Proceedings of Statistics Canada Symposium 2001
Achieving Data Quality in a Statistical Agency: A Methodological Perspective

DATA ACCURACY: HOW GOOD ARE OUR USUAL INDICATORS?

Peter Mariolis, Ph.D.[1]

ABSTRACT

In 2000, the Behavioral Risk Factor Surveillance System (BRFSS) conducted monthly telephone surveys in 50 states, the District of Columbia, and Puerto Rico. Each was responsible for collecting its own survey data. In Maine, data collection was split between the state health department and ORC Macro, a commercial market research firm. Examination of survey outcome rates, selection biases, and missing values for income suggest that the Maine health department data are more accurate. Only 4 of 18 behavioral health risk factors, however, are statistically different by data collector, and for these 4 factors, the data collected by ORC Macro seem more accurate.

KEY WORDS: Telephone surveys; Nonresponse bias; Data quality.

1. INTRODUCTION

In 2000, the Behavioral Risk Factor Surveillance System (BRFSS) conducted monthly telephone surveys on health behaviors and practices in 50 states, the District of Columbia, and Puerto Rico.[2] Although there were common questions, guidelines, and standards, each state was responsible for collecting its own data. Fifty-two states employed 36 different data collection organizations: a unit of the state health department collected the data in 17 states and 1 of 19 different university or commercial contractors collected the data in 35 states.

In Maine in 2000, because of a unique situation, some of the data were collected by a unit of the state health department and the rest by ORC Macro, a commercial market research firm. Although the Maine pairing represents only 1 out of 630 ($_{36}C_2$) possible pairings, it was the first time in the 17-year history of the BRFSS that two data collection entities simultaneously collected data from the same geographical area. This situation enables a comparison of the accuracy of substantive outcomes in relation to commonly used measures of data accuracy.

At least three types of measures are readily available to help assess the accuracy of BRFSS data and are used here: survey outcome rates based on disposition codes, selection bias measures of demographic characteristics, and item missing values. The outcome rates are the CASRO (Council of American Survey Research Organizations) rate, the household detection rate, and the household completion rate. The selection bias measures are for gender and age. The distribution of annual household income, a sensitive variable that usually has relatively large percentages of missing values, is used to measure item missing values. These three sets of outcomes are used to assess overall data accuracy. A fourth set of outcomes examined are prevalence of 13 key behavioral health risk factor variables. These outcomes are used to assess the relationship of the data accuracy measures with substantively important variables.

[1] Peter Mariolis, Centers for Disease Control and Prevention (MS K66), 4770 Buford Hwy, NE, Atlanta, Georgia, USA 30341-3717
[2] Hereafter, "states" refers to the 50 states, the District of Columbia, and Puerto Rico.

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, Guam, Puerto Rico, and the Virgin Islands. It is a (usually) monthly telephone survey primarily used 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 health status, preventive health practices, physical activity, and smoking. CDC coordinates the development of an annual set of core questions which are asked by every state and standardized sets of questions on specific topics (modules) which the states can optionally 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 projects. The sampling frame for the BRFSS consists of all telephone numbers of NXX types 00, 50, 51, 52, and 54, including zero-block numbers. [3] The sample design must be a variant of a random-digit-dialed design. BRFSS guidelines prescribe up to 15 callbacks for unresolved numbers distributed over weekday, weeknight, and weekend calling occasions. One eligible adult is randomly chosen from each household; no proxy interviews are allowed. The states are responsible for 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. The CDC then makes the aggregate data set available to the public. Additional information on the BRFSS, including the specific questions whose results are reported in this paper, may be found at http://www.cdc.gov/nccdphp/brfss.

For data year 2000, the Maine health department was contacted by an outside organization which offered additional funding to increase sample sizes in 10 of Maine's 16 counties. The health department wanted to be responsive but did not have the capacity to significantly increase its number of completed interviews per month. Eventually, the health department continued to conduct its accustomed number of statewide interviews and contracted with ORC Macro to field the oversample of the 10 counties. In the sample design, telephone numbers assigned to each of the 10 counties constitute a stratum and the other 6 counties an 11th stratum.

The data for this study are the telephone numbers assigned to the 10 oversampled counties. Completed interviews for the records assigned to 1 of the 10 strata that are located in one of the other 6 counties were re-coded as "out of scope." The sample records assigned to the Maine state health department and to ORC Macro were identified, and each set of records is analyzed separately. The completed interviews are separately poststratified by gender and six age categories to the total population of the 10 counties.

The formulas for the three survey outcome rates used in this study, in terms of final disposition codes, are presented in Figure 1. The CASRO rate attempts to measure responsiveness among eligible households in the sample able to respond, whether identified as such or not, the household detection rate measures the amount of household contact in the data, and the household completion rate measures responsiveness among identified households.. Survey outcome rates are usually presented unweighted for the sample as a whole but this paper deviates twice from that practice. First, the outcome rates are weighted for the probability of selection of the telephone numbers. Strata differ in their outcome rates, and the mix of strata differs between the health department and the ORC Macro samples. Weighting assures that none of the differences in outcome rates between the health department and ORC Macro outcomes are due to a different mix of strata in their samples. In practice, this adjustment makes very little difference. Second, outcome rates are examined separately for listed, not-listed one-plus block, and zero block telephone

[3] Hundred blocks are sets of 100 telephone numbers with the same area code, prefix, and first two digits of the suffix. Commercial sample providers use telephone directory databases to identify listed household telephone numbers. A one-plus block is a hundred block that contains one or more listed household telephone numbers. A zero block is a hundred block that contains no listed household telephone numbers.

numbers as well as the sample as a whole.[4] All three rates can vary significantly by density status, especially the household detection rate.

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

CASRO Rate

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

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		

Selection biases for gender and age are determined by the difference between the percentage in a gender or age category and the population percentage. Differences in the distributions of gender and age between the Maine health department and ORC Macro data are also examined.

Annual household income is a very sensitive question and, as such, is more likely to have item missing values than other variables are. This paper compares the distributions of income categories and, in more detail, the extent of records with missing values. The two kinds of missing values are Don't know or Not sure and Refusal; these are examined both separately and together. The comparisons are performed by gender because responses to the income question, especially missing values, can vary considerably by gender.

The final set of comparisons is on 13 selected behavioral health risk factors. These 13 variables appear on a report which CDC sends to the states and, as such, have been selected for their public health importance. The percentages of respondents in selected categories in the Maine state health department and ORC Macro data are compared with the distribution of percentages among all states except Maine in the 2000 BRFSS survey.

The statistical significance of differences between Maine state health department and ORC Macro data are determined by a Pearson-like chi-square test for independence using SUDAAN 8.0. Statistical significance was set a $p < .05$.

[4] One-plus blocks are much more likely to contain household telephone numbers than are zero blocks. Most telephone sampling frames exclude zero block telephone numbers. Among one-plus block telephone numbers, listed numbers are more likely to contain household telephone numbers than are not-listed numbers. "Density status" denotes the categories of listed, not-listed one-plus block, and zero block telephone numbers.

3. RESULTS

The disposition code frequencies by density status and data collector are shown in Table 1. The Maine health department released 11,801 sample records, and ORC Macro released 26,967.

Table 1. Frequency (No.) of Disposition Codes by Density Status and Data Collector

	Maine Health Department				ORC Macro			
Disposition Code	Listed	Not Listed, One-Plus Block	Zero Block	Total	Listed	Not Listed, One-Plus Block	Zero Block	Total
01	1,187	334	2	1,523	2,059	625	5	2689
02	375	103	3	481	1,748	510	6	2264
03	449	2,481	4,309	7,239	920	5,416	9,903	16,239
04	334	317	31	682	959	932	158	2049
05	158	796	241	1,195	421	1,917	590	2928
06	90	26	2	118	38	12	0	50
07	261	105	0	366	228	80	0	308
08	13	2	0	15	10	5	0	15
09	15	3	0	18	44	19	0	63
10	12	33	25	70	4	28	220	252
11	58	19	0	77	80	16	0	96
12	13	4	0	17	11	3	0	14
Total	2,965	4,223	4,613	11,801	6,522	9,563	10,882	26,967

The distribution of completed interviews by age and gender is presented in Table 2. The Maine health department completed 1,523 interviews, and ORC Macro completed 2,689.

Table 2. Distribution (No.) of Completed Interviews and Missing Values, by Age, Gender and Data Collector

Completed Interview, by	Maine Health Department			ORC Macro		
Age Group (Years)	Male	Female	Total	Male	Female	Total
18-24	37	56	93	90	100	190
25-34	90	120	210	171	247	418
35-44	147	194	341	241	322	563
45-54	140	178	318	229	353	582
55-64	104	123	227	156	210	366
65+	123	207	330	195	361	556
Don't Know or Not Sure	0	1	1	0	0	0
Refused	1	2	3	1	13	13
Totals	642	881	1,523	1,083	1,606	2,689

The Maine health department has statistically higher CASRO and household completion rates than ORC Macro did, except for zero block numbers, which contain very few households (Table 5). The largest difference is 14.5 percentage points in the overall, weighted CASRO rate; the smallest difference is 7.5 percentage points in the weighted household completion rate for not-listed, one-plus block numbers. The differences in household detection rate by data collector are smaller and not statistically significant except for a 3.2 percentage point difference for the listed, weighted rate and a 1.7 percentage point difference for the overall, unweighted rate.

Table 3. Survey Outcome Rates by Density Status and Data Collector			
Survey Outcome Rates	Maine Health Department (%)	ORC Macro (%)	P Value
CASRO Rate			
Overall, unweighted	59.3	46.1	
Overall, weighted	60.0	45.5	
Listed, weighted	57.3	42.1	
Not-listed, one-plus block, weighted	53.8	43.6	
Zero block, weighted	32.5	35.6	
Household Detection Rate			
Overall, unweighted	22.0	20.3	.00
Overall, weighted	11.5	10.8	.06
Listed, weighted	68.3	65.1	.01
Not-listed, one-plus block, weighted	16.0	15.0	.29
Zero block, weighted	0.13	0.14	.85
Household Completion Rate			
Overall, unweighted	58.6	49.0	.00
Overall, weighted	57.3	47.4	.00
Listed, weighted	58.3	47.5	.00
Not-listed, one-plus block, weighted	55.2	47.7	.00
Zero block, weighted	19.9	37.3	.32

The percent female and the percent in each age category, weighted by the design weights, are not significantly different between the Maine health department and ORC Macro except for the age group 18-24 years (Table 4). ORC Macro has larger selection biases for females and the age group 65+ years, and the health department has larger selection biases for the other five age categories. The population percentages are outside the two standard error confidence intervals for percent female and percent age 18-24 years for both data collectors, which indicates that these are statistically significant biases. The population percentage for the age group 35-44 is also outside the confidence interval for the state health department.

Table 4. Selection Biases for Gender and Age and Distributions of Gender and Age, by Data Collector								
Category	Selection Bias		Design Weight			Confidence Intervals (±2SE)		Population Percentage
	MHD*	ORC Macro	MHD* (%)	ORC Macro (%)	P Value	MHD*	ORC Macro	
Female	4.25	5.16	56.20	57.12	.68	52.76-59.64[§]	54.38-59.86[§]	51.95
Age (Years)								
18-24	-5.68	-3.13	6.76	9.31	.05	4.94-8.58[§]	7.51-11.11[§]	12.44
25-34	-1.44	0.08	15.72	17.24	.36	13.14-18.30	15.14-19.34	17.16
35-44	3.19	1.63	24.26	22.70	.41	21.26-27.26[§]	20.42-24.98	21.07
45-54	2.16	2.02	21.01	20.87	.93	18.05-23.97	18.61-23.13	18.85
55-64	1.49	0.69	13.65	12.85	.62	11.49-15.81	11.03-14.67	12.16
65+	0.36	-1.28	18.68	17.04	.29	16.16-21.20	15.14-18.94	18.32

*Maine Health Department [§] Interval excludes population percentage

The two statistically significant differences in income involve missing values: the combined missing category for men and the refusals for women (Table 5).

Comparisons of Maine health department and ORC Macro data on 13 key health risk factors are presented in Table 6. Four of the comparisons are statistically significant: women aged 40 years or older who report that they have never had a mammogram or breast examination (6.53 percentage point difference), women aged 50 years or older who report that they have not had a mammogram or a breast examination in the past two years (15.21 percentage point difference), respondents who report that they are currently smoking (4.07

percentage point difference), and respondents who report a medium or high chance of getting HIV/AIDS (2.84 percentage point difference).

Table 5. Annual Household Income in U.S. Dollars by Gender and Data Collector

Category	Maine Health Department (%)	ORC Macro (%)	P Value
Male			
<10K	2.09	3.61	.16
10 to <15K	2.19	3.46	.12
15 to <20K	6.80	4.58	.14
20K to <25K	10.93	9.44	.50
25K to <35K	19.67	17.07	.33
35K to <50K	18.13	18.53	.88
50K to <75K	17.25	20.61	.22
7.00 93K1.867			
KPno w(orN)46(o)-1.3 tSCere			
20Kt 5K 423			
25K to <35K	58.53	4(814)-3952.9(.33)T0 -1.1566 TD(35K)4.2(t)5	
50K to <75K	23493	73.26 991	
7.00 63K 7o991			
KPno w(orN)46(o)-1.3 tSCere			
e.8.1fufs8263(e.8.1dK)-17-10.856	10291		

ORC Macro value of 24.04% current smokers would fall at the 57th percentile. Finally, the Maine health department value of 3.58% of respondents with a medium or high change of getting HIV/AIDS would be the minimum value, and the ORC Macro value of 6.42% would fall at the 73rd percentile. An examination of quantiles excluding the nine other states for which ORC Macro collected data in 2000 gives virtually the same results.

Table 7. Quantiles of Health Risk Factors Significantly Different by Data Collector, All States Except Maine, BRFSS, 2000

Never Had Mammogram and Breast Exam (Women aged 40+ years)

Quantile	Estimate
100% Max	24.3661
99%	24.3661
95%	22.8793
90%	21.8680
75% Q3	19.6037
50% Median	17.6754
25% Q1	15.5713
10%	13.6158
5%	13.1187
1%	11.2633
0% Min	11.2633

No Mammogram and Breast Exam in Past 2 Years (Women aged 50+ years)

Quantile	Estimate
100% Max	37.9436
99%	37.9436
95%	35.8481
90%	35.4864
75% Q3	32.6483
50% Median	28.9652
25% Q1	26.3662
10%	21.3631
5%	19.0107
1%	18.2203
0% Min	18.2203

Current Smoker

Quantile	Estimate
100% Max	30.4962
99%	30.4962
95%	27.1851
90%	26.1167
75% Q3	24.9085
50% Median	23.1951
25% Q1	20.7464
10%	19.6643
5%	17.2385
1%	12.8512
0% Min	12.8512

Chances of Getting HIV/AIDS Medium or High (Aged 18-64 years)

Quantile	Estimate
100% Max	11.74100
99%	11.74100
95%	9.23939
90%	7.89847
75% Q3	6.62581
50% Median	6.07783
25% Q1	5.16553
10%	4.82102
5%	4.33777
1%	4.17341
0% Min	4.17341

4. DISCUSSION

The disjunction between the measures of data accuracy, especially the survey outcome rates, and the apparent accuracy of the substantive results is striking. The CASRO and household completion rates clearly indicate more non-response in the ORC Macro data than in the Maine health department data. The household detection rate is also more favorable for the health department. The analysis of selection biases and missing values give a more mixed picture but on the whole suggest that the health department's data are of greater accuracy. The behavioral risk factor data overturn this assessment.

Keeter et al. (2000) and Curtin et al. (2000) recently suggested that large differences in response rates in telephone surveys may not have much of an effect on most substantive outcomes. The results in this paper support that position. Of 10 independent comparisons (grouping the mammogram variables and the physical activity variables, since they are based on some of the same variables), 7 show no statistically significant difference. Further, although the difference in percentage of current smokers is statistically significant, the difference of 4 percentage points is about equal to the size of the interquartile range between states. Thus, its substantive significance is at best marginal.

The major exceptions to the pattern of no differences in outcome despite large differences in response rates are the mammogram and breast exam variables. Here are very large and highly statistically significant differences in spite of substantial subsetting and thus decreased statistical power. The difference in the percentage of respondents reporting a medium or high chance of contacting HIV/AIDS is also large.

The comparison of the mammogram, smoking, and HIV/AIDS prevalences with the prevalences in all other states shows that the Maine state health department results are much more extreme than the ORC Macro results. In three of four cases, the health department results would be the smallest or second smallest value. By contrast, three of the four ORC Macro prevalences would fall between the 53rd and the 57th percentiles. This comparison cannot prove which estimate is more accurate. Nevertheless, it seems reasonable to assume that a pattern of more extreme values would indicate a greater likelihood of bias than a pattern of less extreme values would. On this basis, the ORC Macro results for all four statistically significant differences seem—contrary to the indications given by the data accuracy measures—to be more accurate than the Maine health department results are.

The results in this paper support two extensions of the position that low response rates may not be strongly related to substantive outcomes. First, the cited studies compared outcomes from a single data collector under high and low response rate conditions. This enabled them to assume that, when there was a difference in outcomes, higher response rates meant more accurate data. This study compared outcomes from two different data collectors and found that the data collector with the higher response rate seems to have less accurate data. So, not only can large differences in response rates not make a difference in substantive outcomes, larger response rates need not indicate more accurate data when there are differences in substantive outcomes. Second, these studies dealt exclusively with response rates. This paper has shown that other types of outcome rates and other types of indicators of data accuracy may not be strongly, or even positively, related to data accuracy. Further research, of course, is needed to see how generalizable the findings in this study are.

The results in this paper do not mean that commonly used indicators of data accuracy are irrelevant to data accuracy. Higher response rates, for example, do indicate less of a potential for bias from non-response. Furthermore, both of the two previous studies found some substantive differences between conditions of high and low response rates. Other things equal, higher response rates are better than lower response rates. But one lesson to be learned from these studies, including this one, is that "other things" are rarely equal and that any single indicator of data accuracy is only one of many different imperfect indicators--that is--there are many different specific causes of non-sampling, including nonresponse, error in surveys. This lesson, in turn, implies that only a comprehensive set of indicators that are related to at least the most important plausible causes of non-sampling error in a survey can hope to provide a basis for a realistic assessment of non-sampling error in that survey. The field of survey methodology is a long way from being able to provide such a comprehensive set of indicators.

REFERENCES

Curtin, R., S. Presser, and E. Singer (2000), "The Effects of Response Rate Changes on the Index of Consumer Sentiment", *Public Opinion Quarterly*, 64, pp. 413-428.

Keeter, S., C. Miller, A. Kohut, R. M. Groves, and S. Presser (2000), "Consequences of Reducing Nonresponse in a National Telephone Survey", *Public Opinion Quarterly*, 64, pp. 125-148.