COMMERCIAL TELEPHONE SAMPLES AND COMPONENT OUTCOME RATES

Claude Comeau, Peter Mariolis, Ph.D. [1]

ABSTRACT

Outcome rates in telephone surveys are usually based on an entire sample. Telephone samples from commercial sample vendors, however, contain identifiable subsets of records with very different probabilities of obtaining particular dispositions. In such a case, component outcome rates could vary in ways unrelated to rates based on the entire sample. The 2000 Behavioral Risk Factor Surveillance System (BRFSS) survey is used to examine the degree to which selected outcome rates by state for different subsets of records correlate with corresponding global rates. Although correlations tend to be large, some cases are not, making it worthwhile to examine component outcome rates.

KEY WORDS: Telephone surveys; Outcome rates; Data quality.

1. INTRODUCTION

Outcome rates in telephone surveys are usually based on an entire sample. Telephone samples from commercial sample vendors, however, generally contain identifiable subsets of records with very different probabilities of obtaining particular dispositions. In such a case, component outcome rates could vary in ways unrelated to rates based on the entire sample. Two important distinctions in this regard are between listed and not-listed telephone numbers and between one-plus block and zero block telephone numbers.

Commercial sample vendors have access to databases that identify listed household telephone numbers. All other telephone numbers are not-listed. A hundred block is a set of 100 telephone numbers with the same area code, prefix, and first two digits of the suffix. From the listed number databases, the vendors can determine the number of listed household telephone numbers in any hundred block. One-plus block telephone numbers are telephone numbers (which may be listed or not-listed) from hundred blocks with one or more listed household telephone numbers, whereas zero block telephone numbers are telephone numbers (all of which are not-listed) from hundred blocks with no listed household telephone numbers. Thus, we can distinguish among listed household telephone numbers, not-listed one-plus block telephone numbers, and zero block telephone numbers.

We examine three rates by component: the CASRO (Council of American Survey Research Organizations) response rate, the household detection rate, and the household completion rate. (CASRO, 1982) The CASRO rate is the number of completed interviews divided by an estimate of the number of eligible households in the sample.[2] The household detection rate is the number of identified households divided by the total number of records in the sample. The household completion rate is the number of completes divided by the number of identified households in the sample. The household detection and household completion rates reflect the two key aspects of survey participation distinguished by Groves and Couper

---

[1] Claude Comeau, Comeau Associates, 177 Ball Hill Road, Milford, New Hampshire, USA 03055
Peter Mariolis, Centers for Disease Control and Prevention (MS K66), 4770 Buford Hwy, NE, Atlanta, Georgia, USA 30341-3717

[2] The number of eligible households is estimated by assuming that the proportion of eligible households among records whose status is unknown (no answer and busy dispositions) equals the proportion of eligible households among records whose status is known. The number obtained by this calculation is added to the number of identified eligible households in the sample.

(1998)—contact and cooperation. The usefulness of component outcome measures is not exhausted by these three rates; we also mention actual or possible uses of other outcome measures.

This paper uses data from the 2000 Behavioral Risk Factor Surveillance System (BRFSS) survey to examine the associations among global and component CASRO, household detection, and household completion rates. The data in this paper are used to address a single basic question: Do component outcome rates vary (linearly) in ways unrelated to global outcome rates?

2. DATA AND METHODS

The BRFSS survey is a joint venture of the Centers for Disease Control and Prevention (CDC) and health departments in the 50 states, the District of Columbia, and Puerto Rico.[3] It is a (usually) monthly telephone survey whose purpose is to track the prevalence of behaviors related to chronic diseases and preventive health practices among the civilian, non-institutionalized population 18 years of age or older in each state. Health topics include diet, smoking, physical activity, and alcohol consumption. CDC coordinates the development of a set of core questions which are asked by every state and standardized sets of questions on specific topics which the states can choose to ask; in addition, each state is free to ask any additional questions it chooses. CDC also coordinates the development of standards for sample designs and data collection procedures and provides technical assistance to the states. The sampling frame for the BRFSS consists of all telephone numbers of NXX types 00, 50, 51, 52, and 54, including zero-block numbers. BRFSS guidelines prescribe up to 15 callbacks for unresolved numbers distributed over weekday, weeknight, and weekend calling occasions. One adult is randomly chosen from each eligible household; no proxy interviews are allowed. The states are responsible for data collection. In 2000, 36 states contracted out data collection to commercial or university survey research organizations; in the other 16, a unit of the state health department conducted the data collection. Once the data are collected and initially edited, they are sent to CDC. CDC conducts further editing and, at the end of each year, weights the data and returns them to the states along with several reports. CDC then makes the aggregate data set available to the public. Additional information on the BRFSS may be found at http://www.cdc.gov/nccdphp/brfss.

Although BRFSS data are published in an annual data set, data collection is conducted monthly by each state. (There are exceptions: Michigan collects data quarterly and some states may miss up to 4 months. Also, most surveys are completed within a single month, but occasionally the survey period may extend into the following month or, rarely, beyond.) Thus, each month's worth of data in each state can be considered a separate survey. That is the approach taken in this paper.

The data for this study consist of 1,674,110 records from 49 states (the District of Columbia and the 50 states except for Minnesota and Wisconsin) that used a list-assisted sample design in 2000. The individual records were aggregated into 564 monthly surveys, resulting in up to 12 records per state. The sample design stratifies by one-plus or zero block status. One-plus block telephone numbers are generally sampled at four times the rate of zero block telephone numbers. Monthly sample records are generated quarterly. Sample composition is measured by the percentages of listed, not-listed one-plus block, and zero-block numbers. The formulas for the CASRO, household detection, and household completion rates in terms of final disposition codes used in the BRFSS are provided in Figure 1.

The analytical strategy used is to determine, for each state separately, the Pearson product-moment correlations between each global outcome measure and its components. This strategy controls for differences in data collectors, population characteristics, and other factors to the extent that they are associated with interstate differences. Each correlation coefficient is based on 4 to 12 records. The main data presented in this paper are the distributions of those correlation coefficients for each outcome measure.

---

[3] Hereafter, "states" refers to the 50 states, the District of Columbia, and Puerto Rico.

Figure 1. Formulas for Outcome Measures Using 2000 BRFSS Final Disposition Codes

CASRO Rate

$$\frac{01}{\left[(01+02+07+09)+\dfrac{(01+02+07+09)}{(01+02+07+09)+(03+05+06+08+11)}\times(04+10)\right]}$$

Household Detection Rate

$$\frac{(01+02+06+07+08+09+11)}{(01+02+03+04+05+06+07+08+09+10+11)}$$

Household Completion Rate

$$\frac{01}{(01+02+06+07+08+09+11)}$$

BRFSS Final Disposition Codes

01	Completed interview	07	Selected respondent not available during the interviewing period
02	Refused interview	08	Language barrier
03	Nonworking number	09	Interview terminated within questionnaire
04	Ring no answer	10	Line busy
05	Not a private residence	11	Respondent unable to communicate due to physical or mental impairment
06	No eligible respondent at this number		

3. RESULTS

For all states and for all months of data submission, a mean of 21.0% of sample records are listed telephone numbers (range, 8.7% - 32.9%) (Figure 2). The mean for an individual state ranges from 10.4% to 30.3%, with an overall mean of 21.0%. The mean range of percent listed within a state is 3.5 percentage points (range, 1.4 percentage points - 5.8 percentage points). Because the sample was stratified by block status, the distribution of one-plus and zero block telephone numbers by state and submission month vary little from month to month. Thus, the variability of the distribution of not-listed one-plus block telephone numbers by state and submission month is almost identical to the variability in the distribution of listed telephone numbers by state and submission month.

The correlations between the global and listed number CASRO rates range from .14 to .99 (median, .89); the correlations for all but two states are .59 or above (Table 1). The correlations between the global and not-listed one-plus block number CASRO rates are smaller and more variable: they range from -.14 to .98 (median, .78) and four states have correlations between -.14 and .24. The correlations between the global and zero block number CASRO rates are even smaller and more variable: they range from -.52 to .76 (median, .12) and, unlike the other distributions, the distribution of zero block correlations is not highly skewed.

The correlations between the global and listed number household detection rates range from -.05 to .999 (median, .74) (Table 2). The correlations between the global and not-listed one-plus block number household detection rates are larger and less variable: although they range from -.63 to .97 (median, .78), the interquartile range is .16 (.24 for listed records) and all but two states have correlations of .35 or above. The correlations between the global and zero block number household detection rates are even smaller and more variable: they range from -.43 to .87 (median, .24) and, unlike the other distributions, the distribution of zero block correlations is not highly skewed.

The correlations between the global and listed number household completion rates range from .25 to .99 (median, .91); the correlations for five states are at or below .61 (Table 3). The correlations between the global and not-listed one-plus block number household completion rates are smaller and more variable: they range from -.06 to .98 (median, .76); four states have correlations between -.06 and .28. The

correlations between the global and zero block number household completion rates are even smaller and more variable: they range from -.53 to .81 (median, .20) and, unlike the other distributions, the distribution of zero block correlations is not highly skewed.

4. DISCUSSION

The results in this paper show that although outcome rates for listed and not-listed one-plus block numbers are in general highly correlated with corresponding global outcome rates, in some cases they are not. This suggests that it can be worthwhile to look at outcome rates separately for listed and not-listed one-plus blocks. The situation is clearer for zero-block numbers: most states have small or moderate associations between zero block and global outcome rates, so calculating separate outcome rates for zero block numbers will generally result in outcome rates for zero block numbers that provide data statistically unrelated to the corresponding outcome rates for all records.

One way to use component rates is to compare them with each other. The CASRO and household completion rates are generally highest for listed numbers, slightly smaller for not-listed one-plus block numbers, and smallest for zero block numbers (data not shown). Deviations from this pattern could be used as a signal to further examine data accuracy. Further work is needed to examine the relationships among component measures and how different patterns can come about.

The differences in component rates also raise the question of how each rate is related to data accuracy. Outcome rates can be affected by factors extraneous to data accuracy, for example the extent to which true nonworking numbers return a tritone. It is possible that outcome rates for listed numbers are less affected by extraneous factors than outcome rates for other component measures. Listed numbers were at one time households. If rates of change from household to non-household status are relatively constant across states, then differences in response rates among listed numbers would more directly reflect differences in true response rates than would differences among global or other component rates. Again, further work is needed to determine the extent to which this may be the case.

The calculation of component rates has been discussed in terms of rates for an entire study. Calculation of component rates may be even more useful when examining outcomes by interviewer. Use of component rates could help control differences in the sample used by individual interviewers. For example, anomalous patterns of outcome rates for individual interviewers could help identify interviewers needing improvement.

Finally, an application of component rates in production may be noted. Because efficiency rates will vary by component, the number of completed interviews from a given set of sample records can be more precisely estimated by calculating and applying efficiency rates separately for each component than by applying global efficiency rates. Commercial sample providers can indicate the component status of sample records on those records on request.

REFERENCES

Council of American Survey Research Organizations (1982), *Report of the CASRO Completion Rates Task Force*, New York: Audits and Surveys Company.

Groves, R. M., and M. P. Couper (1998), *Nonresponse in Household Interview Surveys*, New York: Wiley.

Figure 2. Percent of Sample Records That Are Listed Telephone Numbers, by State and Month of Submission

Percent Listed Telephone Numbers

Table 1. Distributions of Correlations Between Global and Component CASRO Rates

Listed Records

Quantile	Estimate
100% Max	0.994123
99%	0.994123
95%	0.978651
90%	0.970352
75% Q3	0.938315
50% Median	0.889336
25% Q1	0.789230
10%	0.673800
5%	0.589663
1%	0.140783
0% Min	0.140783

```
Stem Leaf              #    Boxplot
9 555666778889        12       |
9 00222233344         11    +-----+
8 56667899             9    *-----*
8 0124                 4    |  +  |
7 677889               6    +-----+
7 33                   2       |
6 7                    1       |
6 3                    1       |
5 9                    1       |
5
4
4
3
3
2
2
2
1 6                    1       *
1 4                    1       *
  ----+----+----+----+
Multiply Stem.Leaf by 10**-1
```

Not-listed One-Plus Block Records

Quantile	Estimate
100% Max	0.976797
99%	0.976797
95%	0.944140
90%	0.924239
75% Q3	0.889132
50% Median	0.780922
25% Q1	0.619279
10%	0.450936
5%	0.159089
1%	-0.141714
0% Min	-0.141714

```
Stem Leaf              #    Boxplot
9 78                   2       |
9 0011122234          10       |
8 556799              6     +-----+
8 12244               5     |     |
7 6789                4     *-----*
7 001                 3     |  +  |
6 5688                4     |     |
6 112234              6     +-----+
5 678                 3       |
5 0                   1       |
4 5                   1       |
4
3
3
2 4                   1
2
1 6                   1       0
1 0                   1       0
0
0
-0
-0
-1 4                  1       0
  ----+----+----+----+
Multiply Stem.Leaf by 10**-1
```

Zero Block Records

Quantile	Estimate
100% Max	0.7634190
99%	0.7634190
95%	0.6074270
90%	0.5710643
75% Q3	0.3672313
50% Median	0.1198445
25% Q1	-0.0850833
10%	-0.2725661
5%	-0.4198396
1%	-0.5205990
0% Min	-0.5205990

```
Stem Leaf              #    Boxplot
7 66                   2       |
6 1                    1       |
5 03578                5       |
4 066                  3       |
3 02478                5     +-----+
2 01589                5     |     |
1 112356               6     *--+--*
0 25679                5     |     |
-0 997640              6     +-----+
-1 9644                4       |
-2 8732                4       |
-3
-4 62                  2       |
-5 2                   1       |
  ----+----+----+----+
Multiply Stem.Leaf by 10**-1
```

Table 2. Distributions of Correlations Between Global and Component Household Detection Rates

Listed Records

Quantile	Estimate
100% Max	0.9992285
99%	0.9992285
95%	0.9756642
90%	0.9351920
75% Q3	0.8488764
50% Median	0.7425905
25% Q1	0.6096933
10%	0.3287779
5%	0.2038621
1%	-0.0469052
0% Min	-0.0469052

```
Stem Leaf      #             Boxplot
10  0          1                |
 9  689        3                |
 9  0014       4                |
 8  55689      5             +-----+
 8  01113      5             |     |
 7  556677     6             |     |
 7  00224      5             *-----*
 6  589        3             | + |
 6  11244      5             +-----+
 5  9          1                |
 5  3          1                |
 4  5          1                |
 4  334        3                |
 3  5          1                |
 3  3          1                |
 2
 2  00         2                0
 1  5          1                0
 1
 0
 0
-0
-0  5          1                0
    ----+----+----+----+
Multiply Stem.Leaf by 10**-1
```

Not-listed One-Plus Block Records

Quantile	Estimate
100% Max	0.969071
99%	0.969071
95%	0.964890
90%	0.947907
75% Q3	0.894336
50% Median	0.783756
25% Q1	0.633399
10%	0.449564
5%	0.354852
1%	-0.628246
0% Min	-0.628246

```
Stem Leaf              #        Boxplot
 9  000022355677      12           |
 8  345556788899      12        +-----+
 7  044566778          9        *--+--*
 6  0033479            7        +-----+
 5  337                3        |     |
 4  256                3        |     |
 3  5                  1        |     |
 2
 1
 0
-0  3                  1           0
-1
-2
-3
-4
-5
-6  3                  1           *
    ----+----+----+----+
Multiply Stem.Leaf by 10**-1
```

Zero Block Records

Quantile	Estimate
100% Max	0.87189142
99%	0.87189142
95%	0.78807310
90%	0.71774440
75% Q3	0.49626542
50% Median	0.23776774
25% Q1	0.00134082
10%	-0.20442645
5%	-0.30046383
1%	-0.43244425
0% Min	-0.43244425

```
Stem Leaf      #             Boxplot
 8  27         2                |
 7  0249       4                |
 6  03         2                |
 5  05669      5             +-----+
 4  1257       4             |     |
 3  33577      5             |     |
 2  003488     6             *--+--*
 1  0255       4             |     |
 0  00588      5             +-----+
-0  865        3                |
-1  5430       4                |
-2  30         2                |
-3  60         2                |
-4  3          1                |
    ----+----+----+----+
Multiply Stem.Leaf by 10**-1
```

Table 3. Distributions of Correlations Between Global and Component Household Completion Rates

Listed Records

Quantile	Estimate
100% Max	0.993897
99%	0.993897
95%	0.979648
90%	0.972723
75% Q3	0.947290
50% Median	0.909449
25% Q1	0.848864
10%	0.614562
5%	0.519546
1%	0.249886
0% Min	0.249886

Stem	Leaf	#	Boxplot
9	5555566777889	13	+-----+
9	00011111222234444	16	*-----*
8	56889999	8	+--+--+
8	01344	5		
7	7	1		
7	3	1		
6			
6	1	1	0
5	6	1	0
5	02	2	*
4			
4			
3			
3			
2	5	1	*
2			

Multiply Stem.Leaf by 10**-1

Not-listed One-Plus Block Records

Quantile	Estimate
100% Max	0.9794462
99%	0.9794462
95%	0.9470094
90%	0.9347133
75% Q3	0.8766865
50% Median	0.7644796
25% Q1	0.6061401
10%	0.4951602
5%	0.0930550
1%	-0.0595995
0% Min	-0.0595995

Stem	Leaf	#	Boxplot
9	588	3	
9	011334	6	
8	56668888	8	+-----+
8	02244	5		
7	66689	5	*-----*
7	34	2		
6	5666689	7		+
6	01	2		
5	88999	5	+-----+
5	04	2		
4			
4			
3			
2	8	1	
2			
1			
1			
0	9	1	0
0	1	1	0
-0			
-0	6	1	0

Multiply Stem.Leaf by 10**-1

Zero Block Records

Quantile	Estimate
100% Max	0.8091746
99%	0.8091746
95%	0.7448083
90%	0.5754738
75% Q3	0.4027052
50% Median	0.2007265
25% Q1	-0.0924296
10%	-0.2909217
5%	-0.4893015
1%	-0.5329234
0% Min	-0.5329234

Stem	Leaf	#	Boxplot
8	1	1	
7	45	2	
6	9	1	
5	08	2	
4	0455699	7	+-----+
3	01244568	8		
2	0235	4	*-----*
1	1446667	7		+
0	445	3		
-0	91	2	+-----+
-1	65	2		
-2	985532	6		
-3	2	1	
-4	9	1	
-5	32	2	

Multiply Stem.Leaf by 10**-1