

BIAS FROM EXCLUDING HOUSEHOLDS WITHOUT TELEPHONES IN RANDOM DIGIT DIALING SURVEYS: FOLLOW UP

Mourad Touzani and John Hall

Mathematica Policy Research, Inc., Princeton, New Jersey 08543-2393

Key Words: Sample design, Community Tracking Survey, Bias Estimate

I. INTRODUCTION

The purpose of this paper is to investigate potential bias resulting from undercoverage of RDD sampling frames. Our method is to compare estimates of a variety of characteristics for households sampled by Random Digit Dialing (RDD) methods and those contacted in person as part of an area probability sample.

A large number of surveys are conducted using RDD samples, and are restricted to only households with telephones. There is usually some concern about the bias in survey estimates due to the exclusion of households without a telephone service and under representation of those with intermittent service. In some surveys, such as the Community Tracking Study (CTS), the concern was so great that it was decided not to rely solely on an RDD survey. This paper utilizes data from the first and third round of the Community Tracking Study; it follows up on a prior study using the first round of the CTS data (Hall et al, 2000). The CTS has an RDD component and an area probability sample for households that lack or have intermittent telephone service.

The approach for this paper is to compare estimates using the entire sample to those that obtained from households in the RDD component. Statistically significant differences indicate that the RDD only estimates may be biased. We also present separate estimates for the field component. We interpret differences using criteria that include: statistical significance, the size of the difference relative to the entire sample estimate, the size of the difference relative to the standard error, and the size of the difference relative to policy considerations. Further, we compare bias estimates in the third round to those in the first round of the survey. A bias that is substantially large relative to the estimate itself would be considered important. However, a small bias can still be important in terms of persons affected and (or) dollars spent.

II. COMMUNITY TRACKING STUDY

The household surveys of the CTS have been funded by the Robert Wood Johnson Foundation. Mathematica Policy Research, Inc. (MPR) was the primary contractor for survey designs, instrument development, sample designs and implementation,

interviewing, weighting, and variance estimation. The household survey instruments covered health insurance, use of health services, satisfaction with care, health status and demographic information.

Ninety percent of the CTS RDD component interviews were conducted in 60 sites, which were selected with probability proportional to size and comprise a national multi-stage sample. Forty-eight sites were selected from MSAs that in 1992 had an estimated population of at least 200,000 persons. Three sites were selected from smaller MSAs and 12 from non-MSA areas. Twelve sites (the high-intensity sites) were randomly selected from the 48 sites comprised of MSAs that in 1992 had a population of at least 200,000 persons. In these high intensity sites, there was a larger allocation of RDD interviews plus the area probability sample, referred to below as the field component.¹

In the first round, the CTS collected data from 26,277 households using random-digit-dialing methodology (the RDD component). In addition, 453 households (the field component) were contacted by a personal visit from a lister/screener and interviewed by having the respondent use a cellular telephone (carried by the lister/screener) to call a central interviewing facility. In the third round, the CTS collected data from 23,995 households using RDD methodology and from 595 households for the field component. All households in the field component reported an interruption in telephone service of at least 2 weeks in the last 12 months. For the first round, we will review comparison between data from 453 households in the field component and data from 21,115 households in the RDD component. For the third round, we will compare data from 595 households in the field component to data from 20,628 households in the RDD component.

III. ANALYSIS OF CTS DATA

The analysis was conducted with SUDAAN software using weights constructed to produce national estimates. The weights of both components

¹The data in this analysis is restricted to the strata (48 sites) that represent MSAs with 1992 population of 200,000 or more. The reason for this restriction is that the field component was conducted only in these strata.

took into account the presence of multiple telephones and the length in any reported interruption in telephone service. The analysis took account of the multistage nature of the CTS design. The variables in the analysis include measures of health care utilization (physician visits, receipt of a flux shot, hospital stay, having a usual source of care, and for women aged 45 to 60, having had a mammogram) cigarette smoking (currently smoking among those whoever smoked), and health insurance coverage status (whether uninsured or covered by Medicaid). Data are presented separately for adults and children. To determine statistical significance, we tested for differences between the field and RDD component using the Chi-Square statistic described in Shah, Barnwell and Bieler (1996, p. 6-18).

Tables 1-4 show results of our analysis of categorical variables estimates for adults. Table 1 presents differences found in the first round. Of the 8 variables, differences between the RDD and the whole sample were statistically significant for 7 (only the estimate of hospital stay was not significant). Data in table 2 shows that, for 6 of these, the absolute value of the estimated bias was at least half as large as the estimated standard error for the RDD sample. Of these 6 variables, two have an estimated bias of over a percentage point (smoking and health insurance coverage). The difference of 1.71 percentage points of those who still smoke (as a percent of those who have ever smoked) represents approximately 1.1 million adults. In the area of health insurance coverage, the RDD sample produced a lower estimate (by 1.54 percentage points) of those without health insurance.

Table 3 presents data estimated for adults from the third round of CTS data for the same 8 categorical variables as in round 1. Differences between the RDD and the whole sample were statistically significant for all 8 variables. Further, data in table 4 shows that the absolute value of the estimated bias was at least as large as the estimated standard error for the RDD sample. Of these 8 variables, three have an estimated bias of over a percentage point. (doctor visits within the 12 months preceding the survey, health insurance coverage and mammogram test). The difference of 1.82 percentage points (as a percent of those without health insurance) represents approximately 2.26 million adults. In the area of doctor visits, the RDD sample produced a higher estimate of those who have ever had doctor visits. The difference of 1.20 percentage points represents approximately 1.5 million adults.

Table 5-8 present findings regarding estimates for children. Table 5 shows results of our analysis for 6 categorical variables estimated for children from the first round data. Statistically significant differences were detected for 5 of the 6. In addition,

the absolute value of the estimated bias was at least half as large as the estimated standard error for the RDD sample for 4 out of the 5. The potential of bias is the largest in Medicaid coverage. Table 6 shows that the RDD sample alone would have underestimated the percent of children on Medicaid by approximately 2 percentage points, which would represent approximately 1.4 million children.

Table 7 presents data estimated for children in round 3 for the same categorical variables as in round 1. 2 differences were statistically significant at the 5% level (Medicaid coverage and uninsurance), while 2 other differences were significant at the 10% level (usual source of care and doctor visits within the 12 months preceding the survey). Table 8 shows that the absolute value of the estimated bias was at least half as large as the estimated standard error for the RDD sample for all 4 attributes. Finally, as in round 1, the potential bias is largest in Medicaid coverage. The RDD sample alone would have underestimated the percent of children on Medicaid by approximately 2.55 percentage points.

IV. CONCLUSIONS

In sum, it appears that omitting the field component would have resulted in some bias in sample estimates for nearly all characteristics examined for both round 1 and round 3 of CTS. The potential bias appears to be largest in measures of health insurance coverage. The size of the bias can be even more substantial in policy terms when translated to persons affected. For example, RDD sample alone would have underestimated children on Medicaid by 1.4 millions.

In general, it appears that measures that are more correlated with income, such as health insurance coverage, tend to have higher absolute and relative biases than those that are less related to income. The bias is large relative to the standard error for most of the characteristics including those that are not closely tied to income. Because RDD households with telephone service interruptions had higher weights than those with none, it is possible that the biases noted above could have been reduced to unimportant levels by giving even higher weights to such households in the RDD sample. The problem with this strategy is that it would have resulted in increased loss of precision due to differential weighting.

REFERENCES

Carlson, Barbara Lepidus, Richard Strouse, and John Hall. The Community Tracking Study: Measuring Changes In The U. S. Health Care System Over Time. 2002 Proceedings of the American Statistical Association, Survey Research Methods Section [CD-

ROM], Alexandria, VA: American Statistical Association.

Strouse, Richard, Barbara Lepidus Carlson, and John Hall. "Report on Survey Methods for the Community Tracking Study's 1998-1999 Round Two Household Survey." Mathematica Policy Research, 2001. Available at www.hschange.com as Technical Publication 34.

Strouse, Richard, John Hall, Frank Potter, Barbara Lepidus Carlson, Peter Cunningham, Joanne Pascale, and Rita Stapulonis. "Report on Survey Methods for the Community Tracking Study's 1996-1997 Round One Household Survey." Mathematica Policy Research, 1998. Available at www.hschange.com as Technical Publication 15.

Carlson, Barbara Lepidus, Richard Strouse, and John Hall. The Community Tracking Study: Measuring Changes In The U. S. Health Care System Over Time. 2002 Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association.

Strouse, Richard, Barbara Lepidus Carlson, and John Hall. "Report on Survey Methods for the Community Tracking Study's 1998-1999 Round Two Household Survey." Mathematica Policy Research, 2001. Available at www.hschange.com as Technical Publication 34.

Strouse, Richard, John Hall, Frank Potter, Barbara Lepidus Carlson, Peter Cunningham, Joanne Pascale, and Rita Stapulonis. "Report on Survey Methods for the Community Tracking Study's 1996-1997 Round One Household Survey." Mathematica Policy Research, 1998. Available at www.hschange.com as Technical Publication 15.

Metcalf, Charles E., Peter Kemper, Linda T. Kohn, Jeremy D. Pickreign. Site Definition and Sample Design for the Community Tracking Study. Technical Publication No. 01 Center For Studying Health Systems Change. October 1996

Hall, John, Genevieve Kenney, Gary Shapiro, and Ismael Flores-Cervantes Bias from Excluding Households without Telephones in Random Digit Dialing Surveys: Results of Two Surveys . 1999 Proceedings of the American Statistical Association, Section on Survey Research Methods. Alexandria, VA: American Statistical Association, 2000.

TABLE 1
COMPARISON OF RDD & FIELD ESTIMATES FOR ADULTS (ROUND 1)

Variable	Total (%)	RDD (%)	Field (%)
In the Last 12 months			
Any doctor visits***	75.36	76.21	54.60
Had flu shot***	25.21	25.71	12.67
Any overnight hospital stays**	3.03	3.00	3.97
Currently			
Has usual source of care***	85.64	86.13	72.93
Smoke now***	52.09	50.38	85.42
Covered by Medicaid***	4.21	3.58	17.9
Uninsured***	17.05	15.51	50.46
Ever had mammogram***	84.04	84.66	48.45

*p value for Chi-square statistic < .1

** p value for Chi-square statistic < .05

*** p value for Chi-square statistic < .01

TABLE 2
BIAS RELATIVE TO ESTIMATE & STANDARD ERROR FOR ADULTS (ROUND 1)

Variable	Bias (%)	Bias/Estimate (%)	Bias/SE (%)
In the Last 12 Months			
Any doctor visits ¹	0.85	1.11	125.00
Had flu shot ¹	0.50	1.94	61.73
Any overnight hospital stays ¹	-0.03	1.00	30.00
Currently			
Has usual source of care	0.49	0.57	76.56
Smoke now ³	-1.71	3.39	356.25
Covered by Medicaid ⁴	-0.63	17.60	350.00
Uninsured ⁴	-1.54	9.93	151.00
Ever had mammogram ²	0.62	0.73	144.19

¹In 12 months prior to interview

²Women age45+

³Among adults who reported ever smoking

⁴Adults 18-64

TABLE 3
COMPARISON OF RDD & FIELD ESTIMATES FOR ADULTS (ROUND 3)

Variable	Total (%)	RDD (%)	Field (%)
In the Last 12 Months			
Any doctor visits***	73.82	75.02	54.38
Had flu shot***	19.54	20.00	12.12
Any overnight hospital stays**	2.00	1.85	4.40
Any surgical procedures	13.85	13.98	11.76
Currently			
Has usual source of care***	83.30	84.01	71.92
Smoke now***	47.58	47.00	56.42
Covered by Medicaid***	3.79	3.01	16.34
Uninsured***	16.57	14.75	45.86
Ever had mammogram***	84.18	85.33	58.63

*p value for Chi-square statistic < .1
 ** p value for Chi-square statistic < .05
 *** p value for Chi-square statistic < .01

TABLE 4
BIAS RELATIVE TO ESTIMATE & STANDARD ERROR FOR ADULTS (ROUND 3)

Variable	Bias	Bias/Estimate	Bias/SE
	%	%	%
A. Adult Characteristics			
Any doctor visits ¹	1.20	1.60	352.94
Had flu shot ¹	0.46	2.30	148.39
Any overnight hospital stays ¹	-0.15	8.11	166.67
Ever had mammogram ²	1.15	1.35	239.58
Has Usual Source of Care	0.71	0.85	221.88
Smoke now ³	-0.58	1.23	148.72
Covered by Medicaid ⁴	-0.78	25.91	520.00
Uninsured ⁴	-1.82	12.34	551.52

¹In 12 months prior to interview
²Women age45+
³Among adults who reported ever smoking
⁴Adults 18-64

TABLE 5
COMPARISON OF RDD & FIELD ESTIMATES FOR CHILDREN (ROUND 1)

Variable	Total (%)	RDD (%)	Field (%)
In the Last 12 Months			
Any doctor visits**	82.12	82.66	72.8
Any overnight hospital stays**	0.77	0.74	1.30
Any surgical procedures**	6.71	6.98	2.15
Currently			
Has usual source of care	94.44	94.71	89.87
Covered by Medicaid**	14.65	12.65	49.24
Uninsured**	11.54	11.03	20.24

**p value for Chi-square statistic < .05

TABLE 6
BIAS RELATIVE TO ESTIMATE & STANDARD ERROR FOR CHILDREN (ROUND 1)

Variable	Bias (%)	Bias/Estimate (%)	Bias/SE (%)
In the Last 12 Months			
Any doctor visits ¹	0.54	0.65	70.13
Any overnight hospital stays ¹	-0.03	4.05	30.00
Any surgical procedures ¹	0.27	3.87	79.41
Currently			
Has usual source of care	0.27	0.28	62.79
Covered by Medicaid ²	-2.00	15.81	298.50
Uninsured ²	-0.51	4.62	71.83

¹In 12 months prior to interview

²Children under 18

TABLE 7
COMPARISON OF RDD & FIELD ESTIMATES FOR CHILDREN (ROUND 3)

Variable	Total (%)	RDD (%)	Field (%)
In the Last 12 Months			
Any doctor visits*	82.66	83.18	75.13
Any overnight hospital stays	0.77	0.77	0.70
Any surgical procedures	7.04	7.05	6.86
Currently			
Has usual source of care*	92.11	92.48	86.74
Covered by Medicaid***	13.82	11.27	50.33
Uninsured***	9.55	8.86	19.40

*p value for Chi-square statistic < .1
 **p value for Chi square statistic < .05
 ***p value for Chi square statistic < .01

TABLE 8
BIAS RELATIVE TO ESTIMATE & STANDARD ERROR FOR CHILDREN (ROUND 3)

Variable	Bias (%)	Bias/Estimate (%)	Bias/SE (%)
In the Last 12 Months			
Any doctor visits ¹	0.52	0.63	86.67
Any overnight hospital stays ¹	0.00	0.00	0.00
Any surgical procedures ¹	0.01	0.14	2.70
Currently			
Has usual source of care	0.37	0.40	86.05
Covered by Medicaid ²	-2.55	22.63	490.38
Uninsured ²	-0.69	7.79	146.81

¹In 12 months prior to interview
²Children under 18