [1] - 104:12
unreliable [4] - 101:2, 134:5, 134:6, 162:11
unsubstantiated [1] - 70:6
unsupported [5] - 70:6, 71:4, 71:5, 104:12, 191:24
up [24] - 11:19, 20:17, 20:19, 40:22, 44:24, 54:7, 66:11, 80:4, 91:11, 91:18, 92:9, 92:12, 96:7, 100:11, 100:19, 126:6, 141:16, 161:12, 163:19, 168:16, 174:22, 176:4, 187:1, 195:22
user [43] - 18:8, 19:20, 26:21, 29:6, 37:19, 37:21, 43:13, 69:6, 81:6, 81:9, 87:5, 87:12, 87:13, 91:13, 92:23, 93:2, 93:3, 95:7, 99:9, 104:17, 108:14, 110:15, 113:3, 121:16, 121:21, 136:24, 142:12, 144:7, 144:9, 144:13, 153:11, 154:22, 159:7, 176:7, 176:17, 177:18, 177:21, 183:2, 183:21, 184:23, 187:1, 190:14, 193:20
users [4] - 68:18, 154:20, 169:9
uses [1] - 155:5
usual [1] - 7:17
utilize [1] - 162:10
utilized [7] - 45:8, 142:6, 142:9, 142:10, 142:11, 155:3, 159:6
utilizes [1] - 169:4
utilizing [2] - 31:6, 197:12

V

valid [5] - 78:14, 123:24, 131:8, 187:18, 197:20
validate [5] - 23:6, 167:8, 169:23, 171:10, 171:12
validated [2] - 147:19, 170:24
validating [1] - 22:23
validation [2] - 70:5, 192:20
validity [2] - 73:12, 73:13
valuation [1] - 56:18
value [6] - 148:9, 160:12, 161:2, 197:3, 197:7, 197:8
values [5] - 71:18, 71:20, 82:3, 189:10, 197:1
variables [1] - 178:10
variety [1] - 38:23
various [7] - 20:8, 33:24, 36:4, 65:1, 93:7, 131:6, 153:22
varying [1] - 111:13

versus [3] - 72:19, 160:23, 188:5
viewed [1] - 79:15
visual [2] - 86:4, 134:16
visualized [1] - 40:21
Vitae [2] - 4:13, 33:2
VOLUME [1] - 1:1
vs [2] - 1:15, 1:21

W

wages [1] - 42:10
wait [1] - 6:13
waive [4] - 7:9, 7:10, 7:11, 7:13
waived [1] - 208:5
walk [1] - 45:19
walked [2] - 14:11, 153:23
wallet [8] - 105:5, 105:11, 111:18, 112:6, 126:14, 129:6, 145:18, 145:19
wants [1] - 172:20
ways [8] - 12:22, 29:19, 38:2, 62:20, 148:10, 148:15, 148:16, 148:18
weeks [3] - 37:15, 37:17, 37:24
weight [1] - 196:8
weighting [2] - 65:11, 67:1
weights [7] - 65:1, 65:3, 65:24, 66:3, 66:14, 67:3, 67:5
whatsoever [1] - 89:4
WHEREOF [1] - 207:12
whole [2] - 121:17, 171:7
wide [3] - 38:22, 40:18, 158:14
wife [1] - 152:22
Winkler [1] - 53:10
winner [11] - 48:5, 49:8, 145:11, 146:21, 169:22, 170:7, 171:14, 172:6, 172:9, 183:23, 184:7
winners [7] - 8:19, 47:19, 109:1, 144:16, 184:4, 184:10, 191:5
WINNERS [2] - 1:17, 1:23
winnings [2] - 146:23, 191:7
wit [1] - 207:5
Witness [2] - 156:5, 208:1
witness [6] - 2:3, 207:8, 208:4, 208:5, 208:6, 208:7
WITNESS [11] - 4:3, 7:5, 7:7, 33:10, 33:14, 76:11, 76:17, 131:20, 198:14, 206:13, 207:12
wondered [1] - 106:20
wonderful [1] - 189:4
word [5] - 29:18, 40:5, 94:8, 119:7, 120:5
words [5] - 6:19, 84:6, 91:24, 122:13, 125:2
works [2] - 45:22, 131:21
world [10] - 104:6, 104:7, 104:14, 126:9, 130:3, 146:4, 162:8, 170:12, 170:13, 187:18
worse [1] - 119:10
wow [1] - 33:10
write [1] - 32:8
writing [1] - 207:7
written [2] - 13:16, 205:9
wrote [3] - 13:5, 13:13, 141:1

Y

Y-M-P-A-C-T-U-S [1] - 68:24
year [2] - 36:11, 36:12
years [8] - 37:6, 38:18, 51:2, 52:17, 53:13, 76:2, 123:5, 138:13
yield [1] - 163:2
Ympactus [5] - 50:10, 68:23, 68:24, 69:7, 157:15
yourself [3] - 32:9, 46:5, 72:9
yup [8] - 14:24, 32:1, 32:21, 55:24, 56:10, 145:3, 145:8, 195:4

Z

zero [17] - 146:7, 157:7, 160:1, 168:13, 197:2, 201:24, 202:8, 203:7, 203:8, 204:3, 204:4, 205:5
zeroes [1] - 161:4

EXHIBIT 13

Originally posted: 2022-08-04. View source code for this page here.

Fuzzy Matching and Deduplicating Hundreds of Millions of Records with Splink

A fast, accurate and scalable record linkage library that supports multiple SQL backends



Article updated 2022-08-04

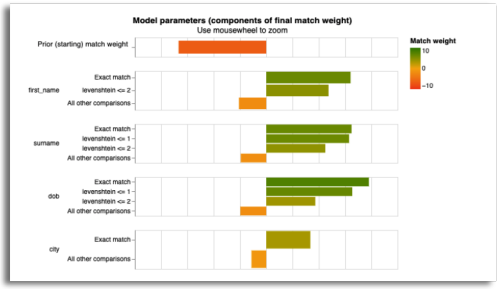
Summary

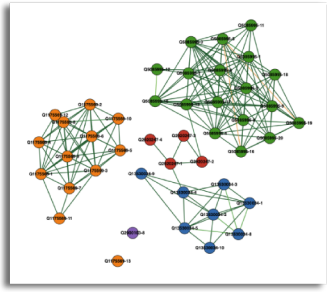
Splink is a Python library for probabilistic record linkage (entity resolution). It supports running record linkage workloads using the `Apache Spark`, `AWS Athena`, or `DuckDB` backends.

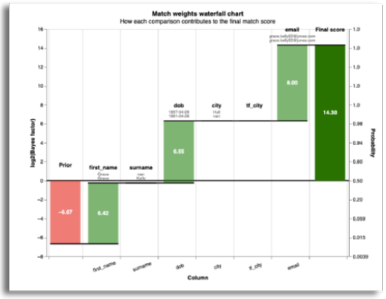
Its key features are:

- It is extremely fast. It is capable of linking a million records on a modern laptop in under two minutes using the `DuckDB` backend.
- It is highly accurate, with support for term frequency adjustments, and sophisticated fuzzy matching logic.
- It is capable of linking very large datasets (100 million records +) using the `Spark` or `AWS Athena` backends.
- It has a simple but highly customisable interface, so users can tackle the majority of record linking and deduplication problems
- Training data is not required because models can be trained using an unsupervised approach.

- It supports the full workflow of data linkage, from exploratory analysts, through to model prediction, diagnostics and quality assurance.
- Is robust, with an automated suite of unit and integration tests.







Problem statement

A common data quality problem is to have multiple different records that refer to the same entity but no unique identifier that ties these entities together:

row_id	first_name	surname	dob	city
1	lucas	smith	1984-01-02	London
2	lucas	smyth	1984-07-02	Manchester
3	lucas	smyth	1984-07-02	
4	david	jones		Leeds
5	david	jones	1990-03-21	Leeds

For instance, customer data may have been entered multiple times in multiple different computer systems, with different spellings of names, different addresses, and other typos. The lack of a unique customer identifier presents challenges at all stages of data analysis — from basic questions such counting the number of unique customers, to feature engineering of customers' details for machine learning purposes.

There is a large body of theoretical and empirical work into this problem. The solution usually involves computing a new unique identifier column which allows entities to be linked and grouped, using a process of statistical estimation, machine learning and/or rules-based logic.

However, there is a lack of free software that can tackle this problem at the scale of millions of records—the size typically seen in large organisations. Solving the problem usually involves generating very large numbers of record comparisons and so is ill-suited to in-memory solutions

in R or Python. Distributed computing frameworks like Apache Spark, or backends that are parellelised and not memory-bound like DuckDB, are a much better fit.

We have released a free and open source library called `splink`, that implements the Fellegi-Sunter/Expectation Maximisation approach, one of the key statistical models from the data linking literature. This is an unsupervised learning algorithm which yields a match score for each pair of record comparisons.

What it does:

In a nutshell, Splink takes messy input data in the table above and produces pairwise predictions of which records match:

row_id_l	row_id_r	match_probability
1	2	0.9
1	3	0.85
2	3	0.92
4	5	0.7

It then clusters the predictions to produce an estimated unique id:

cluster_id	row_id
a	1
a	2
a	3
b	4
b	5

It makes no assumption about the entity type (person, company, case, etc.), so it can be used to link or most types of data.

Try it out

You can try out the library in a Jupyter notebook using our Binder link here. These demos illustrate how to use the library, but note they are running in local mode on free servers, so don't expect great performance.

How it works

Splink is an implementation of the Fellegi-Sunter model. The software generates pairwise record comparisons using an approach called blocking, and computes a match score for each pair which quantifies the similarity between the two records.

The match score is determined by parameters are known as partial match weights. These quantify the importance of different aspects of the comparison.

For example, a match on date of birth lends more evidence in favour of two records being a match than a match on gender. A mismatch on postcode may provide weak evidence against a match because people move house, whereas a mismatch on date of birth may be stronger evidence against the record being a match.

This simple idea has a lot of power to build highly nuanced models. Partial match weights can be computed for an arbitrary number of user-defined scenarios, not just a match or non match. For example, a partial match weight can be estimated for a scenario where postcodes do not match, but are within 10 miles of each other.

These partial match weights are combined into an overall match score, which represents the weight of evidence that the two records are a match.

The library uses unsupervised learning (the Expectation Maximisation algorithm) to estimate these match weights. You can learn more about the theory in my interactive training materials.

A more detailed video description of how this all works can be found here.

Sample Code

We have tried to design an interface which is simple but nevertheless can be adapted to most record linkage and deduplication problems.

In the following code we:

- Specify a data linkage model
- Estimate its parameters
- Use the model to compute pairwise match scores
- Cluster the match scores into groups to produce an estimated unique person id

```
from splink.duckdb.duckdb_linker import DuckDBLinker
from splink.duckdb.duckdb_comparison_library import (
```

```
    exact_match,
    levenshtein_at_thresholds,
)

import pandas as pd

df = pd.read_csv("./tests/datasets/fake_1000_from_splink_demos.csv")

# Specify a data linkage model
settings = {
    "link_type": "dedupe_only",
    "blocking_rules_to_generate_predictions": [
        "l.first_name = r.first_name",
        "l.surname = r.surname",
    ],
    "comparisons": [
        levenshtein_at_thresholds("first_name", 2),
        exact_match("surname"),
        exact_match("dob"),
        exact_match("city", term_frequency_adjustments=True),
        exact_match("email"),
    ],
}

linker = DuckDBLinker(df, settings)

# Estimate its parameters
linker.estimate_u_using_random_sampling(target_rows=1e6)

blocking_rule_for_training = "l.first_name = r.first_name and l.surname = r.sur
linker.estimate_parameters_using_expectation_maximisation(blocking_rule_for_tra

blocking_rule_for_training = "l.dob = r.dob"
linker.estimate_parameters_using_expectation_maximisation(blocking_rule_for_tra

# Use the model to compute pairwise match scores
pairwise_predictions = linker.predict()

# Cluster the match scores into groups to produce an estimated unique person id
clusters = linker.cluster_pairwise_predictions_at_threshold(pairwise_prediction
clusters.as_pandas_dataframe(limit=5)
```

Feedback

We are grateful to all our users who have contributed and provided feedback so far. Please continue to do so by:

- Starting a discussion if you have questions about how to do something
- Raising an issue if you've found a bug or would like to request a new feature
- Raising a pull request if you'd like to fix a bug or add a feature

Or I'm @robinlinacre on Twitter.

Back to homepage.

This site is built using Observable HQ and Gatsby.js. Source code here.

EXHIBIT 14

other ways.[1]

III. SAMPLING

§ 4:13 Generally

Researchers rarely collect data on every single instance of the objects of study. That is called a *census*. They usually *sample* those objects. Agricultural researchers sample the corn in a field; they do not measure each and every ear. The same goes for every other kind of empirical researcher. Sampling not only is less expensive and less time-consuming, under most circumstances it is more accurate than a census. With a proper sampling design, resources can be directed at collecting the most accurate data on a smaller number of people, things, or events. Indeed, demographers evaluate how well the United States census is conducted by comparing the results of the census to samples.

§ 4:14 Units of analysis

The first step in sampling is to decide what is to be sampled, that is, what is the unit of analysis, what is the level of aggregation. For example, does one collect data about individual people or about aggregations such as cities or nations? About workers, organizations, or industries? About rocks, planets, or solar systems? Some things exist only at higher levels of aggregation. For example, the way a corporation is organized cannot be discerned by examining individuals, but only the structural relationships of groups of individuals to other groups.

These choices may have consequences for the statistical analyses that can be conducted and the conclusions that are drawn. Sometimes, when the phenomenon of interest can be studied by observing things at different levels of aggregation, different conclusions result from one using one unit of analysis rather than another.

§ 4:15 Types of sampling

Typically, the goal of sampling is to learn about an entire population of things by looking at a subset of them. The key to accomplishing this is to select the sample in such a way that it is representative of the population. Then, what one learns about the sample is likely to be true also for the population.

The methods used to do this are known collectively as probability sampling. *Probability sampling* involves selecting cases

---

[Section 4:12]

[1]For further details about these descriptive statistics, see §§ 5:1 et seq.

from the population in such as way that there is a known probability of any case appearing in the sample. This permits the use of probability theory to draw inferences about the nature of the population. Following are some common kinds of probability sampling.

Simple random sampling involves drawing a sample from the relevant population so that every member of the population has an equal chance of being selected into the sample. For example, if one wanted to measure the incidence of a certain disease among students at a school, a sample of students could be chosen at random from a hat or a computer could generate a random subset of students. Then the sample could be contacted for whatever testing was needed. *Systematic sampling* is similar, but instead of choosing at random, a random starting point would be selected and then every, say, 10th student in the student directory would be chosen (providing a 10% sample of the student body). These methods work fine if every member of the population of interest is known and the population is homogeneous.

A *stratified sample* is one in which subgroups of the population have been specified in advance, and then sampling takes place from within each stratum. This is helpful when some groups within the population occur in small numbers, and a researcher wants to be sure that enough of them are drawn so that a large enough sample of them is obtained to draw trustworthy inferences about the subgroups in the population. One could draw samples proportionate to the size of each stratum. For example, 40% of the sample could be drawn from the stratum that contains 40% of the population and 60% of the sample from the stratum that contains 60% of the population. Alternatively, one might wish to draw a "disproportionate stratified sample." For example, a sample of 250 from the 10% minority and 250 from the 90% majority. Simple random sampling would have yielded only about 50 for the minority, and that might have been too few. The two subsamples from this stratified sampling would then be weighted (one receiving nine times the weight of the other) so that in the end the statements made about the population would be accurate.

The two preceding methods can be combined in various ways to deal with more complicated circumstances, such as where the population is far flung, heterogeneous, and the identities of elements (usually, names of people) are not known. An example of this is *multi-stage cluster sampling*. Imagine that we wanted to study the health of people within the U.S. We do not have a list of every individual to choose them at random, and even if we did it would be inefficient to try to visit those selected, dispersed all over the country. But we could obtain a list (a "sampling frame") of all counties in all states. We could draw a representative sample of counties from that list. The list might even be strati-

fied in some way, such as by region of the country. (That provides the first stage of the sampling design.) From within the selected counties, we could draw a random sample of people to be tested, perhaps through random digit dialing of telephone numbers and inviting them to be tested. (Selecting those individuals is the second stage.) Thus, the people who become the sample are "clustered" in selected counties around the country.

Two lessons should be drawn from the sampling designs described. A particular sampling design can be devised to suit the nature of what is to be sampled by combining different sampling methods (random and stratified, in different stages). And each such design is, nevertheless, a probability sample because one would know the likelihood of selecting any element at any stage.

Various methods of *non-probability* sampling have been developed. One of these is *purposive sampling*, reflecting that the researcher has a particular purpose in mind for the way the sample is chosen. For example, in a study to discover how physicians learn about new drugs, researchers began with pharmacy records showing which physicians prescribed the drug in a given community. Then they interviewed the physicians, asking, among other things, which of their professional friends they also socialize with, and asked those friends who their professional friends were. The particular method described here is called "snowball sampling." As the sample grew, the researchers could track the diffusion of awareness of the new drug through the emerging friendship network. (It was found that doctors were more likely to learn about new medications from their friends than from medical journals.) It should be apparent that one of the more traditional probability sampling methods would not have been as useful to answering the research question.[1]

§ 4:16 Selection bias

The major defect of any sampling project would be that it fails to select representative elements from the population of interest. Most commonly this results from selection bias.

Selection bias refers to a sample being drawn in a way that makes it unrepresentative of the population to which inferences are to be made. The problem of selection bias is most easily understood by considering several examples. The reader might ponder why conclusions from each of the following samples are likely to be misleading. (Following in parentheses are likely

---

[Section 4:15]

[1]More detailed discussion of several aspects of sampling can be found in §§ 7:1 et seq.

answers):

— A criminologist set out to study criminals by conducting extensive interviews with a random sample of inmates in a state's prisons. (Whatever the study finds can tell us only about those criminals who were caught and incarcerated.)

— Doctors learned about the nature of the disease histoplasmosis by studying patients who came to their hospitals with the disease, and concluded that it was a rare disease which was almost always fatal. (By contrast, public health researchers sampled the public at large and found the disease to be far more common and to lead only rarely to serious harm. Doctors in hospitals saw only the few patients who suffered from the disease seriously enough to require medical attention.)

— During World War II a study was done to determine where additional armor might be placed on planes to protect them from anti-aircraft fire. To do the study, planes returning from missions were examined to see where they had taken the most hits from enemy fire. (The planes that were shot down but could have been saved with additional armor were the ones that would have been the most informative, but, of course, those never returned.)

— A graduating class organizing committee sent an anonymous questionnaire to their members prior to their 25th reunion. It included the question of the classmate's income. When the responses that came back were averaged, the committee was surprised to learn how successful their graduating class members had been in life. (A "response rate" of less than 100%, even a low one, is not in itself a problem. The issue, as usual, is whether those who responded are representative of the population. In the present example, if the more successful alumni are more likely to send back their questionnaires, the average income for the class will appear higher than it actually is.)

§ 4:17 Statistical aspects—Sample size

How large should a sample be to yield valid results? The answer is somewhat counter-intuitive. First of all, the absolute size of the sample is what is important, not the size of the sample relative to the size of the population. (The mathematics of this fact are presented in most textbooks on the statistics of sampling.) Second, the researcher usually has a better idea of what the needed sample size was after the data are collected than before. Third, the larger the sample size the better. The problem researchers face, however, is that while a real gain in accuracy can be achieved when going from n=10 to n=25 or to n=100, as larger numbers are added to the sample, the marginal gain in accuracy shrinks quite considerably. So a researcher has to ask

whether the extra trouble and expense involved in adding another 100 or 500 is worth the gain in accuracy that will be achieved.

How large a sample needs to be depends upon three things:

1. The homogeneity of the variable to be measured in the population. For example, suppose a warehouse holding cans of soup had been flooded and all the labels washed off. But it was known that only a single flavor (that is the variable of interest) of soup was stored per warehouse. How many cans have to be opened to answer the question of what flavor soup was in all of the cans in the warehouse? The obvious answer is: one. (But if the warehouse held numerous different kinds of soup, a larger sample would have to be gathered.) The homogeneity of the variable often cannot be known until after the data are collected.

2. How narrowly the researcher needs to zero in on the answer. That is, how narrow the researcher wants the "confidence limits" to be around the statistical statements that can be made about the population. (Depending on the research question, sometimes plus or minus 10 or 20 percentage points will do well enough; other times it will not do at all.)

3. How confident the researcher needs to be that the obtained range around the population parameters is correct—logically, this depends upon the nature of the research questions, although typically researchers follow the convention of aiming for 99% or 95% confidence.

§ 4:18 Statistical aspects—Confidence limits

Since samples are collected not for their own sake, but for the purpose of inferring back to the population from which the samples came, methods have been developed to allow researchers to perform that task of statistical inference. Suppose a regulator needed to determine the average amount of carbon monoxide being discharged per day per automobile in a given city. The researchers might sample cars and days. Suppose they found the mean amount in their sample to be "50 units." By knowing: (a) the variation (heterogeneity) in their sample; (b) the sample size; and (c) the level of confidence the researcher wishes to have, the researchers can also calculate, for example, that the true population parameter falls within the range of, say, 50 units plus or minus eight units, with 95% confidence. That is, there is a 95% probability that the true population mean falls somewhere be-

EXHIBIT 15

THE AMERICAN STATISTICIAN
2019, VOL. 73, NO. S1, 56–68: Statistical Inference in the 21st Century
https://doi.org/10.1080/00031305.2018.1529623

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

The Role of Expert Judgment in Statistical Inference and Evidence-Based Decision-Making

Naomi C. Brownstein[a,b,c], Thomas A. Louis[d], Anthony O'Hagan[e], and Jane Pendergast[f]

[a]Department of Biostatistics and Bioinformatics, Moffitt Cancer Center,Tampa, FL; [b]Department of Oncologic Sciences, University of South Florida, Tampa, FL; [c]Department of Behavioral Sciences and Social Medicine, Florida State University, Tallahassee, FL; [d]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; [e]School of Mathematics and Statistics, The University of Sheffield, Sheffield, UK; [f]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC

ABSTRACT

This article resulted from our participation in the session on the "role of expert opinion and judgment in statistical inference" at the October 2017 ASA Symposium on Statistical Inference. We present a strong, unified statement on roles of expert judgment in statistics with processes for obtaining input, whether from a Bayesian or frequentist perspective. Topics include the role of subjectivity in the cycle of scientific inference and decisions, followed by a clinical trial and a greenhouse gas emissions case study that illustrate the role of judgments and the importance of basing them on objective information and a comprehensive uncertainty assessment. We close with a call for increased proactivity and involvement of statisticians in study conceptualization, design, conduct, analysis, and communication.

ARTICLE HISTORY
Received March 2018
Revised September 2018

KEYWORDS
Bayesian paradigm; Clinical trials; Collaboration; Elicitation; Scientific method; Subjectivity

1. Introduction

As participants in the October 2017 Symposium on Statistical Inference (SSI), organized and sponsored by the American Statistical Association (ASA), we were challenged to host a session and write a paper inspired by the question, "Do expert opinion and judgment have a role in statistical inference and evidence-based decision-making?" While we work from different perspectives and in different statistical paradigms (both frequentist and Bayesian), there was a resounding "yes!" among us, with ample common ground in our thinking related to this infrequently discussed and often under-appreciated component of statistical and scientific practice.

Expert judgment is a feature of both frequentist and Bayesian inference. Judgments are subjective, and subjectivity is a topic that has generated heated debate in the Statistical community (Gelman and Hennig 2017). The subjectivity in a Bayesian prior distribution has been criticized by frequentists. Implicit claims are that frequentist methods are objective and that they, "let the data speak for themselves." However, frequentist methods also require some components of subjectivity in the choice of a model and of estimators, test statistics, etc. These choices are subjective in the sense that every expert builds knowledge and judgment into their own personal framework of understanding, which is not likely to be identical to that of anyone else. Expert judgment is clearly needed for valid statistical and scientific analyses. Yet, questions regarding how, when, how often, and from whom judgment is helpful, rather than leading to biased, misleading and nonreproducible results, is less clear.

The ASA commissioned this special issue of *The American Statistician* to stimulate "a major rethinking of statistical inference, aiming to initiate a process that ultimately moves statistical-science—and science itself—into a new age." We consider the roles of expert judgment in scientific practice. This article is a distillation of our common ground, along with our advice, warnings, and suggestions for incorporating expert judgment into science, while maintaining the integrity and scientific rigor of the research. Note that although we were asked in the Symposium to consider expert opinion and judgment, we avoid the word "opinion" because it risks being equated to uninformed rhetoric. We emphasize "judgment" that should be informed, carefully considered and transparent. While we might still disagree on specific details or the relative importance of various points, we present a strong, unified statement related to this key component of statistical practice. Additional literature on expert judgment in statistics may be found in a companion paper (Brownstein 2018).

Our article is organized in the following sections where we share our thoughts on when and how expert judgment has a legitimate and necessary role in scientific inquiry. Section 2 presents the role of judgment in the scientific method and more generally, in four stages of scientific inquiry, inference, and decision-making. In Section 3, the four stages are examined in more detail, focusing on their needs for expert judgment and the qualifications implied by "expert." Two case studies are presented in Section 4, with emphasis on the principled and scientific application of expert judgments. Finally, Section 5 summarizes our key conclusions.

CONTACT Naomi C. Brownstein naomi.brownstein@moffitt.org Department of Behavioral Sciences and Social Medicine, College of Medicine Florida State University, 1115 West Call Street, Tallahassee, FL 32306.

Supplementary materials for this article are available online. Please go to www.tanfonline.com/r/TAS.

© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

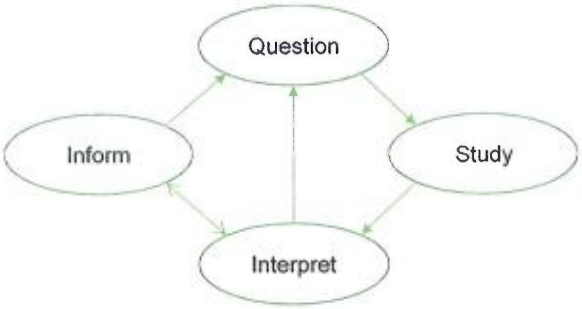


Figure 1. The cycles of inference and decision.

2. The Cycles of Inference and Decision in Science

The practice of science is often described in terms of the scientific method, defined in Oxford University Press (2018d) as involving, "systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses." More generally, one might describe the scientific method as a collection of methodologies and processes of gathering observable, measurable evidence using experimentation and careful observation, and of analyzing the evidence to draw conclusions, inform decisions and raise new questions. The purpose may be to inform about pathways or mechanisms by which results are obtained or to aid in prediction or estimation of quantities of interest. A thorough review of the history of the scientific method is found in Anderson and Hepburn (2016). While we believe that not all scientific inquiry, and certainly not all decision-making, falls within that prototypical guide, it is important to understand how the scientific method fits into our framework. Namely, science is data-driven in the ascertainment of objective information to draw scientific conclusions, yet with expert judgment feeding into the processes.

We base our four-stage graphic (Figure 1) on the depiction of Garland (2016), adding the Inform stage to accommodate decision-making and broadening the definitions of the other stages. It illustrates four stages in the perpetual process of scientific inquiry and evidence-based decision-making: Question, Study, Interpret, and Inform. These stages are formulated in highly general terms, because the practice of scientific inquiry, and hence the remit of statistical analysis, is also very wide.

We first describe activities that comprise each of these stages. Our framework allows us to underscore the varied roles of expert judgment throughout the perpetual cycles of the scientific method.

1. *Question*: Scientific inquiry can be characterized as beginning with one or more questions. A Question might be a formal scientific hypothesis arising either out of observation of real-world phenomena or from the current status of scientific knowledge and debate. On the other hand, a Question could be posed outside the scientific community, such as a request for evidence-based input to inform an impending policy decision. It might also simply express a wish to estimate more accurately certain quantities or parameters in current scientific theories.

2. *Study*: To address the Question scientifically, it is necessary to gather evidence in a rigorous manner. In the Study stage, we include all aspects of study design, including design of observational studies, experiments, questionnaires, systematic literature reviews, and meta-analyses. The Study may also involve sequential design or the design of a number of distinct, possibly concurrent studies. We also include in this stage the conduct of the study, resulting in some form of data.

3. *Interpret*: In the Interpret stage, data resulting from the Study are employed to address the Question. Typically, this may involve descriptive statistics and statistical inference, such as parameter estimation and hypothesis testing. In a Bayesian analysis, the primary statistical result might simply take the form of a posterior distribution. However, the "Interpret" stage should also embed the findings of the analysis into the wider body of science to which it refers, thereby updating that body of knowledge. In doing so, the wider implications of the findings will emerge.

4. *Inform*: The Interpret stage will often suggest new Questions, and a new cycle of scientific investigation will thereby be initiated. First, however, the Interpret stage usually will be followed by the Inform stage. For a formal scientific study, findings should be formally written and communicated in peer-reviewed outlets, such as conferences, journals, and books. In fact, peer-review may lead to revisions in the Interpret stage before formal publication of the findings. Subsequent examination and evaluation of published studies by the scientific community may in turn lead to new interpretations of existing studies and a new Question, leading to a new Study. Where the Question is a request for input to a decision, the Inform stage is when the results of the Study are communicated to facilitate the decision-making process. New Questions may arise based on the output produced in the Inform stage. Alternatively, the original Question may need to be revisited in one or more future Studies, especially when the evidence is not yet adequate to merit a robust conclusion.

3. Science and Subjectivity

In all stages of the scientific method defined in Section 2, expert judgment and knowledge are required. The relevant definitions are first presented. First, knowledge is defined (Oxford University Press 2018b) as, "facts, information, and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject." Judgment is defined (Oxford University Press 2018a) as, "the ability to make considered decisions or come to sensible conclusions." By contrast, opinion (Oxford University Press 2018c) is, "a view or judgment formed about something, not necessarily based on fact or knowledge."

We consider that judgment implies using information and knowledge to form an assessment or evaluation. While it is possible that opinions can be similarly well-informed, by definition, opinion does not necessarily imply that external information was incorporated into the evaluation. It may simply be a view, judgment, or appraisal that fits within a belief system or is com-

fortable for some other reason. While "judgment" is normally thought of as being based on observable truth, "opinion" can be simply based on preference. The subjectivity in this article refers to judgment of experts based on knowledge, skill, and experience. We examine the roles of expert judgment in each of the four stages presented in Figure 1, paying particular attention to integration of statistical and *content* expertise. We use content expertise to refer to the discipline expertise in which the Question arises. For more detail on expert scientific judgment, please see Brownstein (2018).

3.1. The Question Stage

When developing the Question, we rely heavily on the expert judgment of the content experts. The Question may arise from identification of a barrier or problem in need of an answer, a quest to understand the "why" behind some phenomenon, event, or process, or simply to better quantify some parameter, effect or disturbance. For example, one might ask "Why are some people able to fend off the negative impacts of an HIV infection while others cannot?" Knowledge of the literature and what other experiments or studies others have done to address this Question or related Questions is critical. Where the Question arises from a request for scientific input to inform a decision, the content experts serve key roles in formulating specific questions, such as "What can we say about the toxicity of this pollutant for fish in European rivers?", or "Which areas in this catchment will be flooded if the catchment experiences a once-in-100-years weekly rainfall?"

While the content experts have primary responsibility to develop the Question, statisticians can elicit clarity on the framing of the Question by asking pertinent questions from their perspective. Inquiry from the statistician serves not only to establish and confirm the statistician's understanding of the Question and its scientific context, but also to translate the question to a statistical framework, which may guide analytic decisions in the Study Stage. Indeed, strong listening and communication skills are critical for both the content and statistical experts! Barry Nussbaum, 2017 President of the ASA, shared his mantra, "It's not what you said. It's not what they heard. It's what they say they heard" (Nussbaum 2017). Communication can be improved by echoing back your understanding of how you interpreted what you heard, and asking others to echo their understanding of the points you have made. Producing a written summary of the Question, evidence-based justification for the importance of the Question, and evidence needed to answer the Question allows the research team members a chance to check on how well they are communicating and where points of disagreement may still exist.

If the Question is seeking information on potential pathways or mechanisms, evidence for or against competing theories is presented, and decisions must be made on the rationale for how the Question will be pursued. The content experts play the main role here, but statistical expertise can be helpful when framing the Question to bring up potential statistical issues with the proposed approach.

Part of defining the Question is determining what evidence, measures and parameters would be useful and adequate to arrive at an answer. Properties of those measurements, including validity, reliability, cost, and distributional properties are considered. This is an area in which both content and statistical experts can contribute. For example, content experts may be focused on what data would be needed and whether primary data collection would be needed or existing secondary data would suffice. If the study needs to collect primary data, there will be a need for discussion of exactly what information will be desired and how to elicit that information. In turn, the statistician will seek to better understand properties of the desired measurements. Moreover, the statistician may highlight unmeasured influences, the impact of missing data, and whether sources of bias, confounding, or variability could be reduced or eliminated by appropriate study design or data collection methods.

In addition, a Study may involve more than one Question. In this case, discussions are needed regarding which Questions are considered primary or secondary, the interrelatedness or independence of the Questions, and whether any of the Questions can be addressed jointly. Statistical and content experts should collaborate in making these decisions.

Thinking through the issues that can (and will!) arise when defining the Question and information needed to address it requires effective collaboration and team effort. Effective collaboration requires mutual respect for team members and personal authenticity; an understanding of the strengths and skill sets of each member; an understanding and buy-in for a common goal; a willingness not only to hear and understand what another has said, but to embrace ideas different from yours; a willingness to ask for evidence and assumptions behind a statement of fact or belief; reliability and consistency in thought and behavior; a willingness to offer alternative ideas or approaches; a willingness to compromise, when appropriate. All of this rely on fluency, both written and oral, in the language used, including technical and discipline-specific terms. Cognizance of acronyms and language shortcuts is the first step to reducing them to enhance communication. We recommend that researchers create a written summary of decisions made in such discussions, and make sure such documents are accessible electronically to all study members for review and editing.

3.2. The Study Stage

Once the Question and pertinent measures are defined, the approach to gathering information (data) must be developed. Can the Question be answered in one study, or will it take a series of planned studies? What studies have been done before? Could any component of prior studies be replicated with modification or improvement in this study? Did previous studies report unanticipated problems that could be avoided in this study?

Much of the work in planning a study involves a collaborative effort among all members of the research team. Those researchers who will be "on the ground" collecting information will have expertise on what is feasible and what is not. The statistician can offer their expertise and judgment on many aspects of study design and implementation, such as strengths and weaknesses of different study designs, questionnaire development, psychometric properties of data collection instruments, issues surrounding sources and control of error, replication,

operational randomization, and blinding. The content experts will share their knowledge and judgments on the target scope of inference, such as whether to estimate the 100-year flood plane for a large geographic area or just one river; for all people with a disease or only those in a local area who meet defined inclusion criteria. Content experts bring to the discussion additional information that can be considered in the study design, perhaps stemming from theorized or understood pathways, mechanisms, or concurrent influences by which observed outcomes can differ.

If working within the Bayesian framework, statisticians help elicit information from the content experts to feed into the prior distribution. Those approaching the problem from a frequentist perspective will also be looking to prior studies and expert judgment when developing a study design. No matter what analysis approach is used, the ultimate goal remains the same: to collect enough high quality information to effectively address the Question. Here, the word "quality" can encompass many aspects, including reduced variability of measures and removal or control of sources of biases, along with proper data collection and maintenance systems that protect the integrity of the data.

During the Study stage, prior to data collection, there should be enough information to create a study protocol, data monitoring plan, and a statistical analysis plan (SAP), upon which everyone agrees (Finfer and Bellomo 2009; Ellenberg, Fleming, and DeMets 2003; Ott 1991). Such SAPs are becoming more common and recognized by funders as an important part of the research process. While not every study will require these formal documents, the idea is that there is a common understanding on how the study will be conducted; how the data will be collected, monitored, and maintained; and the analytic approach used to address the Question. With the goal of transparency of the body of work to be accomplished, good documentation and data provenance are important components of scientific inquiry.

Unfortunately, forms of scientific malpractice, such as repeating the analysis with a slightly altered question in an attempt to achieve statistically significant results—a practice called *HARKing* (hypothesizing after the results are known)—or testing for many associations without prior hypotheses (p-hacking) are commonplace in the scientific literature (Kerr 1998; Head et al. 2015). A well thought-out study design and SAP can help safeguard against urges to reanalyze the data later after obtaining disappointing results. However, it should be noted that the data monitoring process itself involves judgment, as exemplified in Section 4.1.3 and discussed further elsewhere (Pocock 2006).

A related practice of publishing a Registered Report, in which methods and proposed analyses are preregistered and peer-reviewed prior to conducting the research, is gaining momentum. The intent is to eliminate questionable research practices, such as selective reporting of findings and publication bias, by provisional acceptance for publication before data collection begins. Currently, the Center for Open Science lists 121 journals that use the Registered Reports publishing format as an option in regular submissions or as part of special issues (Center for Open Science 2018).

Often, before a study can begin, funding must be obtained. When there are commissioned studies, often researchers need to write a formal proposal to seek funding. A grant proposal provides an opportunity to present not only the justification for the study but also detail on the study process, analysis plan, and how the desired information will address the Question. The grant review process brings in a new external set of experts with judgments of their own on the merits of the proposal. Feedback for a proposal can identify strengths and weaknesses of the proposed work. Reviewers' expert judgments on whether a particular study should be funded are intended to weed out studies without strong support and justification for both the importance of the Question and the development of the Study.

As the study progresses, important discussions and decisions are likely to be made at research team meetings. It is important that statisticians, like other key members of the team, are present at the table to listen, question, offer insight and expertise, and retain knowledge of decisions made. Written summaries of research meetings, documentation of any decisions made that were not anticipated in the Question or (earlier in the) Study stages or are necessary to successfully implement the study protocol, all promote transparency and reproducibility. Additional documentation for the Study stage that should be available for all team members to view may include computer code management systems, clearly documented analytic code, well-considered file structures and naming conventions, and common file space for all members to see such documentation. Subsequently, the team will synthesize information from the Study stage for use in the Interpret and Inform stages.

3.3. The Interpret Stage

The methodology used to analyze the data and the specifications of the assumed model feed into the interpretation of the data. Some analytic methods produce estimated probabilities that relate directly to the Question; others may provide information that helps address the Question, but perhaps more indirectly. The chosen methodology may produce parameter estimates that need to be understood in context of the model, and the alignment of the model to the observed data and prior beliefs needs to be considered.

The expertise of the statistician is needed both to understand the nuances of proper interpretation of the analytic results in context of the executed study, assumptions made, and modeling used and to guard against overinterpretation. For example, a statistician may help protect the team from common misconceptions and malpractice, such as tendencies to extend inferences to populations outside those under study or to interpret association as causation, or failure to consider the impact of unmeasured confounders, mediators or moderators in the interpretation of results. The statistical expert can protect against improper use or interpretation of a p-value, discuss the difference between clinical and statistical significance, and highlight the potential impact of missingness mechanisms and violations of statistical and causal assumptions on the results. If working in a Bayesian framework, the statistician can also discuss the impact of potential expertise bias and over or underprecision in specification of the prior.

Similarly, it is not appropriate for statisticians to focus solely on analytic methods and numeric results, or for content experts to delay involving statisticians until the Interpret stage. By not being involved in other study aspects and discussions, the statistician is poorly positioned to make modeling choices, recognize

possible biases, and interpret findings in the context of the study as conducted, which could differ from the study as planned.

As the analytic results are interpreted in the framework of the question and the Study protocol, both the content and statistical experts help the study team blend the (properly interpreted) new findings into their existing knowledge and understanding. This process will likely include team discussions of the clinical meaning and impact these findings might have in the population under study, and how the results could be explained or interpreted within the current framework of understanding. The discussion will include the strength of the evidence, based on posterior distributions, point estimates of key parameters, confidence or credibility intervals, etc. and consistency with prior studies, as well as potential weaknesses or caveats of the study posed by both the statistical and content experts.

3.4. The Inform Stage

Once the analytic results have been interpreted within the framework of the study design and measures used to address the Question, it is time to assess what was learned and share that information more broadly. Of course, those who developed the Question and often those who funded the study will need and expect a complete summary of the work done, describing how the results have informed the Question. Indeed, there may be interest in the work outside of academia, such as patients curious about new therapies for their conditions, policy-makers seeking to understand actions that may yield societal benefit, and others simply wondering about current topics and trends in various fields of science.

Scientific reports or publications, where the process, methods, results, and conclusions of a study can be shared broadly, are important tangible outcomes of the Inform stage. In each Inform stage, the team members build on what they knew previously and share what they learned from the findings, whether or not the results obtained were anticipated or exciting. To guard against publication bias, null results, in particular, must be communicated, despite disincentives for doing so (Franco, Malhotra, and Simonovits 2014; Easterbrook et al. 1991).

Strong communication and documentation skills in the Inform stage are paramount to ensure that all components of a study are documented as the study progresses and are presented clearly, completely, and with as much transparency as possible. While the statistician is frequently tasked with managing the sections describing the quantitative methods and results, they should also collaborate throughout the report to provide their input to interpretation and implications of the findings (see Section 3.3). Additional documentation from the Study stage, such as clearly written and readable computer code, can also be included as supplementary material.

The process of creating a scientific manuscript and undergoing the peer-review process for publication is another place in which statistical and content expert judgment enters into both the Interpret and Inform stages. Comments from others on drafts of the manuscript can lead to revised interpretation in light of new information or perspective from that feedback before submission of the manuscript for publication. Once submitted, additional expertise is gathered from the peer reviewers, such as calls for clearer evidence of claims or interpretations made, challenges made to stated justifications of assumptions or interpretations, call for greater transparency or detail in information presented, or citations to related work which may blend into the discussion section. Reviewer comments can greatly impact the quality and clarity of information presented in the final publication.

Recently, journal editors are considering re-examining their review processes, focusing acceptance or rejection on strength of evidence, not p-values. Locascio (2017b) argues for a results-blind review system in which the reviewer makes a preliminary decision on the strength of the manuscript without knowing the results. Final decisions would not allow rejection on the basis of p-values. Commentary on the feasibility of this approach may be found elsewhere (Marks 2017; Grice 2017; Hyman 2017; Locascio 2017a). We encourage journal reviewers to examine the appropriateness of the methods for the study under consideration, rather than accepting justification of methods simply based on their publication and use elsewhere.

When the Question has arisen to facilitate decision-making, the primary purpose of the Inform stage is to convey the scientific evidence to the decision-maker, after it has been assembled and analyzed in the Study and Interpret stages. Here, too, communication skills are particularly important. Governments are increasingly but not exclusively reliant on evidence for policy- and decision-making (Oliver and de Vocht 2017; HM Treasury 2015; Oliver, Lorenc, and Innvær 2014; LaVange 2014), and there is much current interest in the challenges of communicating uncertainty to decision-makers (Cairney and Oliver 2017; National Academies of Sciences, Engineering, and Medicine et al. 2017; National Research Council et al. 2012a,b). Despite these challenges, it is essential to measure uncertainty and important to try to communicate it as effectively as possible. Understanding the perspectives of decision-makers (i.e., priorities and goals, options under consideration, risks/benefits), what processes they must follow, and time-constraints are helpful in such communications.

As depicted in our graphic of the scientific method (Figure 1), learning and discovery is cyclical. Once we address one Question, new Questions often arise. Researchers may be interested in the work of other authors in similar fields, perhaps using information from published studies to inform their next study. Sometimes the results obtained are not definitive, or not adequate for robust decision-making, and ways to redirect the next investigation of the same or a revised version of the Question are planned. Other times, a replication Study is conducted based on the same Question simply to see if the results remain qualitatively similar, despite inevitable lack of perfect duplicability in all aspects of the study environment (Lindsay and Ehrenberg 1993).

3.5. But Is It Science?

When expert judgments are used in any stage of a scientific inquiry, the outcomes contain subjective elements. The inescapable conclusion is that science itself has a subjective component, aspects of which can be communicated probabilistically, and should be interpreted according to the theory of subjective probability (Anscombe and Aumann 1963). That is, while there may be objective information on which probabilities are

based, there also will always be between-individual (scientist) variation.

There can be heated opposition to the notion of subjective science, with objectivity promoted as fundamental to the scientific method, and subjectivity considered as anathema to a true scientist. According to this viewpoint, subjectivity is unscientific, biased, possibly prejudiced. But, science cannot be totally objective. Scientists propose new or amended theories, choose experimental designs or statistical analysis techniques, interpret data, and so on. Although these judgments are subjective, expert judgments result from a synthesis of related prior knowledge and experiences based on observable evidence and careful reasoning. Indeed, making informed subjective judgments is one of the key features that distinguish a top scientist from a lesser one. Statistics similarly involves subjective judgments, as others have recently argued (Gelman and Hennig 2017).

In the design portion of the Study stage, by definition, information is not yet available from the study being planned. Instead, study design decisions are *preposterior* or *preanalysis* and must be based on (prior) external data and judgment. In this sense, regardless of the subsequent analysis, one could consider the design phase as automatically Bayesian. Designers employ varying degrees of formalism in developing the study design and statistical models. A formal Bayesian approach can be used to either develop a frequentist design (Bayes for frequentist) by, for example, finding a sample size or other design components that ensure

$$P(\text{power} > \text{goal} \mid \text{design \& assumptions}) > \gamma,$$

(see e.g., Shih 1995 for an implementation), or to ensure that Bayesian properties are in an acceptable region (Bayes for Bayes). We recommend practitioners consider additional use of Bayesian approaches, even if only to provide a vehicle for documenting the roles of judgments and as a platform for sensitivity analyses.

All researchers, irrespective of their philosophy or practice, use expert judgment in developing models and interpreting results. We must accept that there is subjectivity in every stage of scientific inquiry, but objectivity is nevertheless the fundamental goal. Therefore, we should base judgments on evidence and careful reasoning, and seek wherever possible to eliminate potential sources of bias.

Science also demands transparency. Ideally, it should be standard practice in all scientific investigations to document *all* subjective judgments that are made, together with the evidence and reasoning for those judgments. Although we recognize that this may be impractical in large studies with many investigators, we believe that it might be facilitated with suitable collaborative working tools. We suggest, therefore, that journals might consider requiring that authors provide such documentation routinely, either in an appendix to a submitted paper or in online supplementary material.

4. Case Studies

We present three case studies, each exhibits complexities that require collaborative input from content experts and statisticians. They highlight the roles that expert judgments play and the steps that were taken to ensure that judgments were as objective and scientific as possible.

4.1. Expert Judgment in Randomized Clinical Trials

4.1.1. Background on Bayesian Clinical Trials

Bayesian approaches to clinical trial design, conduct and analysis have been shown to offer substantial improvements over traditional approaches in a variety of contexts. See Abrams, Spiegelhalter, and Myles (2004) and Berry et al. (2010), for a range of examples. The Bayesian approach formalizes using prestudy data and expert judgments in design, conduct, and analysis. Importantly, the formalism provides an effective language for discussing interim and final results, for example, by supporting statements such as, "in the light of accruing data, the probability that treatment A is better than treatment B by at least 5 percentage points is 0.95 ..." It provides a natural way to compute and communicate predictive assessments such as futility, that is, in the light of current information, what is the probability that, ultimately, the trial will not be definitive. The Block HF and TOXO studies outlined below illustrate many of these characteristics.

4.1.2. The Block HF Study

The BlockHF study (Curtis et al. 2013) provides an example of the benefits of embedding evaluations in the Bayesian formalism. It was an adaptive randomized trial using Bayesian criteria, with specification of all features dependent on collaboration amongst clinical and statistical experts. The abstract states,

> We enrolled patients who had indications for pacing with atrioventricular block; New York Heart Association (NYHA) class I, II, or III heart failure; and a left ventricular ejection fraction of 50% or less. Patients received a cardiac-resynchronization pacemaker or implantable cardioverter defibrillator (ICD) (the latter if the patient had an indication for defibrillation therapy) and were randomly assigned to standard right ventricular pacing or biventricular pacing. The primary outcome was the time to death from any cause, an urgent care visit for heart failure that required intravenous therapy, or a 15% or more increase in the left ventricular end-systolic volume index.

Two pacing regimens, biventricular (BiV) versus right ventricular (RiV) were compared using a Cox proportional hazards model. Analyses were stratified by two cardiac devices, with information on the two, stratum-specific hazard ratios (HRs) combined for monitoring. The Bayesian monitoring rules addressed patient safety, stopping for futility, and stopping if the treatment comparison was sufficiently convincing (probability that a clinically meaningful difference was sufficiently high), based on combining evidence over the two stratum-specific HRs. From the statistical analysis section,

> An adaptive Bayesian study design allowing up to 1200 patients to undergo randomization was used, featuring sample size re-estimation and two interim analyses with prespecified trial-stopping rules .... An intention-to-treat analysis served as the primary analysis for all outcomes.

Table 1. θ = a weighted average of stratum-specific log hazard ratios, each comparing BiV versus RiV pacing; PP_0 = posterior probability that the study objective has been met, PRR = posterior probability the study objective has not been met.

	Decision boundaries				
	Conclude objective is met and stop study early	Conclude that sample size is sufficient to continue	Determine that sample size is insufficient but elect not to increase sample size	Conclude that sample size must be increased in increments of 175	Stop study for safety
First Interim Analysis	$PP_0 > 0.99$	$0.90 \leq PP_0 \leq 0.99$	$PRR > 0.9$	$PP_0 < .90$ and $PRR \leq 0.9$	$P(\theta > 0 \mid \text{data, prior}) \geq 0.90$
Sample Size Re-estimation Phase	N/A	$0.90 \leq PP_0$	$PRR > 0.9$	$PP_0 < .90$ and $PRR \leq 0.9$	N/A
Second Interim Analysis	$PP_0 > 0.99$	If neither the outcome in column 2 nor the outcome in column 6 occurs, then the study will continue with the current sample size.			$P(\theta > 0 \mid \text{data, prior}) \geq 0.90$

NOTE: See Curtis et al. (2013) for details.

The trial resulted in a "win" for biventricular pacing compared to right ventricular, HR 0.74, 95% credible interval (0.60, 0.90). Table 1 presents the monitoring rules which are based on the following,

- θ = a weighted average of stratum-specific log HRs, each comparing BiV versus RiV pacing.
- $PP_0 = P\{\theta < \log(0.90)|\text{data}\} = \text{pr}\{\text{HR} < 0.90|\text{data}\}$; the posterior probability that the study objective has been met.
- $PRR = P\{\theta > \log(0.90)|\text{data}\} = \text{pr}\{\text{HR} > 0.90|\text{data}\}$; the posterior probability that the study objective has not been met.
- $P\{\theta > 0|\text{data}\} = \text{pr}\{\text{HR} > 1.00|\text{data}\}$; the posterior probability of a safety concern.

For example, if the probability of meeting the study objective is sufficiently high ($PP_0 > 0.99$), stop the study; if moderately high ($0.90 < PP_0 < 0.99$), continue the study with the current sample size target; if too low ($PP_0 < 0.90$) but there is a reasonable likelihood of success ($PRR < 0.90$), increase the sample size. Inclusion of such probability-based rules was an important benefit of using the Bayesian formalism. It supported complex decision rules that communicated effectively with the clinical experts on the monitoring board.

4.1.3. The "TOXO" Trial

When accruing data are consistent with a prior distribution, a trial can be stopped earlier than with traditional, likelihood-based monitoring. However, there are also situations wherein the prior and the data diverge and stopping is delayed, as is the case in the following, post hoc, rerunning of the Community Programs for Clinical Research on AIDS randomized trial of prevention of TOXO conducted in the 1990s. The study compared pharmacological prevention at a subtherapeutic dose with placebo, with all treatment groups carefully monitored for indications of disease. The premise of such prevention studies is that a low dose of a pharmaceutical that is typically used to treat overt disease may also prevent or delay onset. However, even a low dose of a pharmaceutical may induce adverse effects, such as toxicities or resistance to treatment, and consequently, watchful waiting (the placebo "intervention") may be better than the potentially preventive treatment.

Jacobson et al. (1994) detail the several decisions and expert judgments needed to design and implement the trial, including choice of drugs, inclusion and exclusion criteria, clinical endpoints, and monitoring and analysis plans. The abstract of the article reporting results (Jacobson et al. 1994) states in part,

> Pyrimethamine, 25 mg thrice weekly, was evaluated as primary prophylaxis for toxoplasmic encephalitis (TE) in a double-blind, randomized clinical trial in patients with human immunodeficiency virus (HIV) disease, absolute CD4 lymphocyte count of < 200/microL [CD_4 lymphocytes fight disease; a low level indicates immunodeficiency] (or prior AIDS-defining opportunistic infection), and the presence of serum IgG to Toxoplasma gondii." ..."There was a significantly higher death rate among patients receiving pyrimethamine [compared to control] (relative risk [RR], 2.5; 95% confidence interval [CI], 1.3–4.8; $P = .006$), even after adjusting for factors predictive of survival. The TE event rate was low in both treatment groups (not significant). Only 1 of 218 patients taking [the control intervention] but 7 of 117 taking aerosolized pentamidine for prophylaxis against Pneumocystis carinii pneumonia developed TE (adjusted RR ..., 0.16; 95% CI, 0.01–1.79; $P = .14$). Thus, for HIV-infected patients receiving trimethoprim-sulfamethoxazole, additional prophylaxis for TE appears unnecessary.

The Data and Safety Monitoring Board monitored the trial at prespecified, calendar-determined dates using the O'Brien and Fleming (1979) boundaries. Early in the trial, these boundaries require substantial evidence (e.g., a small p-value) to stop and make a decision; as the trial approaches the predetermined maximum number of follow-up visits, the criterion is close to that for a fixed sample size trial.

The full, calendar time indexed database was available for the after-the-fact example of how monitoring might have proceeded using a Bayesian approach. This illustrative analysis evaluated the "TOXO or death" endpoint using the Cox (1972) proportional hazards model with adjustment for baseline CD_4 count. The illustrative trial was stopped when the posterior probability of benefit or the posterior probability of harm became sufficiently high. Importantly, prior elicitation occurred while the trial was being conducted, before any outcome information was available to the elicitees.

4.1.4. Model for the "TOXO" Trial

The HR (relative "TOXO or death" event rate between the two treatment groups) was modeled using a log-linear model with covariates treatment group ($z_{1j} = 1$, if participant j received pyrimethamine; $z_{1j} = 0$, if placebo), and CD_4 cell count at study entry (z_{2j}). The CD_4 covariate adjusted for possible differences

in immune status at baseline between the two groups. Specifically, the log hazard-ratio is,

$$\log(\text{hazard ratio}) = \beta_1 z_{1j} + \beta_2 z_{2j}$$

with $\beta_1 < 0$ indicating a benefit for pyrimethamine. A flat prior was used for the CD_4 effect (β_2), and a variety of priors were developed for the pyrimethamine effect (β_1). The choice of the Cox (1972) model and the use of a noninformative prior for β_2 were judgments of the statisticians. Though the Cox model has become the default choice when modeling the time to an event, it is based on important assumptions, namely that the hazard functions for both interventions are proportional and censoring is noninformative. As such, the model should only be adopted after careful consideration by experts, such as in this example. The choice of a noninformative prior distribution for β_2 reflects the clinicians' and statisticians' judgments that no information was available in advance of the trial on the association of CD_4 with TOXO incidence.

4.1.5. Elicitation in the "TOXO" Trial

As described in Carlin et al. (1993) and Chaloner et al. (1993), prior distributions for β_1 were elicited from five content experts—three HIV/AIDS clinicians, one person with AIDS conducting clinical research, and one AIDS epidemiologist. Elicitation occurred while the trial was in progress, only pretrial information was available to the elicitees. Two additional priors were included in the monitoring—an equally weighted mixture of the five elicited priors and a noninformative flat prior (equivalent to using the normalized, partial likelihood as the posterior distribution).

Elicitation targeted potentially observable, clinically meaningful features and then transformed responses to parameters of the Cox model. That is, rather than directly elicit a prior for the hazard ratio, each elicitee was asked to report their best estimate of the probability of TOXO or death in two years under placebo (P_0), then draw a picture of the distribution of the two-year probability under pyrimethamine, conditional on the estimate under placebo ($[P_{\text{pyri}} \mid P_0]$). Then, for each elicitee these conditional distributions were converted to a prior distribution for the log(hazard ratio) using the transformation, $\beta_1 = \log(1 - P_0) - \log(1 - P_{\text{pyri}})$.

Figure 2 displays the elicitation results. At trial initiation, there was little known about the baseline incidence of TOXO, and so the content experts based their distributions on general expertise and analogy with other endpoints and contexts. The range of the five reported two-year incidence probabilities under placebo was wide, but elicitees believed that TOXO had a non-negligible baseline incidence. All elicitees were optimistic regarding pyrimethamine's benefit, with experts C and E the most optimistic, placing all of their probability distribution for incidence under pyrimethamine way below their estimate of incidence under placebo.

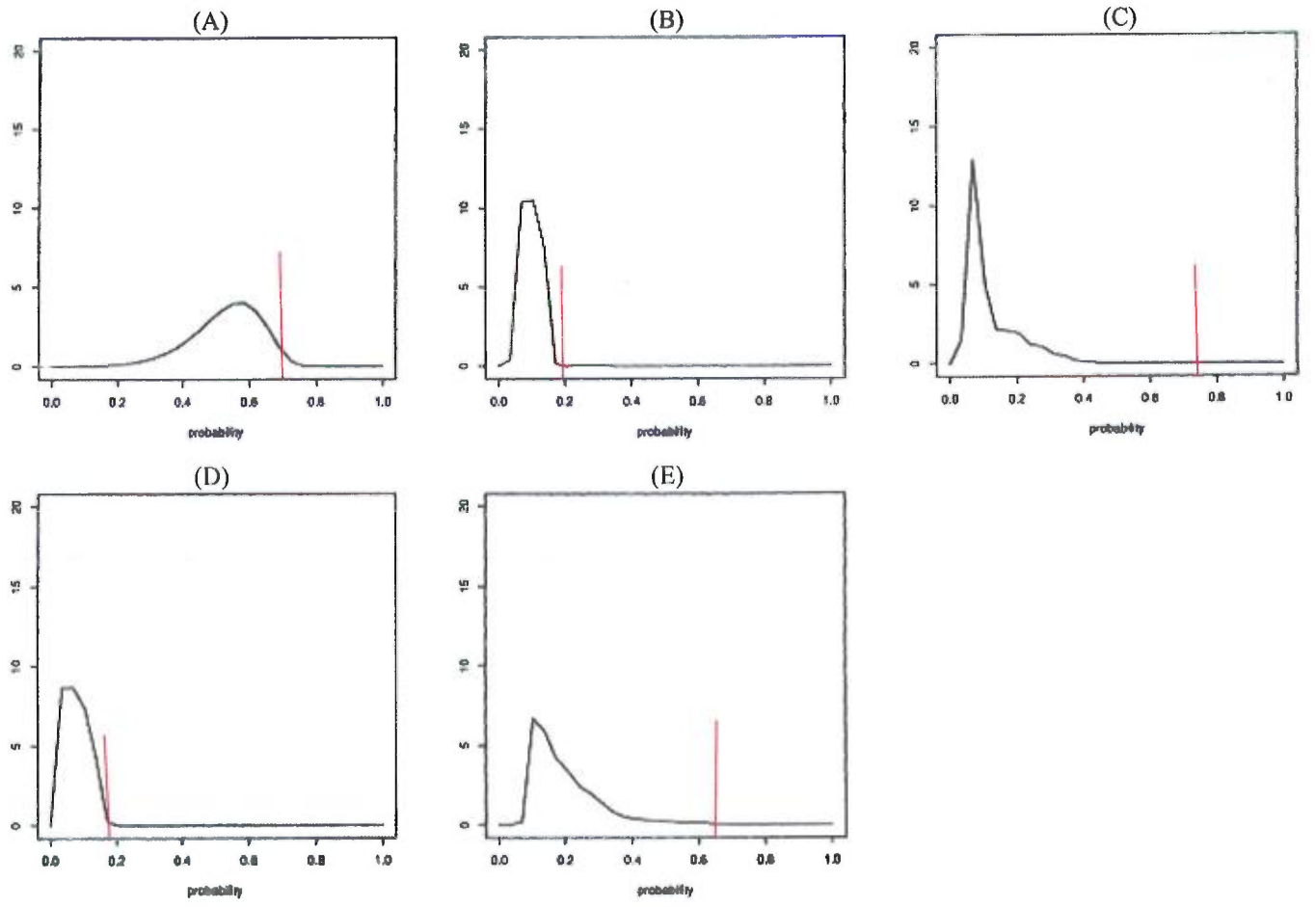


Figure 2. Elicited priors for the five elicitees. The red vertical line at P_0 is the "best guess" two-year incidence of TOXO or death under placebo. The smoothed densities are for two-year TOXO or death incidence under pyrimethamine, conditional on the placebo rate.

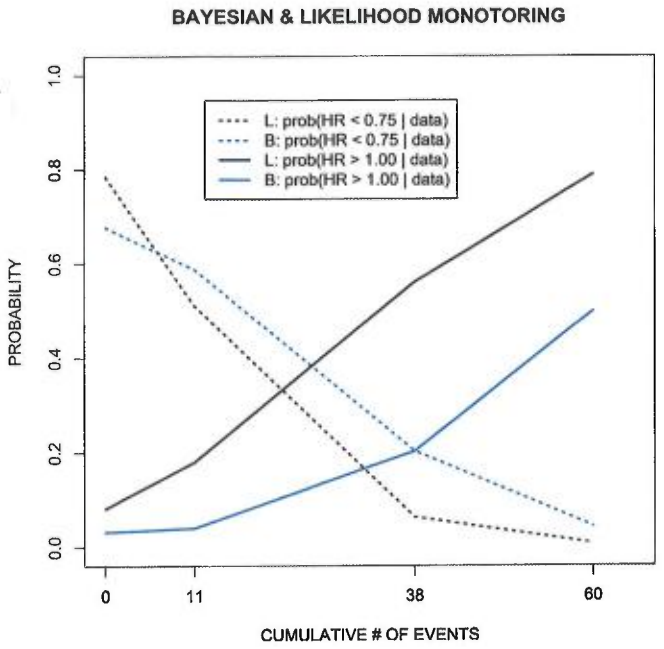


Figure 3. Posterior probability of benefit (hazard ratio $\leq$ 0.75) and harm (hazard ratio $>$ 1.0) for the mixture prior (blue lines) and for monitoring based on the partial likelihood (black lines), which is equivalent to using a flat (improper) prior. Monitoring was at calendar dates (15 Jan 1991, 31 Jul 1991, 31 Dec 1991, 30 Mar 1992) with the X-axis indicating the cumulative number of toxoplasmosis or death events from the combined arms.

While a degree of optimism may be needed to motivate conducting a trial, ethics require that there be sufficient equipoise (uncertainty) regarding which treatment is superior in order to initiate a trial. The elicited priors in this example probably express too strong a belief in the efficacy of pyrimethamine to be used in monitoring an actual trial, but using them in this illustrative monitoring exercise with comparison to likelihood-based monitoring, effectively highlights issues we discuss in Section 4.1.7.

4.1.6. Monitoring Results for the "TOXO" Trial

The actual trial was monitored at calendar dates (15 Jan 1991, 31 Jul 1991, 31 Dec 1991, 30 Mar 1992). At the December 31, 1991 meeting, the monitoring board recommended stopping the trial for futility, because the pyrimethamine group had not shown significantly fewer events, and the low overall event rate made a statistically significant difference in efficacy unlikely to emerge. Additionally, an increase in the number of deaths in the pyrimethamine group relative to the placebo indicated a safety issue.

For the illustrative, after the fact monitoring example, Figure 3 displays posterior probabilities of benefit (HR $\leq$ 0.75, equivalently, $\beta_1 \leq \log(0.75)$ in the Cox model), and harm (HR $>$ 1.0; $\beta_1 > 0$) for an equally weighted mixture of the five elicited prior distributions (denoted by "B"), and for a flat prior that generates partial likelihood-based/traditional monitoring (denoted by "L"). For example, $\beta_1 = \log(0.75)$ indicates that the hazard rate (event rate) in the pyrimethamine group is 75% of that in the control group; $\beta_1 = 0 = \log(1.0)$ indicates equal rates.

As displayed in Figure 2, elicitees believed that TOXO had a nonnegligible incidence and that pyrimethamine would have a substantial prophylactic effect. Thus, each prior and their mixture are to varying degrees far from the accruing information in the trial. Consequently, monitoring based on the partial likelihood, which can be considered "flat prior Bayes," gives an earlier warning of harm compared to monitoring based on the mixture prior. The mixture required considerably more information to overcome the a priori optimism of the elicitees. Also, the posterior probability of harm based on monitoring with any single prior in Figure 2 would lag behind that based on the partial likelihood, with use of prior A, B, or D giving an earlier warning than use of priors C or E.

4.1.7. Discussion of the "TOXO" Trial

Traditional, non-Bayesian monitoring depends on several judgments. For example, trial design including maximum sample size depends on assumptions about baseline event rates, the HR, principal efficacy and safety endpoints, the statistical model, etc. In the Bayesian approach, a subset of these features are given prior distributions, in the TOXO example by independent priors for the slopes on the log HR and on the baseline CD_4 value. These priors, along with other design features determine the monitoring frequency, the shape of monitoring boundaries (flat, increasing, decreasing) and associated values. Boundaries are necessary for all monitored trials, determining them can be either Bayesian or frequentist.

The after-the-fact analysis highlights ethical issues generated by real-time use of expert knowledge. It shows that if the elicited priors had been used in the actual clinical trial monitoring, then due to their optimism it is very likely that trial stopping and other decisions would differ from those produced by likelihood-based (flat prior) monitoring. However, in early phase clinical applications, and in nonclinical applications, use of such optimistic priors can be appropriate and effective.

In the example, stopping would have been delayed, highlighting the question of whether it would be ethical to continue beyond a traditional stopping point due to prior beliefs that pyrimethamine would have a strong, prophylactic effect. Of course, if pyrimethamine had performed well, stopping would have been earlier than that based on traditional methods, also raising an ethical issue for some. And, if the priors had been pessimistic, but the data moderately optimistic, the Bayesian analysis would have tempered enthusiasm (delayed stopping) until convincing, positive, data had accrued.

The TOXO example shows one possible effect of using prior distributions in clinical evaluations designed to be definitive (Phase III). The example is based on the mixture that equally weighted the five prior distributions. The posterior distribution is also a mixture, but the posterior weights give more influence to the priors that are most compatible with the data, so there is some degree of automatic adjustment. Other options include separately monitoring with each prior and then at each monitoring point determine stopping based on the "majority rule" or requiring unanimity. Each choice will produce its own operating characteristic, so extensive simulations are needed to understand properties.

To summarize, in clinical trial monitoring prior information can have two main effects:

1. If prior information conflicts with the available data at the time of monitoring, it may suggest continuing the trial, at least until the next review. This situation can arise when the experts are more optimistic than the emerging data, as in the TOXO trial, or when they are more pessimistic. In either case, continuing the trial long enough to obtain sufficient evidence to convince the content experts may have an added benefit that the results would translate to practice relatively rapidly.

2. If prior information supports the available data at the time of monitoring, it may suggest terminating early, on the grounds that sufficient evidence now exists to make a decision.

An increasing number of trials are utilizing prior information (Abrams, Spiegelhalter, and Myles 2004; Berry et al. 2010), and the advent of such trials emphasizes the importance of prior judgments being made as carefully and as objectively as possible. Judgments should be sought from a sufficient number of content experts to properly represent the range of views. Determining the number and type of elicitees is, itself, an exercise in study design. Of course, prior distributions should be elicited carefully and rigorously, as described, for instance, in O'Hagan (2018).

4.2. Expert Judgment in Environmental Modeling

4.2.1. Background: UK Carbon Flux

This case study describes the estimation of a parameter in a complex environmental modeling problem. As a party to the Kyoto protocol, the United Kingdom is committed to specific target reductions in net emissions of greenhouse gases. Monitoring progress toward these targets is challenging, involving accounting for numerous sources of emissions. One potential mitigating factor is the ability of vegetation to remove carbon dioxide from the atmosphere, thereby acting as a "carbon sink." However, accounting for this is also extremely complex. During the day, through the process of photosynthesis, vegetation absorbs carbon dioxide and releases oxygen, using the carbon to build plant material. Photosynthesis requires sunlight, chlorophyll in green leaves and water and minerals gathered by the plant's roots, so the efficiency of carbon removal depends on factors such as weather, leaf surface area, and soil conditions. Conversely, carbon is released at night, carbon extracted from the atmosphere and converted to biomass will eventually be released as the plant ages and dies, and carbon in leaf litter is released by microbial action in the soil. The Sheffield Dynamic Global Vegetation Model (SDGVM) was built with mathematical descriptions of all these processes to predict net carbon sequestration due to vegetation (Woodward and Lomas 2004). For a given site, the model takes many inputs describing the type of vegetation cover and soil at the site, together with weather data, to estimate Net Biosphere Production (NBP), that is, the net decrease in atmospheric CO_2, from that site over a given time period.

This case study used SDGVM to estimate the total NBP for England and Wales in the year 2000. It is important to recognize that there is inevitably uncertainty about all the model inputs. Uncertainty about inputs induces uncertainty about model outputs, and the study sought to quantify the output uncertainty in the form of a probability distribution for the total NBP. Details are reported in Kennedy et al. (2008) and Harris, O'Hagan, and Quegan (2010).

Before considering the statistical aspects of this case study, we first note that considerable content expertise had already gone into the development of SDGVM. Based on the available scientific knowledge, expert judgments were made in choosing the structure of the model and the equations that describe each of its biological processes. Care went into these choices to ensure that they were reasonable and scientifically defensible. However, in such modeling, it is usually unrealistic to include every process to the greatest level of detail and complexity that is believed to be applicable. First, the more complex and detailed the model, the longer it will take to compute; for practical reasons, it may be necessary to compromise on complexity. Second, more complex models may be less robust and reliable in their predictions, because at the highest level of detail, there is often less scientific consensus about the equations and the parameters within them. In addition, models with a large number of parameters may suffer from overfitting (Hawkins 2004). Judgments about the optimal degree of complexity to achieve a computable, accurate, and reliable representation of the phenomenon being modeled often demand a particularly high level of expertise.

4.2.2. Input Distributions for the SDGVM

The model was run over 707 grid cells covering England and Wales. For each grid cell, given the land cover in that cell, and given input parameters describing the soil composition and properties of the vegetation types, the model was first "spun-up" for 600 years. That is, the model was initialized with a default set of inputs describing the state of the vegetation in terms of ages, heights, leaf density, etc., and the state of the soil in terms of moisture content, age and quantity of organic matter, etc. Then the model was run forward for 600 years using historic climate data at that site from 1400 to 2000, to stabilize the vegetation and soil at states representative of how that site would have been in 2000. The model was then run forward for one more year using weather data from 2000, and the NBP for the year was computed for each grid cell. The NBP for England and Wales in 2000 is the sum of the NBP values across all the grid cells.

Care was taken to quantify the uncertainty in the various inputs as described previously (O'Hagan 2012). Briefly, details are described below.

- *Soil composition*: A publicly available soil map for England and Wales (Bradley et al. 2005) was used to provide estimates of the soil composition in each grid cell. Because the map was created at a higher resolution than the grid cells used for SDGVM, figures were averaged over each grid cell to provide an estimate. The variance of the same figures over a grid cell, divided by the number of map points in the cell, was used to quantify uncertainty around the estimates for a grid cell. However, the variance was increased by a factor to represent (a) additional uncertainty in the map data and (b) spatial correlation within the cell. The decision to use an increased estimate of uncertainty in the model was based on an expert judgment on the part of the statisticians, in consultation with content experts.

- *Land cover*: A map of land cover was also publicly available (Haines-Young et al. 2000), derived from satellite observation. The original analysis reported in Kennedy et al. (2008) did not quantify uncertainty in land cover. However, unlike the soil map, the content experts felt that the uncertainties in land cover were large and there could be biases in the process by which land cover is inferred from the satellite readings. In a subsequent analysis (Cripps, O'Hagan, and Quaife 2013; Harris, O'Hagan, and Quegan 2010), a statistical model was built to quantify uncertainty in land cover maps derived from remote sensing. The analysis of England and Wales NBP in 2000 was then extended to account for the additional uncertainty. The method makes use of the "confusion matrix," which for the Haines-Young et al. (2000) map was given by Fuller et al. (2002). The confusion matrix is a contingency table based on a large survey of actual, ground-truth, land cover, and shows, for each ground-truth vegetation type, the frequency with which it was classified by the Haines-Young et al. (2000) map in each vegetation type. The statistical analysis required probabilistic inversion of the confusion matrix to derive, conditional on the satellite land cover, the probabilities of the various ground-truth cover. The careful expert judgments of statisticians and content experts are delineated in Cripps, O'Hagan, and Quaife (2013).
- *Vegetation properties*: SDGVM classifies land cover into plant functional types (PFT). For England and Wales, we used four PFTs—evergreen needleleaf trees, deciduous broadleaf trees, crops, and grassland. Each PFT is associated with a set of quantities, including maximum age, stem growth rate and leaf lifespan. However, most of these inputs were missing. Some properties had been estimated experimentally, but only for very few individual species within a given PFT. We therefore used expert elicitation to construct a probability distribution for each parameter. Elicitation is an area where it is particularly important to take care to avoid biases that commonly arise in subjective judgments of probabilities (Kynn 2008). Another article arising from the Symposium addresses issues related to elicitation in detail (O'Hagan 2018).
- *Weather*: Weather data, such as temperature, precipitation, and cloud cover, were available for each grid cell for every day in 2000. There are no doubt errors in these data, due not only to errors in the underlying daily measurements, but also to the fact that those measurements have been interpolated to produce the data at the level of grid cells. Nevertheless, it was felt that uncertainty in these inputs was relatively small and could not be quantified without adding extra assumptions. The decision not to trade (potentially unreasonable) assumptions for (potentially increased) precision for weather data was a judgment made jointly by statisticians and content experts.

4.2.3. Propagating the Input Uncertainty in the SDGVM

To propagate input uncertainty through a mechanistic model such as SDGVM, the simplest and most direct approach is by Monte Carlo. A large sample of random draws are made from the probability distributions of the inputs, and the model is run for each sampled input set. The resulting set of outputs is then a random sample from the output uncertainty distribution. However, like many models that are built to describe complex physical processes, the computational load in running the SDGVM model is substantial, and it would not have been feasible to propagate parameter uncertainty through the model in this way at even one of the grid cells. The problem of quantifying uncertainty in complex computer models has acquired the name uncertainty quantification (UQ), and various tools are available to enable uncertainty propagation. The analysis used Gaussian process emulation (O'Hagan 2006; Oakley 2016), which is probably the most popular UQ technique. Even so, it would not have been feasible to build emulators at every one of the 707 grid cells; 33 were chosen by content experts to represent, in their expert judgment, the range of conditions across England and Wales, and GP techniques were adapted to infer the magnitudes of output uncertainty at the other 674 cells. The computational techniques and accompanying statistical theory to manage the analysis are set out in Kennedy et al. (2008) and Gosling and O'Hagan (2006).

4.2.4. Results of the UK Carbon Flux Study

The content experts were certainly interested in knowing how much uncertainty in the total NBP would be induced by uncertainty in the inputs. Their instinctive approach to estimating the total NBP was to run the model just once with all of the inputs set to their expected values; we call their idea the plug-in estimate. Yet, the NBP output from SDGVM is a highly nonlinear function of its inputs. Therefore, the expected value of NBP, when we allow for uncertainty in the inputs, is not equal to the plug-in estimate. Statisticians incorporated the input uncertainty and computed the expected NBP for England and Wales in 2000 as 7.46 Mt C (megatonnes of carbon). By contrast, the plug-in estimate was 9.06 Mt C. Not only is this a substantial difference, but the standard deviation was estimated as 0.55 Mt C. Therefore, the total NBP was probably in the interval (6.36 Mt, 8.56 Mt C) and very likely to be less than the plug-in estimate.

The result was very surprising for the content experts. It seems that the explanation lies in their estimates of vegetation properties. In effect, the experts had estimated values that were approximately optimal for the plants to grow and absorb CO_2. Any deviation from these values led to lower NBP. For the total NBP to be even close to the plug-in estimate required all the parameter values to be close to their estimates, a joint event that had very low probability.

In this case study, the combination of expert judgments from both content experts and statisticians, applied with as much care, rigor, transparency, and objectivity as possible, led to a scientific result that certainly highlighted the role of expert judgment and the statistical quantification of uncertainty, and also prompted new questions regarding the accuracy of current methods for carbon budgeting, with important implications for the science of global climate change. For example, see Cripps, O'Hagan, and Quaife (2013). The cycle of scientific investigation was thereby renewed.

5. Conclusions

Expert scientific judgment involves carefully considered conclusions and decisions based on deep knowledge and skills. The use

of expert judgment is essential in and permeates all phases of scientific and policy studies. Expert knowledge is information; to ignore it or fail to obtain it incurs considerable opportunity costs. Judgments should be as objective as possible and based on data when available. Anything less is unscientific. Yet, deciding what data are relevant always involves degrees of judgment.

Written documentation that logs decisions is critical for informing stakeholders of how and in which study components (design, conduct, analysis, reporting, and decision-making) the principal judgments had impact. By "impact," we mean that such judgments led to decisions that directly affected (or had potential to alter or influence) any aspect of how the study was conducted and/or results interpreted.

Elicitation of expert judgments to produce probability distributions that represent uncertainty about model parameters can be conducted informally, but such judgments are easily affected by unconscious cognitive biases, such as overoptimism or failure to recall all relevant evidence. When such distributions form an important part of a scientific activity, the expert judgments should be elicited scientifically and as objectively as possible, minimizing relevant sources of bias. Doing so requires a carefully designed process, an elicitation protocol, as fully discussed by O'Hagan (2018).

It should be unsurprising that statisticians have essential roles as scientists, ideally serving as leaders or coleaders in all study aspects. Indeed, virtually all aspects of a study have statistical content, though almost no aspects are solely statistical. Consequently, we advise statisticians to more pro-actively promote the needs for statistical principles to permeate all stages of research studies. Furthermore, we advise research authorities, such as journal editors and funding agencies, to recommend or even require thorough collaboration with one or more experts in statistics throughout the duration of all projects.

The 2017 Symposium will not produce a single position document on statistical practice like the "The ASA Statement on p-Values" that resulted from the 2015 ASA Board meeting (Wasserstein and Lazar 2016). However, we echo the call in Gibson (2017) for statisticians to better advocate for the importance of their involvement throughout the scientific process. Finally, we applaud stakeholders, such as the National Institutes of Health (Collins and Tabak 2014) and the American Association for the Advancement of Science (McNutt 2014) in the USA, and National Institute for Health and Care Excellence (NICE 2012) and HM Treasury (2015) in the UK for leading calls for increased statistical rigor and understanding of uncertainty. We encourage additional funding agencies, journals, hiring and promotion committees, and others to join in the call for higher scientific standards, statistical and otherwise. Science in the twenty-first century and beyond deserves nothing less.

Funding

The authors gratefully acknowledge support, in part, from the following: NCB: N/A; TAL: NIH-NIAID, U19-AI089680; PMA2020 from the Bill & Melinda Gates Foundation; AO'H: N/A; and JP: NIA grant P30AG028716. Authorship order is alphabetical.

References

Abrams, K. R., Spiegelhalter, D., and Myles, J. P. (2004), *Bayesian Approaches to Clinical Trials and Health Care*, New York: Wiley. [61,65]

Anderson, H., and Hepburn, B. (2016), "Scientific Method," available at *https://plato.stanford.edu/archives/sum2016/entries/scientific-method*. [57]

Anscombe, F. J., and Aumann, R. J. (1963), "A Definition of Subjective Probability," *The Annals of Mathematical Statistics*, 34, 199–205. [60]

Berry, S. M., Carlin, B. P., Lee, J., and Müller, P. (2010), *Bayesian Adaptive Methods for Clinical Trials*, Boca Raton, FL: Chapman&Hall/CRC Press. [61,65]

Bradley, R., Bell, J., Gauld, J., Lilly, A., Jordan, C., Higgins, A., and Milne, R. (2005), "UK Soil Database for Modelling Soil Carbon Fluxes and Land Use for the National Carbon Dioxide Inventory," Report to Defra Project SP0511, Defra, London. [65]

Brownstein, N. C. (2018), "Perspective from the Literature on the Role of Expert Judgment in Scientific and Statistical Research and Practice," arXiv no. 1809.04721. [56,58]

Cairney, P., and Oliver, K. (2017), "Evidence-Based Policymaking Is Not Like Evidence-Based Medicine, So How Far Should You Go to Bridge the Divide Between Evidence and Policy?," *Health Research Policy and Systems*, 15, 35. [60]

Carlin, B. P., Chaloner, K., Church, T., Louis, T. A., and Matts, J. P. (1993), "Bayesian Approaches for Monitoring Clinical Trials, With an Application to Toxoplasmic Encephalitis Prophylaxis," *The Statistician*, 42, 355–367. [63]

Center for Open Science (2018), "Registered Reports," available at *https://cos.io/rr/*. [59]

Chaloner, K., Church, T., Louis, T. A., and Matts, J. P. (1993), "Graphical Elicitation of a Prior Distribution for a Clinical Trial," *The Statistician*, 42, 341–353. [63]

Collins, F. S., and Tabak, L. A. (2014), "NIH Plans to Enhance Reproducibility," *Nature*, 505, 612–613. [67]

Cox, D. R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society*, Series B, 34, 187–220. [62,63]

Cripps, E., O'Hagan, A., and Quaife, T. (2013), "Quantifying Uncertainty in Remotely Sensed Land Cover Maps," *Stochastic Environmental Research and Risk Assessment*, 27, 1239–1251. [66]

Curtis, A., Worley, S., Adamson, P., Chung, E., Niazi, I., Sherfesee, L., Shinn, T., and Sutton, M. (2013), "Biventricular Pacing for Atrioventricular Block and Systolic Dysfunction," *New England Journal of Medicine*, 368, 1585–1593. [61,62]

Easterbrook, P. J., Gopalan, R., Berlin, J., and Matthews, D. R. (1991), "Publication Bias in Clinical Research," *The Lancet*, 337, 867–872. [60]

Ellenberg, S. S., Fleming, T. R., and DeMets, D. L. (2003), *Data Monitoring Committees in Clinical Trials: A Practical Perspective*, Chichester: Wiley. [59]

Finfer, S., and Bellomo, R. (2009), "Why Publish Statistical Analysis Plans," *Critical Care and Resuscitation*, 11, 5–6. [59]

Franco, A., Malhotra, N., and Simonovits, G. (2014), "Publication Bias in the Social Sciences: Unlocking the File Drawer," *Science*, 345, 1502–1505. [60]

Fuller, R., Smith, G., Sanderson, J., Hill, R., Thomson, A., Cox, R., Brown, N., Clarke, R., Rothery, P., and Gerard, F. (2002), "Countryside Survey 2000 Module 7. Land Cover Map 2000," Final Report, CSLCM/Final. [66]

Garland, T., Jr. (2016), "Scientific Method as an Ongoing Process," available at *https://en.wikipedia.org/w/index.php?title=Scientific_method&oldid=822947033*. [57]

Gelman, A., and Hennig, C. (2017), "Beyond Subjective and Objective in Statistics," *Journal of the Royal Statistical Society*, Series A, 180, 967–1033. [56,61]

Gibson, E. W. (2017), "Leadership in Statistics: Increasing Our Value and Visibility," *The American Statistician*, DOI: 10.1080/00031305.2017.1336484. [67]

Gosling, J. P., and O'Hagan, A. (2006), "Understanding the Uncertainty in the Biospheric Carbon Flux for England and Wales," Technical Report 567/06, University of Sheffield, available at *http://tonyohagan.co.uk/academic/pdf/UUCF.pdf*. [66]

Grice, J. W.(2017), "Comment on Locascio's Results Blind Manuscript Evaluation Proposal," *Basic and Applied Social Psychology*, 39, 254–255. [60]

Haines-Young, R., Barr, C., Black, H., Briggs, D., Bunce, R., Clarke, R., Cooper, A., Dawson, F., Firbank, L., Fuller, R., Furse, M. T., Gillespie, M. K., Hill, R., Hornung, M., Howard, D. C., McCann, T., Morecroft, M. D., Petit, S., Sier, A. R. J., Smart, S. M., Smith, G. M., Stott, A. P., Stuart, R. C., and Watkins, J. W. (2000), "Accounting for Nature: Assessing Habitats in the UK Countryside," Natural Environment Research Council and

Centre for Ecology and Hydrology, Department of the Environment, Transport and the Regions, available at *https://trove.nla.gov.au/work/17604300*. [66]

arris, K., O'Hagan, A., and Quegan, S. (2010), "The Impact of Satellite-Derived Land Cover Uncertainty on Carbon Cycle Calculations," available at *tonyohagan.co.uk/academic/pdf/CTCDPaper_v6.pdf*. [65,66]

Hawkins, D. M.(2004), "The Problem of Overfitting," *Journal of Chemical Information and Computer Sciences*, 44, 1–12. [65]

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., and Jennions, M. D. (2015), "The Extent and Consequences of P-Hacking in Science," *PLoS Biology*, 13, e1002106. [59]

HM Treasury (2015), "The Aqua Book: Guidance on Producing Quality Analysis for Government," *HM Government, London, UK, nd https://www. gov. uk/government/publications/theaqua-book-guidance-on-producing-quality-analysis-for-government (accessed July 10, 2017)*. [60,67]

Hyman, M. R.(2017), "Can 'Results Blind Manuscript Evaluation' Assuage 'Publication Bias'?," *Basic and Applied Social Psychology*, 39, 247–251. [60]

Jacobson, M., Besch, C., Child, C., Hafner, R., Matts, J., Muth, K., Wentworth, D., Neaton, J., Abrams, D., Rimland, D., and Perez, G. (1994), "Primary Prophylaxis With Pyrimethamine for Toxoplasmic Encephalitis in Patients With Advanced Human Immunodeficiency Virus Disease: Results of a Randomized Trial," *The Journal of Infectious Diseases*, 169, 384–394. [62]

Kennedy, M., Anderson, C., O'Hagan, A., Lomas, M., Woodward, I., Gosling, J. P., and Heinemeyer, A. (2008), "Quantifying Uncertainty in the Biospheric Carbon Flux for England and Wales," *Journal of the Royal Statistical Society*, Series A, 171, 109–135. [65,66]

Kerr, N. L. (1998), "Harking: Hypothesizing After the Results Are Known," *Personality and Social Psychology Review*, 2, 196–217, PMID: 15647155. [59]

Kynn, M. (2008), "The 'Heuristics and Biases' Bias in Expert Elicitation," *Journal of the Royal Statistical Society*, Series A, 171, 239–264. [66]

Vange, L. M. (2014), "The Role of Statistics in Regulatory Decision Making," *Therapeutic Innovation & Regulatory Science*, 48, 10–19. [60]

Lindsay, R. M., and Ehrenberg, A. S. (1993), "The Design of Replicated Studies," *The American Statistician*, 47, 217–228. [60]

Locascio, J. J. (2017a), "Rejoinder to Responses on 'Results-Blind Publishing'," *Basic and Applied Social Psychology*, 39, 258–261. [60]

Locascio, J. J. (2017b), "Results Blind Science Publishing," *Basic and Applied Social Psychology*, 39, 239–246. [60]

Marks, M. (2017), "Commentary on Locascio," *Basic and Applied Social Psychology*, 39, 252–253. [60]

McNutt, M. (2014), "Journals Unite for Reproducibility," *Science*, 346, 679–679. [67]

National Academies of Sciences, Engineering, and Medicine, Division of Behavioral and Social Sciences and Education, Committee on National Statistics, Panel on Improving Federal Statistics for Policy and Social Science Research Using Multiple Data Sources and State-of-the-Art Estimation Methods, Harris-Kojetin, B. A., Groves, R. M., eds. (2017), *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps*, Washington, DC: National Academies Press. [60]

National Research Council, Division on Engineering and Physical Sciences, Board on Mathematical Sciences and Their Applications, and Committee on Mathematical Foundations of Verification, Validation, and Uncertainty Quantification (2012a), *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*, Washington, DC: National Academies Press. [60]

National Research Council, Division of Behavioral and Social Sciences and Education, Division on Engineering and Physical Sciences, Committee on National Statistics, Computer Science and Telecommunications Board, and Panel on Communicating National Science Foundation Science and Engineering Information to Data Users (2012b), *Communicating Science and Engineering Data in the Information Age*, Washington, DC: National Academies Press. [60]

NICE (2012), "The Guidelines Manual: Process and Methods," available at *https://www.nice.org.uk/process/pmg6/*. [67]

Nussbaum, B. (2017), "Introductory Remarks: JSM 2017 Presidential Invited Address," available at *https://ww2.amstat.org/meetings/jsm/2017/webcasts/index.cfm*. [58]

Oakley, J. (2016), "Introduction to Uncertainty Quantification and Gaussian Processes," available at *http://gpss.cc/gpuqss16/slides/oakley.pdf*. [66]

O'Brien, P. C., and Fleming, T. R. (1979), "A Multiple Testing Procedure for Clinical Trials," *Biometrics*, 35, 549–556. [62]

O'Hagan, A. (2006), "Bayesian Analysis of Computer Code Outputs: A Tutorial," *Reliability Engineering & System Safety*, 91, 1290–1300. [66]

——— (2012), "Probabilistic Uncertainty Specification: Overview, Elaboration Techniques and Their Application to a Mechanistic Model of Carbon Flux," *Environmental Modelling & Software*, 36, 35–48. [65]

O'Hagan, A. (2018), "Expert Knowledge Elicitation: Subjective, But Scientific," [submitted]. [65,66,67]

Oliver, K. A., and de Vocht, F. (2017), "Defining 'Evidence' In Public Health: A Survey of Policymakers' Uses and Preferences," *European Journal of Public Health*, 27, 112–117. [60]

Oliver, K., Lorenc, T., and Innvær, S. (2014), "New Directions in Evidence-Based Policy Research: A Critical Analysis of the Literature," *Health Research Policy and Systems*, 12, 34, DOI: 10.1186/1478-4505-12-34. [60]

Ott, M. G. (1991), "Importance of the Study Protocol in Epidemiologic Research," *Journal of Occupational Medicine*, 33, 1236–1239. [59]

Oxford University Press (2018a), "Judgment," available at *https://en.oxforddictionaries.com/definition/judgment*. [57]

——— (2018b), "Knowledge," available at *https://en.oxforddictionaries.com/definition/knowledge*. [57]

——— (2018c), "Opinion," available at *https://en.oxforddictionaries.com/definition/opinion*. [57]

——— (2018d), "Scientific Method," available at *https://en.oxforddictionaries.com/definition/scientific_method*. [57]

Pocock, S. J. (2006), "Current Controversies in Data Monitoring for Clinical Trials," *Clinical Trials*, 3, 513–521. [59]

Shih, J. (1995), "Sample Size Calculation for Complex Clinical Trials With Survival Endpoints," *Controlled Clinical Trials*, 16, 395–407. [61]

Wasserstein, R. L., and Lazar, N. A. (2016), "The ASA's Statement on *p*-Values: Context, Process, and Purpose," *The American Statistician*, 70, 129–133. [67]

Woodward, F. I., and Lomas, M. R. (2004), "Vegetation Dynamics—Simulating Responses to Climatic Change," *Biological Reviews*, 79, 643–670. [65]

EXHIBIT 16

UNITED STATES BANKRUPTCY COURT
DISTRICT OF MASSACHUSETTS

In re: TELEXFREE, LLC, TELEXFREE, INC., and TELEXFREE FINANCIAL, INC., Debtors	Chapter 11 Cases 14–40987–MSH 14–40988–MSH 14–40989–MSH Jointly Administered
STEPHEN B. DARR, TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Plaintiff, v. BENJAMIN ARGUETA, A REPRESENTATIVE OF A CLASS OF DEFENDANT NET WINNERS, Defendants.	Adversary Proceeding No. 16-4006
STEPHEN B. DARR AS TRUSTEE OF THE ESTATES OF TELEXFREE, LLC, TELEXFREE, INC. and TELEXFREE FINANCIAL, INC., Plaintiff, v. MARCO PUZZARINI, A REPRESENTATIVE OF A CLASS OF DEFENDANT NET WINNERS, Defendants.	Adversary Proceeding No. 16-4007

SUPPLEMENTAL REBUTTAL EXPERT REPORT OF JOSHUA W. DENNIS

Respectfully submitted this 14th day of April, 2023

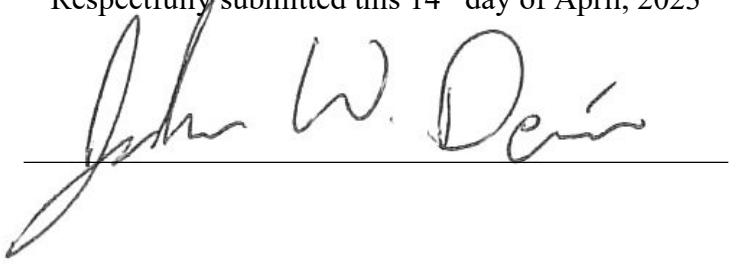


Table of Contents

I. INTRODUCTION 1

II. BACKGROUND 1
- A. TelexFree User Accounts and Participants 2
- B. TelexFree Transactions 4

III. ASSIGNMENT 6

IV. INFORMATION CONSIDERED 6

V. UNDERSTANDING OF TRUSTEE'S METHODOLOGY FOR DETERMINING NET WINNERS AND NET LOSERS 8
- A. Freer Aggregation Process 12
  - i. Step 1: Selection and Transformation of Fields 12
  - ii. Step 2: Cleansing of Fields 13
  - iii. Step 3: Deterministic Linkage 13
  - iv. Step 4: Blocking 14
  - v. Step 5: Calculation of Gammas 15
  - vi. Step 6: Calculation of Match Probabilities 16
  - vii. Steps 7-8: Fixed Threshold and Dynamic Clustering 16
  - viii. Step 9: Convert Clusters of Account Groupings into Clusters of User Accounts 18
  - ix. Step 10: Tabulate Total Net Equity for Each Cluster 18
- B. Martin Net Equity Calculation 19

VI. SUMMARY OF OPINIONS 20

VII. ANALYSIS AND FINDINGS 23
- A. The Freer Aggregation Process is Unreliable 23
  - i. Dr. Freer's Foundational Assumptions Regarding the Special Role of the Name Field is Unsupported and Counter to Case Facts 23
  - ii. Dr. Freer Failed to Appropriately Consider the Quality of the Data Utilized in His Aggregation Process 29
  - iii. Dr. Freer's Results are Internally Inconsistent 37
  - iv. The Results of the Freer Aggregation Process are Inconsistent with his Stated Methodology and Requirements 40
  - v. The Freer Aggregation Process Results in Inconsistent and Speculative Outcomes 45
  - vi. The Only Example Cluster in the Freer Report Contains Questionable Aggregations 47
  - vii. Aggregation Errors Can Result in a Significant Overstatement of Participant Net Equity 48
  - viii. Dr. Freer has Failed to Deliver the Full Scope of His Engagement 52

B. Mr. Martin's Calculation of Net Equity is Unreliable 53
i. The Net Equity Calculation Fails to Account for Participant Investments and Receipts Related to Transfers of Credit 53
ii. The Amount of Cash Invested/Received by Participants as Part of Triangular Transactions or Credits Transfers Can Differ from Amount of Credits Recorded in SIG 58

VIII. CONCLUSION 62

List of Exhibits

Exhibit Number	Description
1	Curriculum Vitae of Joshua W. Dennis
2	Documents Considered
3	User Accounts and Participants by Entity and Type
4	Summary of TelexFree Net Equity by Tier and Transaction Type
5	Summary of TelexFree User Accounts and Participants by Tier
6	Summary of TelexFree Top Net Losers Matched to Net Winners
7	Top 100 TelexFree Net Losers
8	Top 100 TelexFree Net Winners
9.1-9.5	Top Net Loser and Matched Net Winner User Account Detail
10	Frantz Balan & Lusette Balan User Account Detail
11	Alphabet User Account Detail
12	Billybob Arbonne User Account Detail
13	Lopes & Carreiro User Account Detail
14	Zaglia & Ragolia User Account Detail
15	Rejane & Paulo Marinho User Account Detail
16	Maria Neves User Account Detail
17	Sann Devaconcelos User Account Detail
18	Oscar Sosa User Account Detail
19	Claimed Account Aggregations
20.1-20.3	Example Credit Transfer Detail

I. INTRODUCTION

1. I, Joshua W. Dennis, am a Partner in the Boston office of StoneTurn Group LLP ("StoneTurn"). I provide financial and economic consulting services – including the analysis of complex data – to attorneys and companies. I have more than 15 years of experience working with clients and counsel on matters requiring complex data analysis, such as economic damages, forensic accounting, compliance monitoring, and valuation engagements.

2. In addition to experience in data collection, mining, anomaly detection, and reporting of transactional and trading data, I have extensive expertise in financial modeling to quantify and assess lost profits, diminution of business value, and other areas of economic loss. My analyses also frequently include evaluations related to data sufficiency and integrity testing, integration of disparate datasets, data exploration, and visual modeling.

3. In conjunction with my experience related to the quantification of economic damages, I have gained expertise through my certifications and trainings.

4. A copy of my current curriculum vitae, which summarizes my qualifications and professional experience, is attached as **Exhibit 1**.

5. StoneTurn charges $600 per hour for time I spend consulting and developing expert opinions, as well as time that may be spent testifying related to my opinions. Neither my, nor my staff's, nor StoneTurn's, compensation is affected in any way by either the conclusions I have reached or the outcome of this lawsuit.

II. BACKGROUND

6. On June 6, 2014, the Plaintiff, Stephen Darr, was appointed Chapter 11 Trustee ("Trustee") of the bankruptcy estates of TelexFree, LLC, TelexFree, Inc., and TelexFree Financial, Inc., (collectively, the "Company," "TelexFree," or "Debtors").[1] The Trustee alleges that TelexFree

[1] Amended Complaint, Adversary Proceeding No. 16-04006, April 14, 2016 ("Amended Complaint") at 3.

operated a massive Ponzi and pyramid scheme that defrauded hundreds of thousands of individuals in numerous countries out of billions of dollars.[2]

7. The Trustee alleges that roughly 15,000 individuals or entities residing in the U.S. are "Net Winners"[3] who illegally profited from the fraud. These 15,000 Net Winners comprise a class of Defendants ("U.S. Net Winner Class").[4] Additionally, the Trustee alleges that roughly 80,000 individuals or entities residing outside the U.S. are "Net Winners" who illegally profited from the fraud.[5] These 80,000 Net Winners comprise a class of defendants ("Non-U.S. Net Winner Class," collectively with the U.S. Net Winner Class, hereinafter referred to as the "Net Winner Class").

A. TelexFree User Accounts and Participants

8. I understand that an individual could join TelexFree (as a "Participant") by purchasing either a membership plan or individual Voice over Internet Protocol ("VoIP") phone packages. However, VoIP phone packages, which cost $49.90 per package,[6] constituted only a minor portion of the Company's business, as TelexFree primarily generated returns from the membership fees associated with recruiting new Participants.[7]

9. Each time a Participant purchased a membership plan or VoIP package, the Participant established a "User Account" with the Company and was directed to provide a name, address, email address, phone numbers, passcodes, and certain other information.[8]

10. TelexFree used the User Accounts to track the activity of Participants, including "Credits" issued to Participants for becoming a "Promoter" of the business by placing advertisements on certain websites, as well as recruiting additional Participants. TelexFree also used the User Accounts to track certain other activities, such as bonuses, commissions, payments of invoices, use of Credits to satisfy invoices, transfers of Credits, and cash receipts. A Participant could

[2] Amended Complaint at 21.

[3] For the specific definition of Net Winner, see Section V of this report.

[4] Class Certification Order and Approval of Class Counsel, Adversary Proceeding No. 16-4006, October 6, 2016 ("Class Cert. 16-4006") at 3.

[5] Class Certification Order and Approval of Class Counsel, Adversary Proceeding No. 16-4007, August 3, 2017 ("Class Cert. 16-4007") at 2.

[6] Motion by Chapter 11 Trustee for Entry of Order Finding that Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, October 7, 2015 ("Motion for Entry of Order that Debtors Engaged in Ponzi Scheme") at 6.

[7] Amended Complaint at 21.

[8] Amended Complaint at 22-23.

establish more than one User Account, and in practice many Participants had multiple User Accounts, sometimes numbering in the thousands.[9]

11. The User Accounts and related data were stored in TelexFree's Information Management System, SIG.[10] The SIG database contained data for approximately 17 million User Accounts, of which approximately 15 million were associated with at least one paid invoice.[11]

12. An affiliated entity of TelexFree, Ympactus Commercial Ltda. ("Ympactus"), reportedly operated a similar scheme in Brazil, and also stored its data in SIG. Ympactus was seized by the Brazilian authorities in June 2013.[12] The currency field, which identified whether an invoice was denominated in Brazilian reais ("Reais" or "BRL") or United States dollars ("USD"), was utilized to separate out the Ympactus data from the TelexFree data within SIG. User Accounts with Brazilian contact information were denominated in Reais, and thus associated with Ympactus. Conversely, invoices associated with User Accounts with non-Brazilian contact information were denominated in USD and assumed to be associated with TelexFree. Approximately 11 million User Accounts with at least one paid invoice remained after excluding the User Accounts associated with Ympactus.[13]

13. The Company provided Participants with two membership plan options, each of which allowed for the generation of Credits. A one-year membership plan could be purchased by Participants for $339 (for an "AdCentral Plan") or $1,425 (for an "AdCentral Family Plan"), as follows:[14]

- **AdCentral Plan –** $339 for a one-year contract ($50 membership fee plus $289 contract fee). This contract entitled the User Account holder to sell ten VoIP Packages, for which a Participant could receive a commission if the packages were sold. Participants were required to place one internet ad per day and, for each complete week in which the Participant placed the required ads, he or she was entitled to one additional VoIP Package, which could be sold or exchanged

[9] Amended Complaint at 22-23. Also, see e.g., **Exhibit 7** and **Exhibit 8**.

[10] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 11. ("SIG stands for Sistemas de Informacoes Gerenciais, which is from the Portuguese language and translates roughly to 'Information Management System.'")

[11] Expert Report of Timothy J. Martin, April 13, 2018 ("Martin Report") at 11. Also, see **Exhibit 3**.

[12] Martin Report at 7; Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 14-16.

[13] Martin Report at 11. Also, see **Exhibit 3**.

[14] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 7-8.

for $20 in credits with the Debtors. Thus, Participants who posted the required ads were eligible to receive $20 per week for 52 weeks, for a total annual return of $1,040.

- **AdCentral Family Plan –** $1,425 for a one-year contract ($50 membership fee plus $1,375 contract fee). This contract entitled the User Account holder to sell fifty VoIP Packages, for which a Participant could receive a commission if the packages were sold. Participants were required to place five internet ads per day and, for each week in which the Participant placed the required ads, he or she was entitled to five additional VoIP Packages, which could be sold or exchanged for $100 in credits with the Debtors. Thus, the Participants who posted the required ads were eligible to receive $100 per week for 52 weeks, for a total annual return of $5,200.

B. TelexFree Transactions

14. I understand that Invoices issued by the Company to Participants for the purchase of a membership plan could be satisfied in at least the following ways: (1) a Participant could satisfy the invoice by payment in cash directly to the Company (a "Direct Transaction"); or (2) a Participant could satisfy the TelexFree invoice in other ways, such as by having a Participant, usually the recruiting Participant, pay the invoice by redeeming his or her accumulated Credits, and TelexFree would mark the invoice as paid and reduced the recruiting Participant's Credits by a corresponding amount (a "Triangular Transaction"). The Trustee alleges that Participants frequently brought on new recruits through Triangular Transactions as one means to monetize their accumulated Credits, often instead of redeeming Credits directly with the Company for cash.[15]

15. I understand that while invoices associated with the sale of membership plans or VoIP Packages had a face value of approximately $3.07 billion, only $360 million, or roughly twelve percent (12%) of that amount, was paid in cash to TelexFree; the balance of these invoices was satisfied using Participants' Credits.[16]

16. In the case of a Direct Transaction where a Participant satisfies his or her own invoice by payment in cash to the Company, the process generally worked as follows:[17]

[15] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 9-10; Memorandum in Support of Proposed Supplemental Order Respecting Motion by Chapter 11 Trustee for Entry of Order Finding that Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, January 19, 2016 ("Support of Proposed Supplemental Order") at 3-4.

[16] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 10.

[17] Amended Complaint at 23-24.

- The new Participant established an online account.
- The Company's database recorded the data entered by the new Participant and identifies the recruiting Participant, and assigned an identification number to the newly created User Account.
- The Company recorded the purchase, issued a numbered invoice, and marked the invoice as 'pending.'
- The new Participant could pay the invoice by cash, check, cashier's check, money order, wire transfer, or through a third-party online payment processing account. When the invoice was paid, the Company would update the invoice.
- The new Participant could then start building a pyramid underneath the newly created User Account by recruiting other members (or by purchasing new User Accounts themself or through associates), which in turn generated bonuses and commissions for the new Participant.

17. In the case of a Triangular Transaction where a new User Account is opened using accumulated Credits in another User Account, the process generally worked as follows:[18]

- A recruited Participant created a new online account, or used his or her existing account information, to establish a new User Account.
- The Company's database recorded the details entered by the recruited Participant and assigned an identification number to the new User Account.
- The Company recorded the purchase, issued an invoice to the new User Account, and marked the invoice as 'pending.'
- The new recruited Participant forwarded the invoice to the recruiting Participant or to a third-party Participant, who satisfied the invoice with accumulated Credits in an existing User Account.
- The Trustee alleges that the newly recruited Participant always paid the invoice amount in cash to the recruiting Participant (in those cases where there were two separate Participants involved). However, no evidence of such cash payments exists in SIG.

[18] Amended Complaint at 24.

III. ASSIGNMENT

18. Milligan Rona Duran & King LLC ("Counsel") retained StoneTurn[19] to assess the methodology developed and employed by the Trustee in calculating the supposed net funds received by members of the Net Winner Class as a result of their participation in TelexFree.

19. My opinions are based on consideration of case-specific facts and circumstances, as well as on my education, training, and experience. I am not offering any legal opinions.

20. While my work on this matter is ongoing, this report summarizes my current opinions given the information available to date. I may consider any additional materials if produced subsequent to the issuance of this report and modify or supplement my analysis or opinions as necessary. I also reserve the right to address any comments that may be received from an expert retained by the Trustee related to my work in this matter, as outlined in this report and associated exhibits.

21. In connection with my anticipated trial testimony in this action, I may create, from various documents produced in this litigation, demonstrative exhibits which refer or relate to the matters discussed in this report, or in my deposition testimony (if I were to be deposed). I have not yet created any such exhibits as of the date of this report.

IV. INFORMATION CONSIDERED

22. I considered information from a variety of sources, including (i) information produced or filed by the parties in this proceeding, (ii) information provided by relevant parties, and (iii) information from independent research.

23. Specifically, I considered information produced in these proceedings, such as legal filings.

24. Additionally, I considered information provided by relevant parties, such as reports submitted by experts retained on behalf of the Trustee in this matter, as well as any documents considered or cited therein. This includes the Expert Report of Timothy J. Martin, dated April 13,

[19] As used herein, other than references to my education and experience, "I" ,"We" or "StoneTurn" shall mean either me personally or those StoneTurn professionals acting under my supervision. Also, "My", "Our", and "Us" shall refer to actions taken by me personally or by those StoneTurn professionals acting under my direct supervision. All of the conclusions and opinions expressed herein are my own.

2018 (the "Martin Report")[20] and the Expert Report of Dr. Cameron E. Freer, dated April 29, 2022 (the "Freer Report"). I also considered the following data sources and documents:

- TelexFree Sistemas de Informacoes Gerenciais ("SIG") data tables, the management information systems used by TelexFree to track the activity for Participants by User Account.

- TelexFree data tables utilized and modified by Mr. Martin as part of his Net Equity computation[21] ("Huron Data Tables"), including:[22]

    - **rep_id_summary:** This table contains User Account and aggregation information developed by Mr. Martin for the Trustee.

    - **claim_statement:** This table contains the transactions extracted from the TelexFree database, which were then used to determine and calculate Net Equity.

    - **telexfree_accounts:** This table contains the same participant account and aggregation information, but excludes the Ympactus accounts.

    - **telexfree_transactions:** This table contains the transactions extracted from the TelexFree database to calculate Net Equity, but includes only transactions related to Participant Net Equity.

- A summary description prepared by Mr. Martin of the components used to calculate Net Equity, the mechanism for computing Net Equity using the TelexFree database, and the method used to aggregate User Accounts; [23]

- TelexFree data tables considered by Dr. Freer ("Freer Data Tables"), as provided to him by Huron, including:[24]

    - **representante:** This table contains a unique record for each User Account, as well as related information entered by Participants such as a name, address, email address and phone number ("Account Table").

[20] Mr. Martin also issued a Reply Report on September 11, 2020 ("Martin Reply Report").

[21] For the specific definition of Aggregation Process and Net Equity, see Section V of this report, below.

[22] Transmittal letter from Andrew Lizotte to Ilyas Rona, January 5, 2017.

[23] Transmittal letter from Andrew Lizotte to Ilyas Rona, January 5, 2017; and Telexfree Net Equity Calculation and Aggregation.pdf.

[24] See e.g., Martin Report at 9.

- **fatura:** This table contains a unique record for each invoice generated by TelexFree/Ympactus, as well as related details such as the type of plan purchased, the amount, and how the invoice was satisfied ("Invoice Table").
- **transferencia:** This table contains information about each transfer of Credits within TelexFree's system, as well as requests to redeem and withdraw funds ("Transfers Table")
- **bonus:** This table primarily contains information about each accretion of TelexFree Credits into a Participant's User Account ("Bonus Table").

- Data tables containing Dr. Freer's concluded aggregation of User Accounts to a Participant.[25]
- Programming code and other related documentation underlying Dr. Freer's aggregation analysis.
- Electronic Proof of Claim ("ePOC") data tables containing claims and other related information provided by Participants through a portal developed by the Trustee and his advisors.

25. Lastly, I also considered information obtained through independent research, including trade press articles and other publicly available sources.

26. The information I considered is listed in **Exhibit 2**, as well as the footnotes and other citations in this report. I reserve the ability to consider further discovery in this case, including additional expert reports, transcripts of depositions, and documents produced or made available.

V. UNDERSTANDING OF TRUSTEE'S METHODOLOGY FOR DETERMINING NET WINNERS AND NET LOSERS

27. The Trustee initially retained Huron Consulting Services LLC ("Huron"), including Timothy Martin, a Managing Director in the Business Advisory Practice of Huron, to act as a financial and accounting advisor to the Trustee.[26] As part of this retention, Huron and Mr. Martin assisted in the development of a methodology that attempts to identify and quantify TelexFree "Net Winners" and "Net Losers," as defined therein.

[25] aggregation_output.csv; and aggregation_summary.csv. I note that Dr. Freer's aggregation_output.csv file appears to include numerous duplicate records (i.e., where a given User Account appears more than once).

[26] Martin Report at 3-4.

28. Mr. Martin considered a Participant to be a "Net Winner" when the net amount of funds received by the Participant exceeds the amounts expended. Conversely, Mr. Martin considers a Participant to be a "Net Loser" when the net amount of funds expended exceeds the amounts received. The nominal dollar value, whether a net gain or net loss, associated with a given Participant represents their "Net Equity," which I understand the Court in this matter has defined as follows:[27]

> The amount invested by the Participant into the Debtors' scheme, including amounts paid [by the Participant] pursuant to Triangular Transactions, less amounts received by the Participant from the Debtors' scheme, including amounts received [by the Participant] pursuant to Triangular Transactions.

29. I understand that the Court's definition of Net Equity does not include the value of labor for the time spent by Participants or any expense incurred by Participants (e.g., rent, conferences, travel, etc.) from their involvement in TelexFree.

30. In order to analyze the data and determine who is a Net Winner or Net Loser (as defined above), Mr. Martin used a data-matching process to identify owners of TelexFree accounts ("User Accounts") and aggregate multiple User Accounts by Participant ("Aggregation Process"). Specifically, Mr. Martin developed a 13-step Aggregation Process that primarily, if not entirely, utilized deterministic record linkage methods.

31. In response to Mr. Martin's Net Equity Calculation and his related Aggregation Process, I submitted a Rebuttal Export Report on July 31, 2020 ("Initial Dennis Report") that assessed the methodologies and assumptions employed by Mr. Martin. In connection with the Initial Dennis Report and my subsequent testimony as part of an evidentiary hearing,[28] the Court on June 22, 2021 filed a Memorandum of Decision on Class Defendants' Motion to Exclude Expert Witness Testimony of Timothy Martin ("Martin Decision").

32. The Martin Decision laid out the certain legal and factual bases that underpinned Mr. Martin's analysis and opinions, particularly as it related to Mr. Martin's Aggregation Process, and

[27] Supplemental Order Respecting Motion by Chapter 11 Trustee for Entry of Order Finding that Debtors Engaged in Ponzi and Pyramid Scheme and Related Relief, January 26, 2016 ("Supplemental Order"), at 2-3.

[28] Transcript of Video Evidentiary Hearings: Determining the Admissibility and Presumptive Effect of the Aggregation Methodology in Determining Net Winners before the Honorable Melvin S. Hoffman, J.U.S.B.C. November 23-24, 2020.

concluded that "[t]he trustee has not shown by a preponderance of the evidence the reliability of [Mr. Martin's] opinion as to the selection and application of his method for aggregating user accounts to determine in these adversary proceedings the identities and gains of the net winners in the TelexFree scheme. Thus, the expert's opinion is inadmissible."[29]

33. With respect to other issues raised by Defendants related more directly to Mr. Martin's determination of Net Equity, the Martin Decision stated:[30]

> The defendants also raise arguments beyond simply addressing the reliability of Mr. Martin's selection and application of an aggregation methodology, including arguments that relate to Mr. Martin's assumptions and decisions after the aggregation process was complete, when he set out to calculate the gains and losses (net equity) of each alleged participant. Having determined that the reliability of Mr. Martin's aggregation methodology has not been established and thus his expert opinion cannot be admitted, it is unnecessary to address the defendants' additional arguments, which they may choose to raise in the future, if appropriate.

34. As a result of Mr. Martin's exclusion, in or around September 2021 the Trustee retained a second expert, Dr. Freer.[31] Specifically, the Trustee retained Borelian Corporation, of which Dr. Freer is a Principal, to:[32]

> [D]evelop and implement an aggregation methodology for grouping the User Accounts associated with each Participant. Specifically, Borelian was engaged to: (1) develop an aggregation methodology that creates one "cluster" of User Accounts for each Participant such that each User Account appears in precisely one cluster; and (2) provide the set of clusters that results from implementing this methodology on the 10,987,617 User Accounts attributable to TelexFree.

35. Absent from Dr. Freer's assignment is any discussion related to an independent calculation or conclusion as to Net Equity. Instead, it appears Dr. Freer was only tasked with aggregating User Accounts ("Freer Aggregation Process"), and was instructed by counsel for the Trustee to adopt Mr. Martin's concluded Net Equity amounts associated with each User Account.[33] Given that the

[29] Martin Decision at 39-40.

[30] Martin Decision at 40.

[31] Motion by Liquidating Trustee to Schedule Case Management Conference Respecting Supplementation of Expert Reports, September 7, 2021 at 3; and Freer Report at 2.

[32] Freer Report at 2.

[33] BC002505-510 at 505-506 ("I think you've said this before, but we wanted to confirm that the Net Equity calculation by Huron is not under review here. It is our understanding that the Court has approved the Net Equity formula, and

Court appears to have excluded Mr. Martin's opinion in its entirety, it is unclear if, or to what extent, Dr. Freer is able to rely on Mr. Martin's work product, opinions or conclusions, in whole or in part, in forming his own opinions. If it is concluded that Mr. Martin's opinions and related work product should, in fact, be excluded in relevant part, then Dr. Freer's conclusions as to Net Winners and Net Losers, which necessarily utilizes Mr. Martin's concluded Net Equity figures, would likewise have no independent basis. I am not aware of any experts retained by the Trustee at this time other than Mr. Martin who have independently performed a Net Equity calculation as part of this Adversary Proceeding. In any event, Dr. Freer did not perform his own net equity calculations as he admitted they were supplied by Huron.[34] Dr. Freer also confirmed that he did not form any opinions "in detail" regarding whether Martin's calculations were correct or appropriate because he "didn't investigate enough to tell."[35]

36. Notwithstanding this uncertainty, my report addresses: (i) fundamental flaws that existed within Mr. Martin's Aggregation Process and still persist within the Freer Aggregation Process; (ii) new issues Dr. Freer introduced into the Aggregation Process; and (iii) the unreliability of the Net Equity amounts calculated in the now-excluded Martin Report, which were still adopted by Dr. Freer.

---

that Huron has implemented this formula to create the net_equity field for each account in the database. We are hoping that we can continue to use this field without further investigation, even as we examine which accounts should be aggregated…Your assumption is correct. The principal inquiry here is reaching a resolution on the user accounts to be aggregated. As the defendants have pointed out, they may raise issues with respect to the calculation of the net equity, but that is an issue for another day (and is in any event presently outside of your scope."); and Freer Report at 22-23 ("Based on information provided to Borelian from Huron in the form of a rep_id_summary table (the 'Summary Table'), I filtered (via the entity field) the User Account records considered for aggregation down to the 10,987,617 User Accounts attributable to TelexFree and associated with at least one paid invoice. The Summary Table also contains (in the net_equity field) the Net Equity for each User Account as calculated by Huron, which I used in Step 10 to tabulate net equity after clustering.").

[34] Freer Deposition at 45.

[35] Freer Deposition at 48.

A. Freer Aggregation Process

37. The Freer Aggregation Process utilizes information supplied by Participants when registering User Accounts, such as name, email address, street address, cell phone number, and other information ("Potential Identifiers"), which was stored in the *representante* table (i.e., Account Table).[36] Specifically, I understand that Dr. Freer developed a 10-step Aggregation Process, which attempts to confirm which User Accounts belong to the same Participant primarily through the application of a probabilistic record linkage approach known as the Fellegi-Sunter model with Expectation Maximization.[37]

38. Dr. Freer describes the Fellegi-Sunter model as a way to track agreement on multiple fields simultaneously and turn that information into an overall match probability for each pair of records under consideration. The output of the Fellegi-Sunter model is the agreement score for each pair of records under consideration, typically formulated as a "match probability."[38]

i. Step 1: Selection and Transformation of Fields

39. The first step in the Freer Aggregation Process was the selection and transformation of relevant fields from the list of Potential Identifiers. Specifically, he selects 11 of the 44 data fields in the Account Table ("Aggregation Fields"):[39]

- Rep_nome – Full Name
- Rep_email – Email
- Rep_cel – Mobile Phone
- Rep_fone – Telephone
- Rep_end – Address
- Rep_numero – Address Number
- Rep_cep – Address Postal Code
- Rep_login – Username

[36] Freer Report at 5, 23-37 and 44-45.
[37] Freer Report at 5, 13, and 62.
[38] Freer Report at 13-15.
[39] Freer Report at 5 and 40; and 10ksample.xlsx.

- Rep_pwd_secondary – User Created Secondary Password
- Rep_cpf – Social Security or Other Personal Identifier (Other Countries)
- Rep_datanasc – Birthday

40. Other Potential Identifiers available in the Account Table that do not appear to have been utilized in the Freer Aggregation Process include but are not limited to: first name (rep_primeiro_nome); last name (rep_sobrenome); gender (rep_Sexo); company (rep_comp);[40] address neighborhood (rep_barrio); city (rep_cidade); address state (rep_uf); and country code (rep_pais).

41. Additionally, with respect to the rep_login field, Dr. Freer makes a further transformation and truncates it to just the first four characters.[41]

ii. Step 2: Cleansing of Fields

42. The second step of the Freer Aggregation Process was to clean-up and format certain fields. Depending on the specific field, this typically included converting letters to lower case and eliminating certain spaces (e.g., leading spaces, trailing spaces, double spaces).[42]

43. Other cleansing steps were performed on specific fields. For example, the rep_nome field ("Name Field") was replaced with an empty string where it had a length of one, where it was comprised of entirely non-alphanumeric characters, or where it had a length of three and the first and third characters were standard Latin alphabet while the second character was a space (e.g., "a a" or "1 1").[43] For phone numbers (i.e., rep_cel and rep_fone), any non-digit (i.e., 0-9) characters were removed and only the last 7 digits were used.[44]

iii. Step 3: Deterministic Linkage

44. The third step in the Freer Aggregation Process was the application of an initial deterministic record linkage, whereby it combined any User Accounts that were in exact agreement

[40] Consistent with Dr. Freer's observation, this field actually appears to contain various components of the address. (Freer Report at 35.)

[41] Freer Report at 44-45.

[42] Freer Report at 45-46.

[43] Freer Report at 46.

[44] Freer Report at 46.

on cleansed non-null versions of eight the fields: rep_nome; rep_email; rep_cel; rep_fone; rep_end; rep_numero; rep_cep; and rep_login.

45. This deterministic step resulted in 4,367,728 account groupings.[45]

iv. Step 4: Blocking

46. Dr. Freer states that when the dataset is large, it may be computationally unreasonable to run the Fellegi-Sunter algorithm on the set of all pairs directly. For example, in the TelexFree dataset after the application of Step 3, there were approximately 4 million account groupings, resulting in approximately 10 trillion pairs of account groupings to consider. As such, Dr. Freer's fourth step utilized a technique known as "blocking" to reduce the number of pairs under consideration to a much smaller number. Dr. Freer notes that the restriction to a smaller subset of pairs is justified only if every pair that is excluded from the blocking set would have been very unlikely to match, had it been considered. Dr. Freer further states that, importantly, the presence of a pair of records in the blocking set does not signal that the pair is necessarily likely to be a match; rather, it simply means that the pair meets the minimum requirement to be further considered.[46]

47. In this case, for purposes of consideration for potential linkage, Dr. Freer required that User Accounts have "substantial agreement on name." And any account grouping pairs that did not meet this threshold, as determined by Dr. Freer, were deemed to not be part of the blocking set and were thus assigned a match probability of 0.[47]

48. Specifically, Dr. Freer concludes that there is "substantial agreement" with a name if it meets the following criteria:[48]

[45] Freer Report at 49.
[46] Freer Report at 17-18 and 50
[47] Freer Report at 25 and 51.
[48] Freer Report at 53-54.

WRatio(accountGrouping1.*clean_rep_nome*,
accountGrouping2.*clean_rep_nome*) >= 85.5
AND

(*ratio*(accountGrouping1.*clean_rep_nome*,
accountGrouping2.*clean_rep_nome*) >= 70

OR

ratio(*firstLast*(accountGrouping1.*clean_rep_nome*),
firstLast(accountGrouping2.*clean_rep_nome*)) >= 80)

49. As part of this step, Dr. Freer also manually aggregates 62,725 User Accounts with a rep_nome of "John Williams" that also shared other commonalities, such as email domain ("mailinator.com") and street name ("Warry Lane").[49]

50. After the exclusion of John Williams pairs, the output was a blocking set consisting of 117,563,883 pairs of account groupings.[50]

v. Step 5: Calculation of Gammas

51. The fifth step of the Freer Aggregation Process was to consider every pair of account groupings in the blocking set and assign each pair a comparison vector containing the gamma values for the cleaned fields. The gamma values, which depending on the field ranged from 0-1, 0-2, 0-3 or 0-5, were typically determined by applying cutoffs to the results of the string nearness measures (i.e., fuzzy matching) applied.[51] As an example, the criteria for the email address gamma is shown below:[52]

[49] Freer Report at 55.

[50] Freer Report at 55.

[51] Freer Report at 56 to 60. I note that both user birthday and secondary password were compared directly with no fuzzy matching resulting in a binary match (1) or non-match (0).

[52] Freer Report at 57.

gamma_clean_rep_email:

- 3 **if** *email_ratio* == 100
- 2 **if** (*email_ratio* >= 92)
- 1 **if** (*email_beforeAt_ratio* >= 90)
- 0 otherwise

where

- *beforeAt* returns the portion of the string before the first @ character
- *email_ratio* = *ratio*(accountGrouping1.*clean_rep_email*, accountGrouping2.*clean_rep_email*)
- *email_beforeAt_ratio* = *ratio*(*beforeAt*(accountGrouping1.*clean_rep_email*), *beforeAt*(accountGrouping2.*clean_rep_email*))

vi. Step 6: Calculation of Match Probabilities

52. The sixth step of the Freer Aggregation Process was to utilize the Splink Python Library to automatically convert the comparison vectors for each pair of account groupings to match probabilities. [53]

53. The output of this step was an assignment of the match probability to each of the 117,563,883 pairs of account groupings in the blocking set.

vii. Steps 7-8: Fixed Threshold and Dynamic Clustering

54. The seventh and eighth steps of the Freer Aggregation Process involved aggregating pairs of the account groups into clusters. This was done by first creating 149 fixed threshold values, with the fixed thresholds occurring at given percentiles from 0.006 to 100. For any given fixed threshold value, Dr. Freer created a cluster such that each account grouping was placed into one cluster of the clustering by taking the transitive closure of the set of all account grouping pairs whose match probability was at least the given threshold. Higher fixed thresholds that required more closely related pairs would yield a larger number of clusters with smaller average cluster size, while smaller fixed thresholds that did not require the same closeness would yield fewer clusters with a larger average cluster size.[54]

[53] Freer Report at 62.

[54] Freer Report at 62-63.

55. However, due to the heterogeneity of the TelexFree dataset (i.e., some Participants entered information very consistently across may accounts, while other participants entered information that differed more widely), Dr. Freer ultimately elected to forgo a fixed threshold approach in favor of a dynamic clustering process.[55]

56. Dr. Freer states that the dynamic clustering process seeks to select a different fixed threshold for each account grouping in such a way that (a) the choices are compatible across the entire dataset, and (b) the measures of "overclustering" and "underclustering" are balanced with each other.[56]

57. Dr. Freer describes "overclustering" as:[57]

> Namely, define the overclustering measure of a given account grouping (with respect to a proposed clustering) to be the fraction of account groupings within its cluster with which it has a match probability above 0.5. In other words, it is the number of account groupings in its cluster with which it has a match probability over 0.5, divided by the total number of account groupings in its cluster. The average overclustering measure of a cluster (e.g., the fixed-threshold cluster of an account grouping for some particular fixed threshold) is then defined to be the average of the overclustering measures of each account grouping in the cluster. Finally, the overclustering measure of an entire proposed clustering (e.g., a candidate clustering for the dynamic clustering process) is defined to be the average overclustering measure of each account grouping's cluster, averaged across all account groupings. In this way, if the overclustering measure of a given clustering is close to 1, this indicates that the clusters do not contain many pairs of account groupings having low match probability, while if the overclustering measure is close to 0, this indicates that the clusters contain many pairs of account groupings having low match probability.

58. Similarly, Dr. Freer describes "underclustering" as:[58]

> Namely, define the underclustering measure of a given account grouping (with respect to a given clustering) to be the fraction of all account groupings with which it has a match probability above 0.5 that are already within its cluster. In other words, it is the number of account groupings in its cluster with which it has a match probability over 0.5, divided by the total number of account groupings in the dataset with which it has a match probability over 0.5. The underclustering measure of an entire proposed clustering (e.g., a candidate clustering for the

[55] Freer Report at 64-65.

[56] Freer Report at 69.

[57] Freer Report at 67-68.

[58] Freer Report at 68.

> dynamic clustering process) is then defined to be this quantity averaged across all account groupings. In this way, if the underclustering measure of a given clustering is close to 1, this indicates that its set of clusters already contains most of the pairs of account groupings having high match probability, while if the underclustering measure is close to 0, this indicates that there are many pairs of account groupings having high match probability which failed to both be included in any cluster.

59. This balance between overclustering and underclustering was done in the Freer Aggregation Process by choosing, for each account grouping, the fixed-threshold clustering with the smallest threshold that is still consistent with the resulting cluster having average overclustering measure above a determined dynamic clustering parameter. In this case, Dr. Freer determined the dynamic clustering parameter to be 0.60.[59]

60. This step resulted in dynamic clustering of 4,367,728 account groupings.[60]

viii. Step 9: Convert Clusters of Account Groupings into Clusters of User Accounts

61. The ninth step in the Freer Aggregation Process was to determine the final clusters by ungrouping the account groupings created in the various steps and converting each back into User Accounts. This step resulted in 1,566,383 clusters of User Accounts ("Account Aggregations" or "Clusters"), with a mean size of 7.01 User Accounts per cluster.[61]

ix. Step 10: Tabulate Total Net Equity for Each Cluster

62. The tenth and final step in the Freer Aggregation Process was to sum the Net Equity amount associated with each User Account (as determined by Mr. Martin) across all the User Accounts in a given cluster. In total, Dr. Freer aggregated 10,987,617 TelexFree User Accounts to 1,566,383 Clusters. Of those 1,566,383 Clusters, 1,484,702 Clusters are alleged to be Net Losers and 81,681 are alleged to be Net Winners. The Net Winners are claimed to have a total net gain of $1,555,981,482, with an average net gain of $19,049 per Net Winner, while the Net Losers are claimed to have a total net loss of $1,782,171,730, with an average net loss of $1,200 per Net Loser.[62]

[59] Freer Report at 71.
[60] Freer Report at 72.
[61] Freer Report at 72.
[62] Freer Report at 72-73.

B. Martin Net Equity Calculation

63. In the determination of Net, I understand the Trustee, and thus Mr. Martin Equity ("Net Equity Calculation" or "Martin Net Equity Calculation"), considered the following four primary transaction types:[63]

- **Direct Payment:** Transactions in which a Participant makes a payment (via cash, wire transfer, check, payment processor, etc.) to TelexFree in satisfaction of an invoice.
- **Direct Receipt:** Transactions in which a Participant receives a payment from TelexFree or payment processor.
- **Triangular Payment:** Transactions in which a recruiting Participant satisfies the membership fee for a newly recruited Participant using the recruiting Participant's accumulated Credits.
- **Triangular Receipt:** Transactions in which a recruiting Participant is assumed to have received cash from a newly recruited Participant in exchange for using his or her accumulated Credits to satisfy an invoice for the newly recruited Participant.

64. In addition to these four primary transaction types noted in the now-excluded Martin Report, I understand the Net Equity Calculation also includes two other types of transactions: (i) chargebacks; and (ii) manual/purchased Credits. Chargebacks relate to bank or credit card refunds ("Chargebacks") while manual Credits, or purchased Credits as they are referred to in the Huron Data Tables, relate to instances where a participant paid TelexFree directly or through a bank or merchant to purchase system Credits ("Purchased Credits").[64]

65. More specifically, Purchased Credits represent Credits issued to User Accounts unrelated to the purchase of a membership plan and not a result of the placement of advertisements or other components of the compensation scheme. Although I understand that the Trustee alleges that significant amounts of Purchased Credits were issued without consideration, I understand there were other instances in which Purchased Credits were issued to User Accounts in exchange for

[63] Martin Report at 11-12; and Huron TelexFree Analysis of Damages Presentation at 33-35.

[64] Telexfree Net Equity Calculation and Aggregation.pdf; Huron TelexFree Analysis of Damages Presentation at 33-35.

cash payments to TelexFree.[65] As such, Purchased Credits were excluded from the Net Equity Calculation by Mr. Martin unless SIG indicated that a payment was made to a banking institution or via cashier's check, or the value was below $20,000.[66]

66. There were also exchanges of Credits between User Accounts unassociated with the issuance and satisfaction of TelexFree invoices ("Credit Transfers").[67] These Credit Transfers have also been excluded from the Net Equity Calculation by Mr. Martin, and thus the Trustee.

67. The results by Net Winner and Net Loser based upon Mr. Martin's Net Equity calculation in conjunction with Dr. Freer's Aggregation Process are summarized below.[68]

Participant Type	User Accounts	Participants	Total Net Equity
Net Loser	9,150,955	1,484,702	($1,782,171,730)
Net Winner	1,836,662	81,681	$1,555,981,482
TOTAL	10,987,617	1,566,383	($226,190,248)

68. For comparison purposes, the table below the results by Net Winner and Net Loser based upon Mr. Martin's Net Equity calculation in conjunction with Mr. Martin's Aggregation Process.[69]

Participant Type	User Accounts	Participants	Total Net Equity
Net Loser	9,196,914	2,042,027	($1,809,367,883)
Net Winner	1,790,703	95,052	$1,583,177,635
TOTAL	10,987,617	2,137,079	($226,190,248)

VI. SUMMARY OF OPINIONS

69. I have assessed the methodologies set forth in the Freer Report and Martin Report relative to: (i) the Freer Aggregation Process; and (ii) the Martin Net Equity Calculation for Participants,

[65] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 10-11; Huron TelexFree Analysis of Damages Presentation at 34; and Telexfree Net Equity Calculation and Aggregation.pdf at 2.

[66] Huron TelexFree Analysis of Damages Presentation at 34; and Telexfree Net Equity Calculation and Aggregation.pdf at 2. (I am not aware of any analysis performed by Mr. Martin relative why a $20,000 threshold was selected.)

[67] Motion for Entry of Order that Debtors Engaged in Ponzi Scheme at 10-11.

[68] Freer Report, Exhibit 5. Also, see **Exhibit 3**.

[69] Query of the Huron Data Tables. Also, see **Exhibit 3**.

and specifically Net Winners. While I address each of the methodologies separately below, both the Freer Aggregation Process and the Martin Net Equity Calculation contained flawed assumptions that render Dr. Freer's and Mr. Martin's, and ultimately the Trustee's, conclusion regarding the Net Equity amounts supposedly won or lost by a given TelexFree Participant to be unreliable.

70. Relative to the Freer Aggregation Process, key considerations include:

- The Freer Aggregation Process is predicated on Dr. Freer's "understanding" that "it was common practice for recruits to use their own name while using the contact information…of their recruiters, or for family members to share the same contact information, or for a business and individual to share the same contact information,"[70] and as such he required "that any two User Accounts under consideration for possible linkage have at least fairly similar name."[71] However, this understanding is entirely unsupported and counter to real-world case facts. The Court in this matter has already excluded Mr. Martin for making the same assumption without adequate support. Specifically, the Court stated "I conclude that Mr. Martin has not shown that the name field data is sufficient to support his opinion that user accounts should be aggregated primarily by using that data for the purposes of determining who the net winners are and the extent of their liability…[h]e has not shown that his assumptions about the name field data's accuracy are reasonable and not speculative."[72] Further, Dr. Freer does not appear to have specifically considered, let alone tested, the accuracy of the Name Field on which he relies so heavily despite the Court's ruling, and the fact that accuracy is one of the most important dimensions of data quality according to his own citations.

- With respect to the accuracy and consistency of the Potential Identifiers based on the facts of this case, I have also not seen any meaningful evidence that TelexFree had internal controls in place during the registration of a User Account that would have prevented a Participant from entering different variations of their name or a name other than their own, such as spouse's name, family member name, business name, colleague/employer name, or some fictitious, erroneous, or arbitrary combination of numbers, letters, or punctuation. Moreover, Participants sometimes worked in teams, entering User Account information on behalf others, including on shared computers subject to autocomplete. And, importantly, the question of how accurately (or inaccurately) and consistently (or inconsistently) Participants entered the name or any other identifying information across their User Accounts simply cannot be answered by the TelexFree Account data alone, and would almost certainly vary from Participant to Participant. As a result, utilizing an Aggregation Process that requires Participants to have reasonably accurately or consistently entered the same or similar names in

[70] Freer Report at 2.

[71] Freer Report at 25.

[72] Martin Decision at 22.

every instance across all their User Accounts (which could number in the thousands) is unreliable and counter to the facts of this case.

- The results of the Freer Aggregation Process are inconsistent with his stated methodology and requirements. For example, as one of his primary tenets, Dr. Freer states "[b]ecause of the special role that the Name Field plays in the TelexFree dataset…, it is important that no pairs of account groupings that substantially disagree on the name end up matched together."[73] However, there are numerous examples of User Accounts with completely different names that end up grouped under the same Participant according to Dr. Freer. (e.g., Alex Lopes and Jean Carreiro, and Flor M Florido and L & L Home Improvement). Similarly, Dr. Freer states "[o]nly when the names in the TelexFree dataset are near enough…and there are other markers of agreement (such as a matching email address, cell phone number, etc.), should these two User Accounts be treated as belonging to the same Participant and should be aggregated."[74] Once again, there are numerous examples of User Accounts that share no meaningful overlap in Aggregation Fields other than name that the Freer Aggregation Process grouped under the same Participant.

- Manual review of Account Aggregations and ePOC data suggests numerous instances of potential under- and over-aggregation. Even the singular example cluster included within the Freer Report appears to contain aggregation issues. This is highly problematic as a single missed link has the ability to (and often did) dramatically impact the Net Equity associated with Participants.

- As part of his scope of engagement, Dr. Freer was asked to "develop an aggregation methodology that creates one 'cluster' of User Accounts for each Participant." Instead, Dr. Freer performed a purely mathematical exercise that attempts to link together User Accounts that have similar characteristics, and specifically where there is substantial agreement on the values in the Name Field. However, such an approach fails to appropriately consider instances where the resulting Clusters may actually relate to more than one Participant, or where a Participant may relate to more than one Cluster. For example, in his deposition, Dr. Freer explicitly states, "[w]hen the name differs substantially, whether it's the same person or not that entered the accounts, that entered that information, when the name differs substantially, my understanding is those are to be considered as separate clusters." However, the Court in this case already specifically defined Participant as a "person" who purchased a membership plan or VoIP Package in TelexFree, not a "cluster" of User Accounts with similar values in the Name Field. As such, to the extent that the ultimate purpose of Dr. Freer's analysis is a quantification of the supposed net funds received by each TelexFree Participant that comprises the Net Winner Class, Dr. Freer's analysis fails to accomplish this goal absent additional expert analysis. This was also confirmed by Dr. Freer in his deposition where he acknowledged "I'm not aggregating accounts to persons period. I'm aggregating

[73] Freer Report at 51.

[74] Freer Report at 25.

accounts to other accounts in clusters. The assignment to persons is a different thing that might come later."

71. Relative to Mr. Martin's calculation of Net Equity, key considerations include:

- There is no discussion in the Martin Report regarding why Triangular Transactions are included within the determination of Net Equity, but Credit Transfers are excluded when both appear to be substantively similar from a real-world (i.e., cash-based) perspective.

- To the extent any uncertainty exists regarding if, or how much, cash was exchanged as part of Credit Transfers, then such uncertainty necessarily extends to Triangular Transfers, given that the SIG database does not provide evidence of cash payments for either transaction type. This internal inconsistency within the Net Equity Calculation between the treatment of Credit Transfers and Triangular Transactions results in an unreliable Net Equity value.

- A foundational assumption of the Net Equity Calculation is that the Credit values recorded within SIG for Triangular Transactions (and/or Credit Transfers) are "typically" reflective of the amount of cash that was exchanged between Participants. However, the facts in this case suggest that in practice Participants transacted at amounts that could vary greatly, and sometimes at significant discounts, from the amount of Credits shown in SIG due to factors such as specific business arrangements, competition between Participants, or collection issues from new recruits. I understand that certain Participants also worked as part of larger teams with profit-sharing arrangements. Cash payment amounts could be waived, deferred, or borrowed and never repaid. None of this real-world activity is reflected in the SIG data on which Mr. Martin and the Trustee rely, and as a result the Net Equity Calculation can differ greatly from the net amount actually invested and received by Participants.

VII. ANALYSIS AND FINDINGS

A. The Freer Aggregation Process is Unreliable

i. Dr. Freer's Foundational Assumptions Regarding the Special Role of the Name Field is Unsupported and Counter to Case Facts

72. For the Trustee to determine whether a given Participant was a Net Winner or Net Loser, it was first necessary for the Net Equity associated with the one or more underlying User Accounts to be combined into a single Account Aggregation for the Participant that was the financial beneficiary in a complete and accurate manner.

73. In his attempt to accomplish this task, Dr. Freer primarily relied upon a Probabilistic Record Linkage process, and applied it to the User Account information that was entered by

Participants, as reflected in the TelexFree *representante* table (i.e., Account Table).[75] And, according to Dr. Freer, of all the 44 fields in the Account Table, the Name Field (i.e., rep_nome) supposedly plays a "special role." Specifically, he states:[76]

> Because of the pyramid nature of TelexFree's scheme, Participants often recruited friends and relatives into the scheme. **My understanding** of the TelexFree scheme is that, in the course of User Account creation, it was common practice for recruits to use their own name while using the contact information (email address, phone number, street address, and other fields) of their recruiters, or for family members to share the same contact information, or for a business and an individual to share the same contact information. **(emphasis added)**

74. This "understanding," which appears as Introductory Paragraph 5 to the Freer Report, is of paramount importance. It is foundational to the entirety of the Freer Aggregation Process, as described in the Freer Report:[77]

> Only when the names in the TelexFree dataset are near enough (in the sense of string nearness; see Section II.A.3), say "Alice Adams" and "Alice Addams", and there are other markers of agreement (such as a matching email address, cell phone number, etc.), should these two User Accounts be treated as belonging to the same Participant and should be aggregated. Accordingly, **in the aggregation methodology, we require that any two User Accounts under consideration for possible linkage have at least fairly similar name**. (**emphasis added**)

75. Thus, if Dr. Freer's assumption regarding the Name Field is found to be inaccurate, it would render the entirety of his analysis unreliable. Despite the critical importance placed on the Name Field, Dr. Freer provides neither a source for this "understanding" nor any meaningful support elsewhere in his report. However, when asked in deposition about the basis for this "understanding," Dr. Freer stated:[78]

> Generally, the court documents, including primarily, I believe the court's memorandum from June 2021 and possibly from the Martin report or Dennis rebuttal or Martin reply.

---

[75] Freer Report at 22 and 37 ("The aggregation methodology developed by Borelian, under my direction, is mainly probabilistic in its approach because it is necessary to have a probabilistic step in the aggregation methodology to account for near-misses in data field matching, to assign partial credit for various fields based on these near-misses, and to account for all of this information when forming clusters.").

[76] Freer Report at 2.

[77] Freer Report at 24-25.

[78] Freer Deposition at 130-131.

76. As an initial observation, I do agree that Dr. Freer's assumptions regarding the Name Field are consistent with (and as equally flawed as) the assumption made previously by Mr. Martin, who very similarly stated:[79]

> **After a thorough review of the data, the Team selected the full name of the Participant as a starting point**. Because Participants could receive Direct Receipts from TelexFree, it is reasonable to assume that a Participant would include an accurate name when registering a User Account. Additional fields were selected based on the likelihood that their inclusion, combined with other fields, would accurately aggregate User Accounts by Participants. (**emphasis added**)

77. To the extent that Dr. Freer's understanding regarding the Name Field is wholly or partially based upon Mr. Martin, such reliance is unfounded. Whereas Mr. Martin at least tried to provide a rationale – albeit fatally flawed as evidenced by his exclusion by the Court – Dr. Freer makes no such attempt. In fact, Dr. Freer appears to concede this point stating "[w]hile many Participants created multiple User Accounts, some Participants at times entered information that was inconsistent across their various User Accounts."[80] In his subsequent example, Dr. Freer describes two User Accounts opened by the same Participant that may have entered different email addresses but provides no evidence or support addressing why potential inconsistency would not extend to the Name Field as well.

78. By way of comparison, while the value in Name Field appears to have played no meaningful role in the actual operation of a User Account, the email address was how TelexFree contacted Participants, and thus could potentially be more representative of the Participant. For example, the Martin Report on this topic states:[81]

> Email address was selected because TelexFree records, such as invoices, were sent to the email addresses associated with the User Accounts. Therefore, Participants were presumed to more likely provide accurate email addresses than most other data.

[79] Marin Report at 16.

[80] Freer Report at 1.

[81] Martin Report at 16-17.

79. Furthermore, with respect to the Court's memorandum from June 2021 (i.e., the Martin Decision), the Court carefully explained over the course of ten pages why such an assumption regarding the Name Field is at best unreliable absent clear evidentiary support, stating:[82]

> Having concluded that fields would need to be used in some combination, Mr. Martin asserted that he "considered name...to be the most important variable." Martin Tr. 67:23-68:1, 222:19-20. Yet, **he has offered only unsupported assumptions and untestable bases for why he selected the name field, despite its acknowledged flaws, above all other fields to be the primary focus in his aggregation process.** See also Martin Tr. 222:21-24 (agreeing that "if name didn't mean as much to...a particular user, that could skew the results of an aggregation").
>
> As noted, for an expert witness's opinion testimony to be admissible, it must be "based on sufficient facts or data." Fed. R. Evid. 702(b). Given actual and potential data quality and reliability issues discussed above, the defendants assert that, having selected the name field as his starting point and constant, Mr. Martin relied upon insufficient facts or data to support his expert opinion that user accounts should be aggregated primarily by using data from that field to determine the identity of individual net winners and the extent of their net winnings. See Defs.' Mot. 13-15. As also noted, an expert witness's opinion testimony must be based upon reliable principles and methods and the reliable application of such principles and methods. Fed. R.Evid. 702(c)-(d). Defendants suggest that Mr. Martin's process of selecting the name field as his starting point and constant does not meet either of these requirements for admissibility of his expert opinion. See Defs.' Mot. 15-16.
>
> Given the issues identified above, **I conclude that Mr. Martin has not shown that the name field data is sufficient to support his opinion that user accounts should be aggregated primarily by using that data for the purposes of determining who the net winners are and the extent of their liability. He has not shown that his assumptions about the name field data's accuracy are reasonable and not speculative**. He effectively conceded that his initial assumption about participant behavior lacked support. He failed to support his additional assumptions with any reference to the actual data. His own statements and concessions about the name field data's quality issues, along with his failures to show that his efforts to address those issues were well-reasoned, all undermine his position that such data is sufficient to support his opinion. Further, Mr. Martin has offered almost no concrete information about the principles and methods that he used and how he used them to select the name field as his starting point and constant. With no documentation or detailed explanation of his efforts, there is no practical ability to objectively review the reliability of Mr. Martin's analytical techniques. For admissibility, it is not enough for an expert simply to provide his conclusions, he must also show his work so that its reliability, including its

[82] Martin Decision at 13-22.

> adherence to accepted practices in the relevant field, can be meaningfully reviewed. While I need not for purposes of the current dispute conclude that Mr. Martin's selection and use of the name field as a starting point and constant were correct or incorrect, **I must and do conclude that Mr. Martin has not shown the requisite reliability of his selection and use of the name field, which formed the basis of his opinion. (emphasis added)**

80. Given that the Court already rejected Mr. Martin's explanation on this issue, it is unclear why Dr. Freer now believes the Court's decision somehow supports his "understanding" when in fact it does precisely the opposite.

81. As for the Initial Dennis Report, Dr. Freer appears to ignore its express purpose, which was described in the Initial Dennis Report as:[83]

> [O]ne of the factors likely driving what appears to be under-aggregation in certain instances, is the disconnect between the Aggregation Algorithm's requirement that the names be the same or similar across every Participant User Account and the reality that Participants could create User Accounts using a variety of names, such as business name, personal name, or even fictional names. This reliance on consistent naming fails to appropriately account for instances where other Potential Identifiers provide strong indicia that the User Accounts potentially belong to a single Participant.

82. As such, the Initial Dennis Report, as well as the various examples contained therein, are counter to, not in support of, the notion that distinct Participants consistently entered the same or substantially similar names.

83. With respect to the limited additional discussion by Dr. Freer on why he believes the Name Field to be the same or substantially similar across all of a Participant's User Accounts, he states:[84]

> In a typical user-account dataset, email addresses and phone numbers are unique to an individual. However, this is often not the case in the TelexFree dataset. As described in Paragraph 5, a major source of complexity with the TelexFree dataset is the prevalence of User Accounts that contain different names, but the same email address, phone number, address, etc.
>
> As a result, the Name Field (rep_nome) in the TelexFree dataset gains a special meaning that it might not have in another dataset. In particular, it is the *only true differentiator* of Participants – if User Accounts A and B match in all other fields (email address, cell phone number, etc.) but A's Name Field is quite different from

[83] Initial Dennis Report at 27.

[84] Freer Report at 24-25.

> B's Name Field (say, "Alice Adams" and "Bob Baker"), then these User Accounts should be treated as belonging to different Participants and should not be aggregated.

84. As purported support for this assertion, Dr. Freer's cites the Initial Dennis Report (already discussed above) and the affidavits for Frantz Balan and Lusette Balan.

85. Once again, these Affidavits refute rather than support Dr. Freer's understanding. The Affidavit of Lusette Balan confirmed that she had "no involvement in Telexfree" despite her name appearing in multiple User Accounts, and which was further discussed in the Affidavit of Frantz Balan: [85]

> There are numerous other accounts that were set up by me or team members working with me that were not in my name but that use my personal email address and are aggregated to someone else. **For example, the Trustee has aggregated several accounts to my wife, Lusette. This is incorrect because my wife Lusette had no personal involvement with TelexFree.**
>
> Similarly, there are accounts that I did not set up or in any way operate that are aggregated to me, even though they use a different email, street address, or username. Some of those accounts use email address that do not belong to me, such as vjmanigat@gmail.com.
>
> Based on my understanding of the Trustee's aggregation model, I believe the Trustee does not understand that many participants worked in teams, which had shared goals of member recruitment in order to earn money. The Trustee assumes everyone worked alone…
>
> There are a number of reasons why the information used to make the aggregations is unreliable.
>
> First, within my team, everyone's basic personal information like name, address, and phone number was known, so anyone could put an account in someone else's name.
>
> Second, shared computers allowed users to "autocomplete" fields, so even if someone used the computer in our office to setup TelexFree accounts, they could use the autocomplete to enter my information without me necessarily knowing about it or using the account…
>
> Third, there were financial incentives for people to set up accounts that appeared to be someone different, even if you were using the account. This included the

[85] Affidavit of Frantz Balan, July 30, 2020 ("F. Balan Affidavit") at 3-5; and Affidavit of Lusette Balan, July 29, 2020 at 2.

> basic rule that you could not "recruit yourself," meaning that your downlines need to be different people. **(emphasis added)**

86. These affidavits represent real-world examples and directly contradict, rather than support, Dr. Freer's fundamental assumption that the Name Field "is the *only true differentiator*."[86] I also note that Dr. Freer continued to aggregate User Accounts to Mr. Balan that have the email address "vjmanigat@gmail.com" despite Mr. Balan's confirmation that these accounts did not belong to him. Further, Mr. Balan offers an explanation for Dr. Freer's observation regarding the prevalence of User Account that "contain different names, but the same email address, phone number, address, etc."[87] (e.g., people working in teams and/or with shared computers).

87. Dr. Freer has not provided any alternative real-world explanation as to why Participants, when creating User Accounts, supposedly always entered the same or substantially similar names in the Name Field, but frequently elected to utilize inconsistent or inaccurate information when entering every other field - opting instead to use the contact information of friends, relatives, or even fictious, erroneous or arbitrary combinations of numbers, letters, or punctuation.

88. The Court already made it clear that an Aggregation Process founded simply on the unsupported "understanding" that Participants consistently entered the same or substantially similar names across all of their User Accounts is insufficient. Yet, Dr. Freer failed to provide any meaningful support demonstrating why such an understanding is grounded in fact, which renders his Aggregation Process, similar to Mr. Martin's, unreliable for determining Net Equity.

ii. Dr. Freer Failed to Appropriately Consider the Quality of the Data Utilized in His Aggregation Process

89. Regardless of the approach to record linkage selected (e.g., deterministic, probabilistic, hybrid), Dr. Freer states "it is crucial to assess the quality of the data in each field" and "select fields to be used for linkage based on data quality."[88]

90. Further to this point, Dr. Freer describes four dimensions of data quality:[89]

[86] Freer Report at 24-25.
[87] Freer Report at 24.
[88] Freer Report at 5-6.
[89] Freer Report at 7.

> The goal of data quality assessment is to identify fields that are of sufficiently high quality for reliably assessing pairwise matches. The dimensions relevant to data quality include **completeness** (how often the field values are present vs. missing), **consistency** (how consistent the values used for matching are), **accessibility** (inclusion of different kinds of nonoverlapping fields such as name, address, phone, email, etc.), and **believability** (how credible or plausible the actual values are). **(emphasis added)**

91. As support for this statement, Dr. Freer cites Section 3.1 of "Data Matching, Data-Centric Systems and Applications" by Peter Christen (the "Christen Text"). However, Section 3.1, entitled Data Quality Issues Relevant to Data Matching lists six dimensions to data quality, not four. Notably absent from Dr. Freer's consideration are:[90]

- *Timeliness*: How old are the data available. For the matching of two databases, have the data been recorded at the same time or not? This can be a crucial factor to a successful matching because personal information, such as people's addresses, telephone numbers, and even names, change over time. If the data to be matched have been recorded at different points in time then this needs to be taken into account during the matching process.

- *Accuracy*: How accurate are the attribute values in the database(s) used for matching or deduplication? Is it known how the data have been entered or recorded? Have data entry checks been performed, and have the data been verified for correctness using external reference data (such as references of known and valid addresses)?

92. It is unclear whether Dr. Freer's omission of these two data quality dimensions from his report was intentional, but both dimensions are relevant in this case.

93. With respect to *timeliness*, User Accounts were created over the period of approximately February 2012 through April 2014.[91] As such, during this more than two-year period, it is possible that identifying information for certain Participants could have changed, including name.

94. With respect to *accuracy*, the Christen Text states that it, along with consistency, are arguably the most important data quality dimensions for data matching and deduplication.[92] Further, the text suggests one should consider how the data was recorded, whether data entry checks were performed, and whether the data was verified using external references.

[90] Section 3.1 of P. Christen, Data Matching, Data-Centric Systems and Applications, Springer-Verlag, 2012.

[91] This time range reflects the minimum and maximum dates shown in the rep_data_cad field (i.e., User Account creation date) per the *representante* table. Also, see Freer Report at 31.

[92] Section 3.1 of P. Christen, Data Matching, Data-Centric Systems and Applications, Springer-Verlag, 2012.

95. In this case, with respect to how the data was recorded, it was generally inputted manually by millions of individuals worldwide through a website-based registration form. I have not seen any evidence regarding whether this webform was (i) consistently in English or Portuguese; (ii) to if and when the language may have changed over time; or (iii) if the Participant had the ability to manually select the language.[93] As such, there was the potential for language barriers as part of the input process for certain Participants. Also, as discussed above, Participants sometimes worked in teams, entering User Account information on behalf others, including on shared computers subject to autocomplete.[94]

96. With respect to data entry checks, I am not aware of any internal controls that took place during the registration of a User Account that would have validated or confirmed a name, or prevented a Participant from entering a fictious, erroneous or arbitrary combination of numbers, letters, or punctuation in place of a name.[95]

97. This does not support the notion that either the Participant's name or the remainder of the TelexFree Account table on which Dr. Freer relies had a high degree of accuracy, an important dimension of data quality according to Dr. Freer's own citations.

98. The Court also explicitly emphasized the need for an independent assessment of data accuracy, and particularly the Name Field, stating:[96]

> I conclude that Mr. Martin has not shown that the name field data is sufficient to support his opinion that user accounts should be aggregated primarily by using that data for the purposes of determining who the net winners are and the extent of their liability. **He has not shown that his assumptions about the name field data's accuracy are reasonable and not speculative**. **(emphasis added)**

99. I also note that in his deposition Dr. Freer explicitly states that he "didn't make as, formal, a separate assessment" of *accuracy* and instead "completeness, consistency, and believability have to do with accuracy and those I assessed in detail."[97] However, using these other three dimensions of data quality as a proxy for or indicia of *accuracy* is unfounded, as Dr. Freer subsequently

[93] See e.g., Martin Decision at 12.
[94] See e.g., F. Balan Affidavit at 3-4.
[95] See e.g., F. Balan Affidavit at 4-5.
[96] Martin Decision at 22.
[97] Freer Deposition at 106.

admitted stating "it's certainly possible to have complete data that is not accurate, or consistent data this is not accurate, or believable data that is not accurate."[98]

100. Further, with respect to the *believability* dimension of data quality, Dr. Freer states that the "dataset is also generally rather believable, despite a small portion of exceptions (due to garbage like '11111' or value like 'Mickey Mouse')."[99] He then points to a list of 471 names associated with just a single Cluster, which he states reflects "a visual illustration of the general quality of the dataset."[100] Dr. Freer provides no discussion regarding if or how this single Cluster, presumably selected at his own discretion, is representative of the entire dataset.

101. In fact, there are approximately 2,400 unique names associated with more than 200,000 separate User Accounts where the name entered by the user is less than 4 characters.[101] Of these, the vast majority (i.e., over 2,300 names) are Net Losers, the losses of which total more than $11.4 million.[102] Examples include the following:

[98] Freer Deposition at 128.

[99] Freer Report at 24.

[100] Freer Report at 24 and Exhibit 3.

[101] Queries of the Freer and Huron Data Tables.

[102] Queries of the Freer and Huron Data Tables.

Full_Name	# of User Accounts	Aggregate Net Loss
1 1	40,687	($2,063,193)
. .	36,498	($1,986,986)
a a	30,363	($1,551,269)
.	8,329	($535,314)
q q	5,339	($278,741)
W W	3,191	($161,427)
c c	3,009	($151,197)
m m	2,077	($108,882)
x x	2,166	($108,732)
d d	2,021	($102,095)
l l	1,966	($98,403)
Jo	1,849	($92,354)
A B	1,770	($89,421)
0 0	1,667	($84,082)
k k	1,512	($76,497)
s s	1,455	($72,954)
j	1,328	($68,650)
j j	1,298	($65,619)
p p	1,242	($62,076)
f f	1,171	($59,481)
, ,	1002	($59,243)
1	830	($44,638)

102. Additionally, based on my review of the Account Table, I also observed usernames with more than four more characters that are as equally inaccurate. For example, there are over 125,000 TelexFree User Accounts that consist entirely of numerical values, punctuation, or spaces (i.e., no letters), such as "....." or "398985 9849348".[103]

103. Dr. Freer performs a similar analysis stating:[104]

> [N]early all of the name values contain strings consisting entirely of alphabetical characters or the symbols commonly found in names (period, comma, hyphen, space, or apostrophe) – as opposed to numbers or other symbols – which is an indication that the value is a legitimate name. Specifically, 97% of the name field values are both five characters or longer and contain only the characters described above.

[103] Based on query of the Account table. Also, see e.g., rep_id 11655333 and 26451460.

[104] Freer Report at 26.

104. Additionally, Dr. Freer performed a manual review of just 300 Name Field values from this remaining population, or 0.0028% of remaining User Accounts with Name Field values that are both five characters or longer and contain only the characters commonly found in names.[105] Based on this review, he concludes that 98% "look[ed] like a name."[106] While the Freer Report provides no explanation as to what precisely it means to "look like a name," Dr. Freer in his deposition elaborated on this point confirming that "Mickey Mouse" would not have looked like a name to him but "Barack Obama" would have,[107] despite the fact that it's likely both are inaccurate.

105. Taken together, Dr. Freer estimates that in 95% of the User Accounts, the value of the rep_nome field appears to be a name.[108] Assuming the denominator in his percentage is the 10,987,617 User Accounts attributable to TelexFree and associated with at least one paid invoice, this means, by Dr. Freer's own admission, there are over 500,000 User Accounts that do not appear to contain a name in the Name Field.

106. Considering just the approximately 170,000 User Accounts that had at least one number in the Name Field and didn't appear to reflect a business,[109] this accounts for approximately $2.3 million in winnings from User Accounts with positive net equity and $11.0 million in net losses from User Accounts with negative net equity.[110]

107. As such, this does not support the notion that either the Name Field or the remainder of the TelexFree Account Table is "generally rather believable." This is particularly important when considered in the appropriate context, which is the potential impact on a single Participant, not the entire population of Participants.

[105] Calculated as 300 / (97% x 10,987,617). 10,987,617 is the number of User Accounts attributable to TelexFree and associate with at least one paid invoice. Also, see Freer Report at 26.

[106] Freer Report at 26.

[107] Freer Deposition at 111-112.

[108] Freer Report at 25-26.

[109] Business were assumed to be any names that included the following: inc, corp, llc, trust, ltd, consult, service, enterprise or .com

[110] Queries of the Freer and Huron Data Tables.

108. In terms of *consistency*, Dr. Freer contends that this dimension is "generally high" for eight of the Aggregation Fields (including name), albeit with "occasional variation."[111] Yet, he provides little in the way of support for this contention. This is particularly true as it relates to Dr. Freer's most critical Aggregation Field, the Name Field associated with the User Account. On this issue, Dr. Freer simply states:[112]

> It is worth noting that while it may turn out that a name is made up (such as "Mickey Mouse"), this itself is not a critical issue for data aggregation. The relevant issue for aggregation is whether the values in the field support the matches, not whether or not the values in the field reflect the true name. Even when a Participant uses a pseudonym that is totally unrelated to their real name, **it is possible to successfully aggregate their User Accounts if (a) they use the pseudonym consistently,** and (b) not too many other Participants use that same pseudonym without including different information in other fields. In the best case scenario, the true identity of the Participant can, at a later time, be ascertained from the other indicia; in the worst case scenario, the Participant will never be identified, but their records do not contaminate the rest of the aggregation. **(emphasis added)**

109. This statement appears to imply is that *accuracy* and *believability* of the Name Field are of lesser importance to the Freer Aggregation Process, as long as there is *consistency* in the name used by a given Participant across all of their User Accounts, and the remaining Aggregation Fields provide sufficient differentiation from other Participants. However, Dr. Freer poses this as an unsupported conditional "if then" statement, not a statement of fact confirmed through documentation or analysis. Dr. Freer even admits that the data lacked consistency, stating: "The TelexFree dataset presents an additional challenge due to the presence of heterogeneity: some Participants entered their information very consistently across many accounts, while other Participants, for a variety of reasons, entered information that differed more widely (e.g., name variants, typos, multiple email addresses or phone numbers, etc.)."[113]

110. Stated in the converse, if Participants did not enter names consistently or did not provide sufficient differentiation in other Aggregation Fields, which the case facts suggest was likely for certain Participants (especially large Net Winners), then it would not be possible for Dr. Freer to successfully aggregate their User Accounts. And, importantly, the question of how consistently

[111] Freer Report at 24.
[112] Freer Report at 25.
[113] Freer Report at 64.

(or inconsistently) and accurately (or inaccurately) Participants entered the name or any other identifying information across their User Accounts simply cannot be answered by the TelexFree Account data alone, and would almost certainly vary from Participant to Participant. To this point, the Freer Deposition reads:[114]

> Q. Did you make any determinations about the likelihood of data entry scenarios where information like name or any portion of contact information could be flipped between different users?
>
> A. No, and at some level, it's impossible to tell certain things from the dataset itself. If you take a record and you write down all the fields and you tell me, "Person A did it versus Person B did it," of course, information is not in the dataset. I don't know from the dataset, period.
>
> Q. You agree with me that you have at least some understanding that on certain occasions when a user entered data, they entered information that pertained to somebody else?
>
> A. That is, yeah, I mean, as I describe in that paragraph, there's, in fact, several specific scenarios where my understanding was that kind of thing was likely to have happened.

111. To this point, Section 2.6 of the Christen Text is entitled Evaluation of Matching Quality and Complexing, and states:[115]

> To evaluate the completeness and accuracy of a data matching project, some form of ground-truth data, also known as *gold standard*, are required. Such ground-truth data must contain the true match status of all known matches (the true non-matches can be inferred from them). However, obtaining such ground-truth data is difficult in many application areas. For example, when matching a large tax payers database with a social security database it is usually not known which record pair classified as a match refers to a real, existing individual who has a record in both databases. Only further investigation, such as checking extra data about the individual under consideration, or even contacting them, can help determine the truth about such a classified match.
>
> A related problematic issue is the manual classification of potential matches through clerical review. It is often difficult to make a manual match or non-match decision with high confidence if the two records in a potential match pair contain several attribute values that differ from each other. Without further external information, a decision that was made manually might be wrong.

112. As discussed above, and even setting aside the potential issues with the *believability* and *accuracy* of the data, simply assuming without basis or support that a Participant *consistently* entered data, and critically the Name Field, into a webform sometimes thousands of times over is not a reliable basis for an Aggregation Process in this case. This is especially true when (i)

[114] Freer Deposition at 132-133.

[115] Section 2.6 of P. Christen, Data Matching, Data-Centric Systems and Applications, Springer-Verlag, 2012.

TelexFree did not appear to have any substantive verification or internal controls process as part registering a new User Account, and (ii) Participants had little, if any, incentive to enter names (or any Potential Identifiers) accurately. Thus, without quality data as an input, as is the case here, Dr. Freer's output is equally unreliable. The Christen Text concisely explains this issue:[116]

> For any type of data analysis, processing and management, the *garbage-in garbage-out* principle holds. If the quality of the input data is low, then the output generated is normally not a high quality or accuracy either.

iii. Dr. Freer's Results are Internally Inconsistent

113. As one of his primary tenets, Dr. Freer states "[b]ecause of the special role that the name field plays in the TelexFree dataset (see Section III.B.1 above), it is important that no pairs of account groupings that substantially disagree on the name end up matched together."[117]

114. Specifically, Section III.B.1 of the Freer Report states:[118]

> [I]f User Accounts A and B match in all other fields (email address, cell phone number, etc.) but A's Name Field is quite different from B's Name Field (say, "Alice Adams" and "Bob Baker"), then these User Accounts should be treated as belonging to different Participants and should not be aggregated.
>
> Only when the names in the TelexFree dataset are near enough (in the sense of string nearness; see Section II.A.3), say "Alice Adams" and "Alice Addams", and there are other markers of agreement (such as a matching email address, cell phone number, etc.), should these two User Accounts be treated as belonging to the same Participant and should be aggregated. Accordingly, in the aggregation methodology, we require that any two User Accounts under consideration for possible linkage have at least fairly similar name.

115. However, as discussed below, the results of the Freer Aggregation Process are internally inconsistent. As a first example, take his simple statement that "if User Accounts A and B match in all other fields (email address, cell phone number, etc.) but A's Name Field is quite different from B's Name Field…then these User Accounts should be treated as belonging to different Participants and should not be aggregated."[119] In fact, the Freer Aggregation Process does precisely

[116] Section 3.1 of P. Christen, Data Matching, Data-Centric Systems and Applications, Springer-Verlag, 2012.

[117] Freer Report at 51.

[118] Free Report at 24-25.

[119] Freer Report at 24-25.

the opposite, as reflected in the User Accounts aggregated as Cluster 23118541 shown in the table below:[120]

cluster_id	rep_nome	rep_email	rep_numero	rep_end	rep_cidade	rep_pais	rep_fone	Sum of net_equity	Count of rep_id
23118541	A	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (1,047.90)	21
	B	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (12,125.70)	243
	C	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (13,572.80)	272
	D	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (7,634.70)	153
	F	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (5,189.60)	104
	H	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (948.10)	19
	J	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (54,590.60)	1,094
	K	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (1,097.80)	22
	L	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (12,075.80)	242
	M	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (11,676.60)	234
	N	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (10,429.10)	209
	P	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (6,586.80)	132
	R	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (3,293.40)	66
	S	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (7,684.60)	154
	T	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (8,932.10)	179
	V	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (6,736.50)	135
	Z	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (748.50)	15
	(blank)	(blank)	1485	Rua Jo	Xucarana	AF	689958478	$ (12,075.80)	242
Grand Total								**$ (176,446.40)**	**3,536**

116. As part of his Cluster 23118541, Dr. Freer has, in fact, aggregated User Accounts with name "A" and User Accounts with name "B" (along with other letters of the alphabet and even no name at all). This likely occurred because Dr. Freer removed the Name Field from consideration where it only contained a single character.[121] And, while Dr. Freer may have intended his "A" and "B" example in a purely hypothetical rather than literal sense, it still demonstrates an important point. In this example, Dr. Freer aggregates 3,536 User Accounts to a single Participant even though that supposed Participant inconsistently entered his or her name. This outcome is in direct contradiction to his own example and critically his assumption that name is the "only true identifier."[122]

117. This example is also highly relevant when compared to other instances where a Participant may likely have entered a variety of "garbage" names, albeit with more than one character. In the

[120] See **Exhibit 11**.

[121] Freer Report at 46 ("For rep_nome (name), the additional cleaning steps consist of: (a) replacing, by the empty string, any values of length 1.").

[122] Freer Report at 24.

table below, these supposedly different Participants all share substantial commonalities, in whole or in part, other than a clearly fictitious name:[123]

cluster_id	rep_nome	rep_email	rep_end	rep_cep	rep_fone	Sum of net_equity	Count of rep_id
10758965	Team Legendary	billybob.arbonne@gmail.com	1	1	?	$ (1,425.00)	1
			?	?	3477667668	$ (1,425.00)	1
					?	$ (2,101.50)	2
			123 main st	77007	3477667668	$ (19,900.10)	14
			123 main st.	77007	3477667668	$ (21,125.50)	15
		tommylin19@yahoo.com	?	?	?	$ (14,832.40)	13
			123 main st	77007	3477667668	$ (1,425.00)	1
	Team Legendary 10heimer	billybob.arbonne@gmail.com	123 main st	77007	3477667668	$ (1,425.00)	1
	Team Legendary 11heimer	billybob.arbonne@gmail.com	123 main st	77007	3477667668	$ (1,425.00)	1
	Team Legendary 12heimer	billybob.arbonne@gmail.com	123 main st	77007	3477667668	$ (1,425.00)	1
	Team Legendary 13heimer	billybob.arbonne@gmail.com	123 main st	77007	3477667668	$ (2,850.00)	2
	Team Legendary 15heimer	billybob.arbonne@gmail.com	123 main st	77007	3477667668	$ (1,425.00)	1
	Team Legendary 16heimer	billybob.arbonne@gmail.com	123 main st	77007	3477667668	$ (1,425.00)	1
	Team Legendary 17heimer	billybob.arbonne@gmail.com	123 main st	77007	3477667668	$ (1,425.00)	1
	Team Legendary 18heimer	billybob.arbonne@gmail.com	123 main st	77007	3477667668	$ (1,425.00)	1
10778683	Team Legendary	billybob.arbonne@gmail.com	?	?	?	$ (1,025.80)	1
10787619	Team Legendary	peterkwongusa@yahoo.com	1	1	1	$ (1,425.00)	1
			?	?	?	$ (4,203.00)	4
	Team Legenday	billybob.arbonne@gmail.com	1	1	?	$ 7,174.90	1
11089827	chin hao chang	billybob.arbonne@gmail.com	9080 telstar ave #302	91731	3477667668	$ 18,824.40	1
12982273	X Legend X	billybob.arbonne@gmail.com	123 main st	77017	3477667668	$ (1,425.00)	1
13328313	27legendary WC	billybob.arbonne@gmail.com	1337 legend ave	77007	3477667668	$ (1,425.00)	1
14207003	Majestic Tia	billybob.arbonne@gmail.com	123 main st.	77077	3477667668	$ (1,425.00)	1
14394871	Walt Disney	billybob.arbonne@gmail.com	123 main st	551166	11111111111	$ (14,250.00)	10
			123 main st.	77007	3477667668	$ 1,474.90	1
15341909	Agent hydralisk	billybob.arbonne@gmail.com	123 main st	77007	3477667668	$ (11,350.10)	8
	Agent Hydralisk2	billybob.arbonne@gmail.com	123 main st.	77007	3477667668	$ (1,425.00)	1
	Agent Hydralisk3	billybob.arbonne@gmail.com	123 main st.	77007	3477667668	$ (1,425.00)	1
15744519	WI-FI Cindy	billybob.arbonne@gmail.com	123 main st.	77007	3477667668	$ (1,425.00)	1
22032025	King Blizzard	billybob.arbonne@gmail.com	123 main st	0	3477667668	$ (339.60)	1
23079231	Barack Obama	billybob.arbonne@gmail.com	123 main st.	0	1111111111	$ (339.60)	1
23088735	why why not	billybob.arbonne@gmail.com	123 main st	0	1111111111	$ (339.60)	1
23092685	glee 810	billybob.arbonne@gmail.com	123 main st	0	1111111111	$ (339.60)	1
23104679	dark shadow	billybob.arbonne@gmail.com	123 main st	0	1111111111	$ (339.60)	1
Grand Total						**$ (90,087.20)**	**94**

118. Unlike the previous example where Dr. Freer was willing to aggregate different names, in this case, Dr. Freer treats these as 15 separate Participants. Such a distinction is arbitrary, however, and based on the number of characters in a string, not case facts or sound methodology. If a single Participant is willing and able to create multiple User Accounts with "A", "B," or "C" as the name, why is it any less likely that a single Participant created multiple User Accounts with "Barack Obama," "Walt Disney," or "Team Legendary" as the name?

119. And, extending this logic one step further, if a Participant is willing and able to use a range of fictitious names only limited by their creativity, why is it any less likely that a Participant (or someone on the Participant's behalf) either intentionally or unintentionally entered other

[123] See **Exhibit 12**. I also note that these accounts share substantial similarity on Secondary Password.

inaccurate or inconsistent names across their various User Accounts, such as the name of friends, relatives, or businesses? This internal inconsistency provides further indicia of the flawed nature of the Freer Aggregation Process.

iv. The Results of the Freer Aggregation Process are Inconsistent with his Stated Methodology and Requirements

120. Dr. Freer explicitly states that "in the aggregation methodology, [he] require[d] that any two User Accounts under consideration for possible linkage have at least fairly similar name"[124] and that the Name Field was the "only true differentiator."[125]

121. Dr. Freer further describes this requirement as follows:[126]

> Because of the special role that the name field plays in the TelexFree dataset..., it is important that no pairs of account groupings that substantially disagree on the name end up matched together. I use blocking as a mechanism to accomplish this goal, by ensuring that the blocking set contains only pairs of account groupings with substantial agreement on name. In this way, only such pairs (i.e. those with substantial agreement on name) are even considered for potential linkage.

122. However, if this was indeed the intent of his Aggregation Process, then Dr. Freer's results are inaccurate and inconsistent with his desired outcome.

123. For those names that Dr. Freer aggregated under a single Participant, see for example Cluster 119427, which includes the following unique names:[127]

[124] Freer Report at 25.

[125] Freer Report at 24.

[126] Freer Report at 51.

[127] See **Exhibit 13**.

cluster_id	rep_nome	rep_email	rep_end	rep_cidade	rep_pai	rep_fone	Sum of net_equity	Count of rep_id
119427	Alex de carvalho Lopes	carvalopes@gmail.com	Rua Prof° Renato Ribeir	Vitória	UY	552730262158	$ (1,375.10)	1
	Alex Lopes	carvalopes@gmail.com	R Prof Renato Ribeiro d	Vitoria	UY	5527997852189	$ (1,474.90)	2
	Edjane Carreiro	nay_justin@hotmail.com	SW 10Ch Place	derffildbeach	US	15613057841	$ (49.90)	1
		may_justin@hotmail.com	Sw 10Th place	Deerfield Beach	US	5613057841	$ (289.10)	1
	Eliaavne Carreiro	elianevit40@gmail.com	Maplewood Av	Everett	US	17818445344	$ (49.90)	1
	Eliane B Carreiro	elianevit40@gmail.com	elianevit31@hotmail.co	Boston	US	17818445344	$ (975.00)	1
			Maplewood av	Boston	US	17818445344	$ 515.40	1
				Everett	US	16175071974	$ (2,990.00)	2
			maplewood ave	Boston	US	17818445344	$ 710.00	1
			Rua Sao Jorge	Itaciba	US	2732868987	$ (906.50)	1
		elianevit31@hotmail.com	Maplewood av	Boston	US	17818445344	$ (149.70)	1
	Eliane B. Carreiro	elianevit40@gmail.com	43 maplewood ave	Boston	US	17818445344	$ 16,875.74	1
						7818445344	$ (199.60)	1
			43 maplewood ave,	Everett	US	17818445344	$ 210,371.50	1
			Maplewood av	Boston	US	17818445344	$ (1,351.00)	2
			maplewood ave	Boston	US	17818445344	$ (1,975.00)	4
	Eliane Carreiro	elianevit40@gmail.com	maplewood av	Everett	US	17818445344	$ (49.90)	1
			43 maplewood	Boston	US	17818445344	$ 11,257.16	1
			43 maplewood av	Everett	US	7818445344	$ (49.90)	1
						16179083340	$ (49.90)	1
			Maplewood Av	Boston	US	16175071979	$ (49.90)	1
			Maplewood av	2149	US	17818445344	$ (249.50)	1
				Everett	US	17818445344	$ (6,745.90)	8
						2732868987	$ (149.70)	2
						16179083340	$ (249.50)	1
						7817384710	$ (1,445.10)	1
						16175071974	$ (28,045.90)	53
						18578889783	$ (99.80)	1
			maplewood ave	Boston	US	17818445344	$ (49.90)	1
			Rua Sao Jorge	cariacica	BS	(27) 3286-8987	$ (49.90)	1
					US	(27) 3286-8987	$ (49.90)	1
				Itaciba	US	(27) 3286-8987	$ 38,350.90	2
		elianevit31@hotmail.com	maplewood av	Everett	US	17818445344	$ (199.60)	1
			43 maplewood av	Everett	US	2792290673	$ (49.90)	1
			43 maplewood ave	Everett	US	16178422930	$ (49.90)	1
						18573184245	$ (49.90)	1
			Maplewood av	2149	US	17818445344	$ (149.70)	2
				Everett	US	17818445344	$ (848.30)	11
		elianevit40gmail.com	Maplewood av	Everett	US	7818445344	$ (49.90)	1
		Cleidesantos2273@hotma	Maplewood av	Everett	US	16175071974	$ (99.80)	1
		may_justin@hotmail.com	4379 sw 10th place	Deerfield Beach	US	5613057841	$ (49.90)	1
		felishapaz@gmail.com	Maplewood av	Everett	US	16175071974	$ (149.70)	1
	jaqueline lopes	josehelber@gmail.com	.....	.....	US	18578882783	$ (49.90)	1
	Jaqueline Lopes Carreir	josehilber@gmail.com	121 main st	Everett	US	8578882783	$ (5,700.00)	4
	Jean Carreiro	jelite2014@gmail.com	maplewood ave	Everett	US	8572461459	$ (49.90)	1
			Maplewood ave.	Everett	US	8572461459	$ (49.90)	1
			Mapplewood Ave	Everett	US	8572461459	$ (49.90)	1
	Jean de Paula Carreiro	carreirojean@hotmail.com	43 maplewood ave	Boston	US	18572461459	$ (49.90)	1
	Jean P Carreiro	elianevit40@gmail.com	maplewood ave	Boston	US	18572461459	$ (3,925.00)	7
			Maplewood Ave	Boston	US	18572461459	$ (1,375.00)	1
	Jean P. Carreiro	elianevit40@gmail.com	43 maplewood ave	Boston	US	18572461459	$ (1,375.00)	1
			Maplewood av	Boston	US	18572461459	$ 112,639.20	1
	nael lopes	saminha_nunes@yahoo.c	.	Everett	US	.	$ (49.90)	1
Grand Total							**$ 327,228.50**	**140**

124. Manual review of the Name Field for this Cluster, which Dr. Freer contends is the "*only true differentiator*,"[128] suggests that he is potentially over aggregating at least five separate Participants: (i) Alex Lopes; (ii) Eliane Carreiro; (iii) Jacqueline Carreiro; (iv) Jean Carreiro; and (v) Nael Lopes. Even if comparing selected pairs of these names (or variations thereof) yields a high "similarity score" using Dr. Freer's selected fuzzy matching approach, common sense suggests that these are likely not the same person. While each of the three Carreiros and two

[128] Freer Report at 24.

Lopeses may or may not be family members given they share the same last name, the first name clearly differs. And, taking the far ends of the Cluster, Alex Lopes and Eliane B. Carreiro share almost nothing in common with respect to name, let alone substantial agreement. In fact, review of the underlying data suggests Alex Lopes and Eliane Carreiro have no contact information in common whatsoever, with the former living Vitoria, Brazil and the latter in Everett, Massachusetts. This demonstrates a clear failure in the Freer Aggregation Process, as it is it stands in direct contradiction to his stated goal "that no pairs of account groupings that substantially disagree on the name end up matched together."[129]

125. This issue with the Freer Aggregation Process likely lies, at least in part, with his clustering methodology. As a hypothetical, Group 1 may share some commonalities with Group 2, and Group 2 may share some commonalties with Group 3, and Group 3 may share some commonalities with Group 4, the result of which is the aggregation of Groups 1, 2, 3 and 4 together as a single Participant. The potential issue is that taken as a whole, Groups 1 and 4 may share nothing in common, including name.

126. On this point, Section 6.8 of the Christen Text addresses issues managing transitive closure of clusters stating:[130]

> The transitivity of matches can also lead to problems in that 'chains' of records, where individual pairs are classified as matches, are formed. The records at the two ends of the chain can, however, be quite different from each other, and they would not be considered to correspond to a match.

127. Dr. Freer in his deposition further discusses the topic as follows:[131]

> Whereas, in fact, there's heterogeneity in this dataset, where some parts of the dataset do look something like that, I wouldn't say "ideal situation" but easier-to-deal-with situation.
>
> In other parts, there is a greater risk of chaining, as you were describing it. And, of course, if A is joined to B is joined to C is joined to D and so on, by reasonably high numbers, but A is not joined to Z at the end by reasonably high number, it would be bad to include A and Z at the same time.

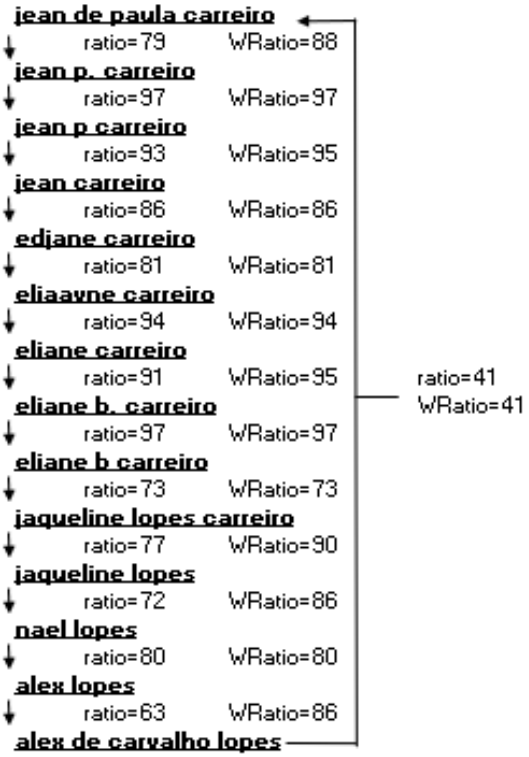
128. These "bad" scenarios are clearly seen in the example discussed above relative to the User Accounts with last name Carreiro and Lopes, where any two specific names may have a relatively high nearness measure (e.g., Eliane Carreiro vs. Edjane Carreiro), but taken at the outer edges the

[129] Freer Report at 51.

[130] Section 6.8 of P. Christen, Data Matching, Data-Centric Systems and Applications, Springer-Verlag, 2012.

[131] Freer Deposition at 202.

names can have a very low nearness measure (e.g., Jean de Paula Carreiro vs. Alex de Carvalho Lopes), as reflected in the chart below:



129. Other examples of names which appear in a single of Dr. Freer's Clusters, but do not appear to substantially agree, are shown below:[132]

cluster_id	Name 1	Name 2
140155	Irene Ramos Fernandes Silva	Nidya Leal
140155	Irene Ramos Fernandes Silva	Ruben Sousa
335971	Flor M Florido	L & L Home Improvement
100223	Heloisa Santamaria	Marialva Sousa de Matos
2522900	Benjamin Argueta	Kenia Venjamin
3194458	Dorali Escobar	Zuly Caicedo
4047480	fatima elena bettencourt de andrade	maria beatriz de andrade martins
4253920	George Saad	El Hachem Georges
3931290	Giovanni Rigatti	go Brignoni
2825686	Luis Miguel Joaquim Rodrigues	Sao Rodrigues
11199437	Regina Celia Cintra Castro	Karina c
9021130	Juan Pablo	Erick Gutierrez
10729205	Maria Do Scorro Prazeres Rodrigues	Ricardo Trancoso
5326578	Patrícia Moutinho	Maria Angela Magalhães Coutinho
12071881	Jose Demoura	Paulo Roberto Ferreira Moura
471741	Kelly Silva	Arcelina Benedita Mota

[132] Review of the Freer Data Tables.

cluster_id	Name 1	Name 2
471741	arcelina	zeike s
5014900	Camel R Quispe Calderon	ayme sias calderon
11096037	Helio Paz	eliton gonçalves paz
10756933	WenXiong Lee	Sui Ning Lee
10630997	fadelly robayo	Ricardo Ranulfo Yaranga Valencia
12266593	w dalman general services	BBC Financial Services LLLC

130. With respect to issues managing transitive closure Section 6.8 of the Christen Text states:[133]

> In real-world databases, the problem of record chains being generated by a pair-wise classification technique seems to occur only rarely because the space of all possible values in the different record attributes is very large. The likelihood that records that are not matching have a high similarity with each other is therefore very small.

131. However, in this case, the Aggregation Fields on which Dr. Freer relies differ from those found in typical "real-world" datasets, such as healthcare data, making transitive closure more prone to potential record chains. Dr. Freer admits as such stating:[134]

> In a typical user-account dataset, email addresses and phone numbers are unique to an individual. However, this is often not the case in the TelexFree dataset. As described in Paragraph 5, a major source of complexity with the TelexFree dataset is the prevalence of User Accounts that contain different names, but the same email address, phone number, address, etc.

132. Dr. Freer contends that another requirement of his Aggregation Process is the following:[135]

> Only when the names in the TelexFree dataset are near enough…**and there are other markers of agreement (such as a matching email address, cell phone number, etc.),** should these two User Accounts be treated as belonging to the same Participant and should be aggregated. **(emphasis added)**

133. However, closer inspection of his resulting clusters shows that Dr. Freer has in fact aggregated User Accounts that may share a similar name and rough geographic location but do not meaningfully share other markers of agreement, such as matching email addresses, cell phone numbers, birthdays, etc. For example, Cluster 5463980 includes 12 User Account and two distinct names, Loredana Zaglia and Loredana Ragolia and no other Aggregation Fields in common.[136]

[133] Section 6.8 of P. Christen, Data Matching, Data-Centric Systems and Applications, Springer-Verlag, 2012.

[134] Freer Report at 24.

[135] Freer Report at 25.

[136] See **Exhibit 14**.

rep_nome	rep_cpf	rep_ numero	rep_end	rep_cep	rep_cel	rep_datanasc	rep_email	Sum of net_equity	Count of rep_id
loredana zaglia	180859	26	via costantino nigra	10080	393332449962	NULL	raumy@me.com	(499.00)	5
							lunadicristallo59@yahoo.it	(249.50)	5
						8/18/1959	raumy@me.com	1,038.80	1
loredana ragolia	20981	31	via romita	93100	3498520663	9/2/1981	geometravilla@tiscali.it	8,999.10	1
Grand Total								**9,289.40**	**12**

134. As another example, Cluster 119427 (discussed above) includes three User Accounts with the name Alex Lopes. These three Alex Lopes User Accounts share no Aggregation Fields with any other User Accounts within that Cluster, including name.

135. This apparent disconnect between Dr. Freer's results and his explicitly stated methodology that requires at least fairly similar names and the existence of other markers of agreement further demonstrates the unreliability of his Aggregation Process.

v. The Freer Aggregation Process Results in Inconsistent and Speculative Outcomes

136. In order to determine the relative similarity of any two names, Dr. Freer relies on certain fuzzy matching algorithms that measure "string nearness."[137] With respect to the Name Field, Dr. Freer primarily relies on two different algorithms, *WRatio* and *ratio*, which he describes as follows:[138]

- *WRatio* is an "attempt to weight (the name stands for 'Weighted Ratio') results from different algorithms to calculate the 'best' score". It uses *ratio* as well as variants of *ratio* that are good at finding whole tokens in any order. (Again, a token is a sequence of non-space characters separated by a space; in "Quick brown fox", the tokens are "Quick", "brown", and "fox".) As a result, I use it for nearness matching on the cleaned version of the *rep_nome* field, which has multiple names in various orders, sometimes with typos (e.g., "maria teresa milagres neves", "maria teresa neves", "marai teresa neves", maria milagres", and "maria t m neves" may all refer to the same name).

- *ratio* is a standard string nearness measure that is part of the difflib Python package; difflib itself is a built-in Python library, which speaks to its ubiquity. ratio is used by both RapidFuzz and FuzzyWuzzy. The ratio nearness measure involves "find[ing] the longest contiguous matching subsequence that contains no "junk" elements; these "junk" elements are ones that are uninteresting in some sense, such as blank lines or whitespace … This does not yield minimal edit sequences, but does tend to yield matches that "look right" to people." ratio is good at avoiding typos, treats various parts of the string the same (unlike *jwsim*), and is good at grouping together edits that occur in a similar part of the string. As a result, I use it for nearness matching in the cleaned

[137] Freer Report at 10 and 38.

[138] Freer Report at 39-40.

versions of *rep_nome, rep_email, rep_cel, rep_fone, rep_cep, rep_numero, rep_login,* and *rep_cpf* fields, as well as the derived quantities of "first and last token of name" and "the part of the email address before the @ sign" that are defined using *firstLast* and *beforeAt* functions in Step 5 of the aggregation methodology.

137. However, fuzzy matching can also yield inconsistent and speculative results, especially in this case given that TelexFree, in part, spread organically through family members or communities that may share last names or even first names.[139]

138. For example, Dr. Freer treats User Accounts with the name Paulo Marinho and User Accounts with the name Rejane Marinho as separate Clusters,[140] despite there being little or no difference across most other Aggregation Fields other than first name.[141] However, based on Dr. Freer's selected fuzzy matching algorithms, these names have a *WRatio* and *ratio* of 67, which would mean that the pair would not have made it into the blocking group as part of Step 4.[142] Thus, it is likely that Rejane Marinho and Paulo Marinho were not aggregated together as a result of their first names not sharing the same letters in the same order, and in conjunction with the fact that their last names (while matching) were relatively shorter. Frantz Balan and Lusette Balan were likely not aggregated for the same reason.[143]

139. By comparison, and as discussed above, Dr. Freer treats User Accounts with the name Jean P. Carreiro and Eliane P. Carreiro as the same Participant, and similar to Rejane and Paulo Marinho, there is little or no difference across most of their other Aggregation Fields. However, based on Dr. Freer's selected fuzzy matching algorithms, the names Jean P. Carreiro and Eliane B. Carreiro have a *WRatio* and *ratio* of 82, which would mean that the pair would have made it into the blocking set as part of Step 4 and receive a rep_nome gamma value of 2 as part of his Step 5 .[144] Thus, it is likely that Jean and Eliane Carreiro were aggregated together as a result of their first names sharing more common letters in the same order, and in conjunction with the fact that their last names matched and were relatively longer.

[139] See e.g., Freer Report at 2.

[140] See Cluster 14635041 (Paulo Marinho) and Cluster 22681651 (Rejane Marinho).

[141] See **Exhibit 15**.

[142] Freer Report at 53.

[143] See **Exhibit 10**.

[144] Freer Report at 53 and 56.

140. This comparison demonstrates how the Freer Aggregation Process can lead to certain User Accounts being aggregated while others are kept apart not based on having "at least fairly similar names" in any meaningful or real-world sense, but arbitrarily based on, for example, whether two people's first names happen to share letters in a similar order.

vi. The Only Example Cluster in the Freer Report Contains Questionable Aggregations

141. The Freer Report contains a single example cluster, which Dr. Freer contends is a "visual illustration of the general quality of the dataset."[145] Specifically, Dr. Freer selected Cluster 132785, which includes User Accounts with the name Maria Neves or variations thereof.

142. However, even inspection of Dr. Freer's one selected example reveals potential issues with his Aggregation Process. For example, User Accounts with Rep_IDs 196685, 1902924 and 16337571 have Maria Neves as the name, but the contact information is for an individual with the email address of mmariashousecleaning@gmail.com, living at 155 Pleasant St., Marlboro, MA 01752 or 998 Pleasant St., Framingham, MA 01701, with the phone number 857-236-7654. None of these or the other Aggregation Fields for these User Accounts match Aggregation Fields found elsewhere in cluster 132785.[146]

143. For comparison, most of the other Maria Neves User Accounts in this cluster have an email address of ttmilagres@gmail.com, living at 2B Brackett Rd, Framingham, MA 01702 with the phone number 508-904-3810.[147] Dr. Freer appears to have failed to consider that two different Participants with the same name may have lived in the same city or state. I have not seen any evidence that Dr. Freer has attempted to perform any additional research or analysis to determine whether these Maria Neveses living at different addresses and with different contact information are, in fact, the same Participant.

144. Similarly, User Accounts 11501403 and 15376207 in this Cluster have the names Maria Lares and Rita Lares, with the contact information shown below:

[145] Freer Report at 24.

[146] See **Exhibit 16**. While City (for the Framingham address) and State match other User Accounts in the Cluster, City and State are not Aggregation Fields and the zip code (which is an Aggregation Field) still differs. I also note that these two User Accounts appear to have been generated around the same time as other User Accounts in this Cluster, making it less likely that it is a single Participant that changed address and phone number.

[147] See **Exhibit 16**.

cluster_id	rep_nome	rep_email	rep_ numero	rep_end	rep_cidade	rep_cep	rep_fone	Sum of net_equity	Count of rep_id
132785	rita lares	calderonr27@yahoo.com	4409	4409 merida st	fort worth	76133	8176924229	$ (49.90)	1
	maria lares	m.adame43@yahoo.com	2101	2101 w fuller ave	fort worth	76133	9132599771	$ (49.90)	1
Grand Total								$ (99.80)	2

145. None of these Aggregation Fields appear elsewhere in the Cluster. In fact, the only potential overlap is the rep_cpf field (i.e., social security number), which has a value of 999999999, and was similar to just five of the remaining 466 User Accounts in the Cluster (and appears over 3,400 times in the data overall).[148] So, in this case, Dr. Freer is aggregating accounts with different names, different contact information, and the only commonality is a clearly fictional social security number used just a handful of times elsewhere in the Cluster.

146. These potential issues identified in the one example Cluster that was specifically selected by Dr. Freer, and supposedly represents a "visual illustration of the general quality of the dataset,"[149] provides further evidence of the flaws in his Aggregation Process.

vii. Aggregation Errors Can Result in a Significant Overstatement of Participant Net Equity

147. With respect to the impact of potential linkage errors, the Statistical Data Warehouse Design Manual states:[150]

> Poor quality (eg if variables are missing, indecipherable, inaccurate, incomplete, inconstant, inconsistent) could lead to records not being linked – missed links – or being linked to wrong records – false links. The impact of these two types of errors may not be equal (eg a missed link may be more harmful than a false link), so this needs to be taken into account when designing a data linkage strategy, especially if the linking has legal or healthcare implications.

148. The numerous flaws in the Freer Aggregation Process, as described herein, also have the ability to significantly impact the Net Equity of supposed Net Winners.

149. For example, Cluster 129528 is comprised of five User Accounts with the name "sann" or "sann r" and is supposedly a Net Winner with Net Equity of $862,565. The largest User Account in this Cluster has an email address of sann@vicss.com and an address at 189 Squire Rd #40, Revere, MA.[151]

[148] Queries of the Freer and Huron Data Tables.

[149] Freer Report at 24.

[150] Kroeze, Harold. "Methodology: Data Linkage." European Commission, February 2017.

[151] See **Exhibit 17**.

150. Comparatively, Cluster 128225, is comprised of 839 User Accounts, many of which include names such as "Sanderley Devaconcelos", "Sanderley Vaconcelos", and "Sann Vasconcelos", an email address of sann@vicss.com and an address at 189 Squire Rd #40, Revere, MA, but this Account Aggregation is supposedly a Net Loser with Net Losses of ($363,822).[152]

151. Given that every Aggregation Field in Cluster 129528 is either the same as, or in substantial agreement with, an Aggregation Field in Cluster 128225, including name (e.g., "sann" vs. "sann vasconcelos"), I see no logical or factual basis for treating these as separate Participants. By appropriately combining these Account Aggregations together, this Participant's Net Equity would decrease substantially from $862,565 to $498,743, a reduction of over 40%.

152. As a second example, Cluster 10838191, a supposed Net Winner with Net Equity of $295,498, is comprised of three User Accounts, all with the name Oscar Enrique Sosa Esculpi, an email address of oscarsosa[####]@live.com or ososa_esculpi1@hotmail.com and an address at 1315 nw 98 ct suite 7, Miami, FL.[153]

153. Comparatively, Cluster 10811451, is comprised of 1,030 User Accounts the majority of which have the name Oscar Sosa, an email address of oscarsosa[####]@live.com or ososa_esculpi1@hotmail.com, and an address at 1315 nw 98 ct suite 7 in Miami, FL. But this Account Aggregation is supposedly a Net Loser with Net Losses of ($242,162). [154]

154. Given that every Aggregation Field in Cluster 10838191 is either the same as, or in substantial agreement with, an Aggregation Field in Cluster 10811451, including name, I see no logical or factual basis for treating these as separate Participants. By appropriately combining these Account Aggregations together, this Participant's Net Equity would decrease substantially from $295,498 to $53,336, a reduction of over 80%.

155. The significant impact of a single missed link on a Participant's concluded Net Equity is also apparent when comparing the Trustee's largest Net Loser and Net Winner Account Aggregations. For example, the top 100 Net Losers in TelexFree comprise a total loss of approximately $25.8 million, and have individual losses ranging between approximately $132,000

[152] See **Exhibit 17**.
[153] See **Exhibit 18**.
[154] See **Exhibit 18**.

and $3.1 million.[155] Comparably, the top 100 Net Winners in TelexFree comprise a total gain of approximately $148.1 million, and have individual winners ranging between approximately $937,000 and $4.1 million.[156]

156. Given the structure of TelexFree, there is a lower likelihood that a Participant would incur significant net losses on the scale reflected by the top Net Losers (i.e., repeatedly buy into TelexFree either directly or through Triangular Transactions but yet receive minimal or no gains).[157] One alternative rationale for the existence of some of these large Net Losers is the Trustee's failure to appropriately consolidate certain Net Loser Account Aggregations with corresponding Net Winners. To test this hypothesis, I manually reviewed the Potential Identifiers for the top five Net Losers (excluding the Cluster 25130588 for John Williams that Dr. Freer treats separately), and attempted to find matching Potential Identifiers among Net Winners. In all five cases, I found one or more Potential Identifiers that matched a Net Winner Account Aggregation. This evidence, which is summarized in **Exhibit 6** and detailed **Exhibits 9.1** to **9.5**,[158] includes instances where the same or similar values for Potential Identifiers, such as address, phone number, login, or email, appeared across both Net Loser and Net Winner Account Aggregations.

157. For each of the potential matches I found, I also queried the Huron Data Tables to determine whether certain transactions, such as Credit Transfers or Triangular Purchases/Receipts, occurred between the matched Account Aggregations. Based on my review of the SIG data in this matter, I found that there was often a high degree of activity between User Accounts held by the same Participant, a factor not considered by Dr. Freer as part of the Aggregation Process. As such, confirming that transactions occurred between the matched Account Aggregations, which it did in most instances, further strengthens the potential connections.

158. By way of example, and as reflected in **Exhibit 9.1**, the second largest Net Loser in TelexFree based on the Trustee's calculation of Net Equity is Du Painting DBA ("Du Painting"), which lost $1,018,291 and comprised 1,248 User Accounts. Across these User Accounts, Potential Identifiers for Du Painting included logins structured as "visionusa[##]," email addresses such as

[155] See **Exhibit 7**.
[156] See **Exhibit 8**.
[157] See e.g., **Exhibit 5**.
[158] Also, see **Exhibits 17-18**.

davidsonteixeira@hotmail.com, physical addresses such as 33 Vernal St., Everett MA, and phone numbers such as 781-484-7474 and 617-501-6788. A separate Participant under the Trustee's Aggregation Process, Parent ID 4032900, Davidson R. Teixeira, is shown to be a Net Winner of $1,544,804 comprised of 81 different User Accounts. Across Mr. Teixeira's User Accounts, every one of the noted Potential Identifiers for Du Painting can be found, which would provide strong indicia that both Account Aggregations represent the same Participant. As a result of what appears to be a missed link by the Freer Aggregation Process, Mr. Teixeira's Net Equity is overstated by more than $1 million, or approximately 193%, a very significant difference.[159]

159. As a second example, and reflected in **Exhibit 9.4**, I found that Shal Group (Cluster 12650741 and Net Loser of $712,458) shared many of the same or similar Potential Identifiers with Dolarex (Cluster 10992563 and Net Winner of $189,701). In fact, this linkage was confirmed by the claims submitted through the ePOC portal, as I can see that a single user (i.e., Saryas Jaff, UserId 26627) submitted two claims, one for Shal Group (ClaimId 42877) and one for Dolarex (ClaimId 42906).[160] Once this under-aggregation is appropriately considered and accounted for, of Dolarex changes from a $190,000 Net Winner to a $523,000 Net Loser under the Net Equity Calculation, a difference of over 275%.[161] This further demonstrates the significant impact a single aggregation error can have on a Participant's Net Equity.

160. In total, and simply based off the manual review of this sample population, I identified approximately $3.9 million in Net Loser aggregations that could potentially reduce the $4.4 million allegedly owed by the matched Net Winners, resulting in a reduction of almost 90%.[162]

161. As discussed above, one of the factors likely driving what appears to be under-aggregation in certain instances, is the disconnect between the Freer Aggregation Process's requirement that the names be the same or similar across every Participant User Account and the reality that Participants could create User Accounts using a variety of names, such as business name, personal name, nicknames, or even fictitious names. This reliance on consistent naming fails to

[159] Calculated as 1,018,291 / (1,544,804 – 1,018,291) = 193%.
[160] Queries of the ePOC Data Tables.
[161] See **Exhibit 9.4**.
[162] See **Exhibit 6**.

appropriately account for instances where other Potential Identifiers provide strong indicia that the User Accounts potentially belong to a single Participant.

162. It is also important to consider the impact of linkage errors in the appropriate context. While a given missed or false linkage may not represent a significant portion of TelexFree Net Equity in total, it can dramatically change the results for a single Participant.

viii. Dr. Freer has Failed to Deliver the Full Scope of His Engagement

163. As part of his scope of engagement, Dr. Freer was asked to "develop an aggregation methodology that creates one 'cluster' of User Accounts for each Participant."[163] Instead, Dr. Freer performed a purely mathematical exercise that attempts to link together User Accounts that have similar characteristics, and specifically where there is substantial agreement on the values in the Name Field.

164. However, such an approach fails to appropriately consider instances where the resulting Clusters may actually relate to more than one Participant, or where a Participant may relate to more than one Cluster. For example, in his deposition, Dr. Freer explicitly states, "[w]hen the name differs substantially, whether it's the same person or not that entered the accounts, that entered that information, when the name differs substantially, my understanding is those are to be considered as separate clusters."[164] However, the Court in this case already specifically defined Participant as a "person" who purchased a membership plan or VoIP Package in TelexFree,[165] not a "cluster" of User Accounts with similar values in the Name Field. The issue with conflating a person and cluster can be seen through the examples such as Sanderley Devaconcelos, Oscar Sosa and Frantz Balan, among others.

165. As such, to the extent that the ultimate purpose of Dr. Freer's analysis is a quantification of the supposed net funds received by each TelexFree Participant that comprises the Net Winner Class, Dr. Freer's analysis fails to accomplish this goal absent additional expert analysis. This was also confirmed by Dr. Freer in his deposition where he acknowledged "I'm not aggregating

[163] Freer Report at 2.

[164] Deposition of Cameron E. Freer, April 4, 2023 ("Freer Deposition") at 137.

[165] Support of Proposed Supplemental Order at 3

accounts to persons period. I'm aggregating accounts to other accounts in clusters. The assignment to persons is a different thing that might come later."[166]

B. Mr. Martin's Calculation of Net Equity is Unreliable

i. The Net Equity Calculation Fails to Account for Participant Investments and Receipts Related to Transfers of Credit

166. Relative to what constitutes Net Equity, I understand the Court in this matter stated the following:[167]

> (i) In determining the amount of claims of Participants, any claim or portion of claim based upon accumulated credits in a Participants' User Accounts as of the Petition Date shall be disallowed;
>
> (ii) The claims amounts of Participants shall be determined on a Net Equity basis, which shall be defined as follows: the amount invested by the Participant into the Debtors' scheme, including amounts paid [by the Participant] pursuant to Triangular Transactions, less amounts received by the Participant from the Debtors' scheme, including amounts received [by the Participant] pursuant to Triangular Transactions[;]
>
> (iii) In determining the amount of a claim of a Participant who has more than one User Account, the activity in all of the Participants' User Accounts shall be aggregated and netted against one another[.]

167. Importantly, while this order notes that claims based on accumulated Credits are disallowed, Net Equity should consider the amount *invested*[168] by a Participant in TelexFree less amounts *received* by a Participant from the TelexFree scheme. Based on how the Trustee and Mr. Martin described their determination of Net Equity, the distinction between accumulated Credits and amounts invested/received appears to be the exchange of actual monetary consideration, such as cash. For example, simply earning Credits through the placement of advertisements on classified websites would not entitle the Participant to claim the value of those Credits, but if that same Participant gave cash to another Participant as part of a Triangular Transfer, then the cash paid would indeed be considered as part of the Net Equity Calculation.

[166] Freer Deposition at 137.

[167] Supplemental Order at 2-3.

[168] For purposes of this report, I am using the term "invest" or "invested" to mean the payment of cash into the TelexFree scheme, such as through Direct Payments, Triangular Payments, Credit Transfers In, or Purchased Credits. No other definition or legal meaning is intended or implied.

168. However, the determination of Net Equity as executed by the Trustee and Mr. Martin does not account for instances where a Participant uses cash to purchase Credits from another Participant for purposes such as setting up additional User Accounts or performing subsequent Triangular Transactions. In both instances, a Participant is investing actual cash into the TelexFree scheme, yet the Net Equity Calculation treats these situations inconsistently by including Triangular Transactions and excluding Credit Transfers.

169. This dynamic between Triangular Transactions and Credit Transfers is also evident within the SIG data. For example, Mr. Balan, together with the accounts that I understand were incorrectly aggregated to his wife,[169] is alleged to be a $1.2 million Net Winner based upon the Net Equity Calculation performed by Mr. Martin, as shown in **Exhibit 10**. Most of those supposed winnings came in the form of Triangular Receipts, which totaled almost $1.5 million. However, in order to help generate the Credits necessary for this volume of Triangular Receipts, Mr. Balan also was a substantial net importer of approximately $836,000 in Credits, which I understand were not provided to Mr. Balan free of cost, as Mr. Martin and the Trustee presumes.[170] By including Triangular Transactions and excluding Credit Transfers, Mr. Martin is artificially increasing the supposed net amounts received by Participants and not properly considering the investments that were necessary in order to generate those supposed receipts.

170. There is no discussion in the Martin Report regarding why Triangular Transactions are considered within the determination of Net Equity but Credit Transfers are excluded when both appear to be substantively similar from a real-world (i.e., cash based) perspective. Additionally, to the extent any uncertainty exists regarding if, or how much, cash was exchanged as part of Credit Transfers, then such uncertainly necessarily extends to Triangular Transfers given that the SIG database does not provide evidence of cash payments for either transaction type. This internal inconsistency within the Net Equity Calculation between the treatment of Credit Transfers and Triangular Transactions results in a concluded Net Equity value that can differ greatly from the net amount actually invested and received by Participants.

[169] L. Balan Affidavit; F. Balan Affidavit at 3.

[170] F. Balan Affidavit at 6-7.

171. Credit Transfers between User Accounts and Credit Transfers as part of Triangular Transactions are also stored together within SIG from a data perspective. I understand that the "Transfer Table" contains information about transfers of Credits between User Accounts and includes a "tipo" that delineates "ten different transfer types, including requests to withdraw funds, payments of invoices with credits, and transfers of credits between accounts. The remaining types relate to chargebacks and other adjustments."[171]

172. Further, the transfer of Credits in exchange for cash is not just a hypothetical scenario that occurs infrequently; rather, the SIG data reflects vast transfers of Credits both between distinct Participants and internally within a Participant's User Accounts. This is especially true for many of the Participants that Mr. Martin determined to be large Net Winners and Net Losers, as shown in **Exhibit 8** and **Exhibit 7**, respectively. For example, the top 100 Net Winners made over $210 million in Credit Transfers Out and over $250 million in Credit Transfers In (i.e., significant net importers of Credits).[172] By comparison, the top Net Losers made $32 million in Credit Transfers Out and $19 million in Transfers In (i.e., significant net exporters of Credits).[173]

173. The inclusion of Credit Transfers is highly impactful to the determination of a Participant's Net Equity as in many cases it can change a Participant's designation as a Net Winners or Net Losers, a sample of which is shown below:[174]

Participant Type	**Cluster**	**Net Equity Excluding Transfers In/Out**	***6 - Credit Transfers In***	***7 - Credit Transfers Out***	**Net Equity Including Transfers In/Out**
NetLoser	13022279	**($1,018,291)**	*($578,327)*	*$845,237*	**($751,381)**
NetLoser	11849449	**($622,342)**	*($1,904,355)*	*$2,254,352*	**($272,344)**
NetLoser	12242105	**($481,228)**	*($621,865)*	*$1,008,813*	**($94,279)**
NetLoser	10525055	**($240,773)**	*($768,138)*	*$1,548,862*	**$539,950**
NetLoser	1020407	**($210,525)**	*($5,147,691)*	*$6,163,099*	**$804,884**
NetWinner	132785	**$4,087,745**	*($7,290,296)*	*$6,634,688*	**$3,432,137**
NetWinner	2522900	**$4,012,999**	*($7,135,275)*	*$3,926,183*	**$803,907**
NetWinner	323967	**$3,147,693**	*($5,155,263)*	*$3,265,763*	**$1,258,193**
NetWinner	1385480	**$1,494,466**	*($3,220,534)*	*$1,554,200*	**($171,868)**
NetWinner	10853999	**$1,308,377**	*($1,581,410)*	*$176,389*	**($96,644)**

[171] Huron TelexFree Analysis of Damages Presentation at 29.

[172] See **Exhibit 8**.

[173] See **Exhibit 7**.

[174] See **Exhibit 7** and **Exhibit 8**.

174. It defies economic logic that one Participant would simply give Credits to another Participant without financial consideration in return (particularly in these large sums) given that Credits could be converted into cash through Direct Receipts or, much more frequently, Triangular Transactions. It is also worth noting that TelexFree appears to have charged the recipient of Credit Transfers a $3 fee per transaction, demonstrating that such activities were not without a cost.

175. There is also direct evidence of Credits transferred in exchange for cash based on descriptions of events noted by users through the ePOC portal, such as:

- Ana Patricia Salmeron de Pereira (ePOC ClaimId 68594)[175]
    - "Through the account of patriciaycarlos **I was buying the credit for other participants with real cash, so i could accumulate a large amount of credit in this account, to allow me to easily create new accounts for new members**. As this was the fastest method. Therefore when Telexfree got shutdown, all of this accumulated money ($90,038) had been distributed and ultimately given to Telexfree through accounts and Voips. So I never actually had this amount of money, therefore all I am owed for this account is the original $1425 that I paid."[176] **(emphasis added)**
    - The transactions reflected in SIG confirm that Ms. Salmeron de Pereira had $100,237 in total Credit Transfers In, of which $83,507 came from her login "patriciayalex" or "patriciaycarlos."[177]
    - If Credit Transfers In and Out were considered as part of Net Equity, Ms. Salmeron de Pereira's Net Winnings of $104,998 would be reduced to $31,856, a decrease of ($73,142), or 70%.
- Marcel McFarren (ePOC ClaimId4282)[178]
    - "**I gave money to a friend of mine Diego Alonso Lizama Zegarra on several opportunities who transfered telexfree credits to my accounts in order to purchase new accounts and VOIP plans**. I currently don't remember or have records of all the transfers I received or the dates nor individual amounts. But **in the time I participated in telexfree I paid in cash [$30,383.40] US dollars that where transferred to my account from other user accounts**. Wich

[175] See **Exhibit 20.1**.

[176] Queries of the ePOC Data Tables: UnmatchedAccount table, UnmatchedAccountId=6961, Description field.

[177] Queries of the Freer and Huron Data Tables: Cluster 12902527.

[178] See **Exhibit 20.2**.

[$26,193.60] were used in opening user accounts and voip accounts, [$597.00] US dolars in user account maintenance and [$3,592.80] US dolars in voip renewals."[179] **(emphasis added)**

- The transactions reflected in SIG confirm that Mr. McFarren had $35,028 of Credit Transfers In from User accounts Mr. McFarren.[180]
- If Credit Transfers In and Out were considered as part of Net Equity, Mr. McFarren's Net Winnings of $2,736 would be reduced to a Net Loss of ($21,924), a decrease of ($24,660) or approximately 900%.

- Jose Norford (ePOC ClaimId 150603)[181]
  - "This claim indicates that I owe approximately [$202,698.80] however I should be receiving a reimbursement of approximately [$32,775.00]."[182]
  - The transactions reflected in SIG confirm the determined Net Equity to be $202,698.80 under Mr. Martin's Aggregation Process (i.e., the basis for ePOC). However, when Credit Transfers In and Out are considered, Mr. Norford's Net Equity under Mr. Martin's Aggregation Process is reduced to $39,118.12.[183]
  - If Credit Transfers In and Out were considered as part of Net Equity under the Freer Aggregation Process, Mr. Norford's Net Winnings of $199,749 would be reduced to Net Winnings of $23,844, a decrease of ($175,905) or 88%.

176. This exclusion of Credit Transfers fails to appropriately consider the "amount invested" or "amounts received" by Participants within the determination of Net Equity, and, as a result, the Net Equity determined by Mr. Martin is not reliable for the over 435,000 Account Aggregations that took part in Credit Transfer transactions, of which approximately 73,000 are currently alleged Net Winners.[184]

[179] Queries of the ePOC Data Tables: UnmatchedAccount table, UnmatchedAccountId=23413, Description field.

[180] Queries of the Freer and Huron Data Tables: Cluster 13094257.

[181] See **Exhibit 20.3**.

[182] Queries of the ePOC Data Tables: UnmatchedAccount table, UnmatchedAccountId=13886, Description field.

[183] Queries of the Freer and Huron Data Tables: Cluster 10789133.

[184] Queries of the Freer and Huron Data Tables.

ii. The Amount of Cash Invested/Received by Participants as Part of Triangular Transactions or Credits Transfers Can Differ from Amount of Credits Recorded in SIG

177. With respect to Triangular Transactions, Mr. Martin describes the process as follows:[185]

> To redeem credits for a recruited Participant a new User Account was opened for the recruited Participant and the recruiting Participant satisfied the resulting membership fee with credits held by the recruiting Participant (a "Triangular Transaction").
>
> Based upon interviews with Participants and others, the recruited Participant **typically** paid the membership fee in cash to the recruiting Participant who, in turn, satisfied the recruited Participant's membership fee with their own credits. **(emphasis added)**

178. This description is exemplary of why the Net Equity Calculation is not reliable. Mr. Martin and the Trustee assume that the amount of Credits transferred as part of a Triangular Transaction is equivalent to the amount of cash invested or received by Participants. However, the assumption that cash was exchanged is based on interviews with what could only be a very small percentage of the total 2.1 million TelexFree Participants according to Mr. Martin (or 1.6 million TelexFree Participants according to Dr. Freer), many of whom are outside of the U.S., and even then, cash exchanges only "typically" occurred.[186]

179. Mr. Martin does not provide any specific guidance regarding what "typically" means in this context. For example, was cash exchanged between Participants as part of Triangular Transactions 90% of the time, or less than 75%? In those instances where cash was exchanged, was it always the exact amount of, and equivalent to, the amount of Credits recorded within SIG, the data source which forms the basis for the Trustee's determination of Net Equity? Given the massive size and volume of Triangular Transactions that occurred on a global scale (i.e., more than $2.7 billion dollars and 13.3 million transactions for Triangular Receipts alone),[187] much of which I understand was transacted in physical cash,[188] the assumption that Credits recorded in SIG are necessarily equivalent to investments or receipts by Participants is unsupported.[189]

[185] Martin Report at 7.

[186] See, **Exhibit 3**. I understand that Counsel requested the notes from these interviews, but was informed that notes were not compiled.

[187] See **Exhibit 4**.

[188] See e.g., F. Balan Affidavit at 5.

[189] F. Balan Affidavit at 5.

180. In direct contradiction to this assumption, I understand that Frantz Balan, a supposed Net Winner according to the Trustee,[190] affirmed the following:[191]

> It is a mistake to assume that 100% of the face amount of invoices were collected 100% of the time. In practice, promoters like me rarely received the full amount in cash. I would estimate for me and the promoters I worked with or spoke to that at best this happened only 10% of the time.
>
> Most often, promoters were only able to ask for steeply discounted payments, as low as $250 for an AdCentral Family membership. Most commonly, the amount requested ranged from 50% to 75% of the invoiced amount.

181. Additionally, I understand that Participants were often part of larger teams working alongside or employing other Participants. In these instances, I understand each team, and each pair of Participants within a team, had their own compensation arrangements and there was no universal payment practice.[192]

182. As one approach to teaming, I understand that it was the role of the lower-level recruiting Participants of the team to identify and signup new members; however, the responsibility of satisfying the resulting invoices with Credits generally fell to the higher-level team leads in the TelexFree pyramid structure. The recruiting Participants would also collect the cash from the new recruits (frequently at a discount relative to the amount of Credits, as noted above), and then distribute a percentage of the cash collected to the lead Participant and keep the remainder as a fee.[193] In this way, Participants were actually sharing compensation in a manner not reflected in the SIG data.

183. The information within SIG would not reflect this real-world dynamic, making it a potentially unreliable source for purposes of determining Net Equity. As a result, the Trustee's concluded Net Equity for the lead Participant would be overstated relative to cash, and the Net Equity for the recruiting Participant would be understated.

184. Also, as it relates to teaming, I'm not aware of anything that would have prevented two people from potentially sharing User Accounts and splitting any cash investments and receipts

[190] Queries of the Freer and Huron Data Tables.

[191] F. Balan Affidavit at 5.

[192] F. Balan Affidavit at 4.

[193] F. Balan Affidavit at 6.

associated therewith. There is evidence that this may have occurred based upon logins such as in the "patriciayalex" or "patriciaycarlos."[194]

185. Further, as part of my manual review of top Net Winners and Net Losers (discussed above), I explored certain of the underlying transaction detail to help me better understand the potential connections between top Net Losers and Net Winners. For example, the largest supposed Net Loser, "John Williams," is alleged to have lost over $3.1 million (Cluster 25130588).[195] This Account Aggregation according to Dr. Freer is comprised of 62,725 underlying User Accounts, every one of which is a VoIP Phone Plan with a value of $49.90 purchased through a Triangular Payment.[196] The chance that a single person bought in excess of 62,000 VoIP Phone Plans for over $3.1 million during the approximately two-week period of March 30, 2014 to April 13, 2014[197] (a process that would likely have taken one person months of full-time work to accomplish since each VoIP Phone Plan required the creation of a new User Account) does not stand up to common sense logic.[198]

186. An alternative explanation lies in the timing of the first User Account in the "John Williams" Account Aggregation (i.e., March 30, 2014). I understand that on or around March 9, 2014, TelexFree changed its compensation plan such that Participants were required to sell at least one VoIP Phone Plan, which many did through Triangular Transactions.[199] It is appears that "John Williams" is the result of this modification to TelexFree's policy. Whether the creation of these 62,725 User Accounts were done by a service company, an automated software product, or by some other means is not clear from the data, but it does introduce significant uncertainty regarding whether "John Williams" truly made over 62,000 cash payments of $49.90 to what the SIG data shows to be over 1,800 unique counterparty Participants in the span of less than month.[200] To the extent that such amounts were not actually paid by "John Williams" to the counterparty

[194] Queries of the Freer and Huron Data Tables: Cluster 12902527.

[195] See **Exhibit 7**.

[196] Queries of the Freer and Huron Data Tables.

[197] Queries of the Freer and Huron Data Tables.

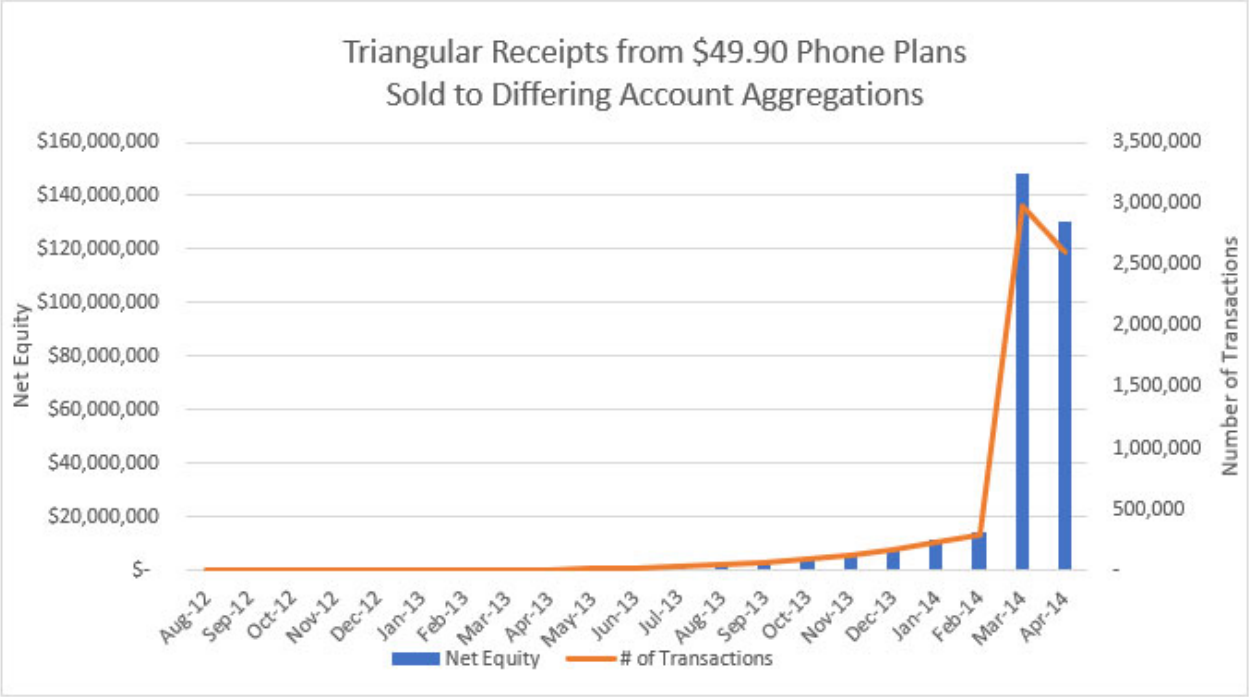
[198] This same fact pattern exists for other Account Aggregation as well, including at least "Ming Hua" (Cluster 21036267). See **Exhibit 7** and **Exhibit 9.2**.

[199] www.sec.gov/litigation/complaints/2014/comp-pr2014-79.pdf; www.justice.gov/usao-ma/pr/former-president-telexfree-sentenced-billion-dollar-pyramid-scheme; queries of the Huron Data Tables.

[200] Queries of the Freer and Huron Data Tables.

Participants, then the Net Equity for those Participants is necessarily overstated. This is due to the fact that each supposed $49.90 Triangular Payment by "John Williams" would have a corresponding Triangular Receipt associated with the counterparty Participant in the SIG data, which is a component of the Net Equity Calculation.

187. In fact, when all TelexFree Triangular Receipts of $49.90 are charted over time a demonstrable spike beginning in March 2014 becomes clearly apparent, as shown in the table below:



188. During just the post-March 9, 2014 period, there are over 5.4 million Triangular Receipts for $49.90 between User Accounts supposedly belonging to different Participants, resulting in supposed net winnings of more than $270 million.[201] As noted above relative to "John Williams," there is significant uncertainty (and no support in the SIG data) regarding whether Participants actually received the associated cash despite the increase to their Net Equity.

189. Additionally, it is not clear to what extent the Net Equity calculation considers the potential impact of foreign exchange rates. Mr. Martin identifies over 1.7 million Net Losers outside the

[201] Queries of the Freer and Huron Data Tables.

United States across more than 200 different countries.[202] However, I understand that the Credit amounts within SIG are intended to reflect U.S. dollars (i.e., USD). As such, when non-US Participants presumably exchange cash in local currency, it is neither clear what exchange rate may have been used at the time, nor how changes in those exchange rates between the time of the transactions and today may disproportionately impact certain Participants.

190. For example, Mr. Martin indicates that there were approximately 950,000 victims in Brazil.[203] The exchange rate today is approximately 5 Brazilian Reais (i.e., BRL) to 1 USD. By comparison, the exchange rate in 2014 was approximately 2.4 BRL to 1 USD.[204] Therefore, depending on whether you are a Net Winner or Net Loser, the amount of your claim or liability could differ from the original net transaction amount invested or received by as much as 100%.

191. The potential inconsistency between the actual value of the cash exchanged and the amount of Credits recorded for Triangular Transactions within SIG is equally true with respect to Credit Transfers. Given this uncertainty regarding a foundational assumption of the Net Equity calculation, which is that the Credit amounts reflected in SIG are a reliable and accurate proxy for actual cash investments and receipts, the resulting Net Equity amounts concluded by the Trustee are equally unreliable in this instance.

VIII. CONCLUSION

I have assessed the methodologies set forth in the Freer Report and Martin Report relative to: (i) the Freer Aggregation Process; and (ii) Mr. Martin's Net Equity Calculation for Participants, and specifically Net Winners. As addressed in detail above, both the Aggregation Process and the Net Equity Calculation contained flawed assumptions that render Dr. Freer's and Mr. Martin's, and ultimately the Trustee's, conclusion regarding the Net Equity amounts supposedly won or lost by a given TelexFree Participant to be unreliable.

[202] Huron TelexFree Analysis of Damages Presentation at 8 and 18.

[203] Huron TelexFree Analysis of Damages Presentation at 18.

[204] https://www.ofx.com/en-us/forex-news/historical-exchange-rates/yearly-average-rates.