

## SOME EFFECTS OF UNDERCOVERAGE IN A TELEPHONE SURVEY OF TEENAGERS

Steven L. Botman and Karen Allen  
NCHS, 6525 Belcrest Rd., Hyattsville, MD 20782

Survey designers sometimes use a sampling frame that incompletely covers a target population. This paper examines preliminary coverage effects that result from using a large, face-to-face nationally representative population sample as the sampling frame for a smaller telephone survey. In this case, the smaller survey is a national telephone survey of persons aged 12-18 years (teens). The final teen sampling frame encompassed teens in households with an available, valid telephone number. The telephone numbers were usually identified from the larger survey.

In the larger survey, some households with a telephone refused to report their telephone number. Some telephone numbers in the sample were also unlisted. Some household telephone numbers not available in the frame, however, could be obtained elsewhere. Obtaining telephone numbers for these sample teens required nontrivial survey efforts, and more importantly these efforts were not always successful. The problem is both undercoverage and noncoverage.

In this paper, such efforts are assumed to have identified telephone numbers for a negligible number of households. With later data we can assess this assumption.

This assumption, however, is not unreasonable. That is, we likely found some, but not many, such telephone numbers for individuals who did not have a valid available household number in the frame. There is anecdotal information that households with unlisted telephone numbers disproportionately refuse to provide their telephone number in the larger survey. If in this case a respondent to the larger survey refused to provide their household number, they would be also unlikely to provide the larger survey with a telephone number for a household contact. In this case, we could not obtain a household telephone number for a sample teen.

With this working assumption on frame coverage, an estimated 11.1 percent of teenaged population is not covered by the frame. By this we mean if we sum the sampling weights in the frame for teens in households with an available telephone number, we would estimate 88.9 of the teen population. Also most (an estimated 7.3 percent of the teens) resided in households without a telephone.

This assumption has several implications. We decreased the population coverage rates from 92.7 (= 100 - 7.3) percent, the estimated proportion of the teenage population residing in telephone households, to 88.9

(= 100 - 11.1) percent. We also likely reduced the differences in proportion of the covered and noncovered population with a characteristic of interest.

The noncovered population encompasses the teen household population without a telephone and the teen household population with a telephone but for whom we have no telephone number. The former group is predominantly low income. The latter group is not. We were uncertain for teens of the effects on data of such incomplete frame coverage.

Incomplete frame coverage may introduce error in survey data (see, for example, Groves, 1989). Incomplete coverage of a target population, however, is only one potential source of survey error. Nonresponse bias, for example, may also introduce error in survey data. The coverage analysis in this paper also provides clues on designing post-survey weighting adjustments for such a survey. The final estimator will include weighting adjustments that attempt to reduce bias due to incomplete population coverage of the frame.

In this paper, to characterize teenagers by telephone status, we use information from the larger survey. For this paper, we take the Teenage Attitude and Practices Survey (TAPS) as an example of a survey whose sampling frame partially covers a target population. The larger survey is the National Health Interview Survey (NHIS). Large differences in the health status of teens were not identified in the covered and noncovered populations.

One limitation in this analysis is that for teens, these two surveys use different respondent rules and interview modes. In the NHIS, survey data for most individuals under age 18 are obtained by proxy from a knowledgeable, adult family member in the household. In TAPS, survey data are obtained using self response. In the NHIS, data are obtained by a face-to-face, household interview. In TAPS, data are usually obtained by a telephone interview.

Section 1 introduces the paper, including a description of the TAPS data collection strategy. Section 2 examines frame coverage. Section 3 provides data on the teens by telephone status based on the larger survey. Section 4 summarizes the paper.

### SECTION 1. INTRODUCTION

Considerable research has been conducted on coverage effects in telephone surveys (see, for example, Thornberry and Massey's paper on "Trends in U.S. Telephone Coverage across Time

and Subgroups" in Groves (1988)). Thornberry and Massey estimate that for 1985-1986, 7.2 percent of persons in the population lived in households without a telephone number. Telephone coverage of the population is not random. The telephone coverage of low income households is lower than that for other households.

For the TAPS, a preliminary estimate is that 11.1 percent of the population of teens is dropped from the frame. An estimated 6.8 percent of teens live in households reporting no telephone, an estimated 3.5 percent of teens live in households with a telephone but would refuse to report their telephone number, and an estimated 0.8 percent of teens live in households with unknown telephone status.

In two ways, this differs from the usual situation with telephone surveys. Most telephone surveys are based on random digit dialing (RDD) samples. In RDD surveys, all telephone households are in the frame.

In this survey, not all teens in households with telephones are in the frame. In this example, we effectively drop from the frame households either refusing to provide a telephone number or claiming no telephone number. Also, in most telephone surveys, the frame contains limited information. In this case, we had basic health and demographic data on teens in the frame, regardless of telephone status.

The National Center for Health Statistics (NCHS) designed the TAPS to produce information on smoking-related characteristics of teens. NCHS selected the TAPS sample from teenagers identified in NHIS. In fact, the TAPS sample is comprised of all teens identified in the NHIS from the third quarter of 1988 to the second quarter of 1989.

Under a contractual relationship with NCHS, the Bureau of the Census was the TAPS data collection agent data. Computer assisted telephone interviewing (CATI) was used for the TAPS. The CATI system was operated at a single location. Such a data collection strategy permits data collection efforts to be closely monitored.

An attempt was made to collect some data on sample teens dropped from the frame. By this we mean teens who would have been sampled for the TAPS CATI interview except that we had no available, valid telephone number. For these teens, NCHS mailed an abbreviated survey questionnaire. In general, obtaining high response rates in a households mail survey is difficult.

Obtaining high response rates in a mail survey of teens is even more difficult. It likely that teens see little value in participating in any survey. This TAPS mail sample was targeted to households without a telephone or households that refused to provide a phone number to the NHIS. With

lower response rates in the mail survey there is less protection against response bias. Preliminary results indicate a TAPS CATI response rate of 80 percent.

## SECTION 2. FRAME COVERAGE

As noted, an estimated 11.1 percent of the teen population live in households without an available telephone number. This is, of course, larger than that for the overall population in households without a telephone. Thus, it is appropriate to examine the coverage of the frame.

Table 1 shows the coverage of the teenage population by several socio-demographic characteristics. By region, the lowest coverage is in the South. By urban/rural status, the lowest coverage is in rural areas. By Metro Status, coverage is lowest in the central city areas and in non metropolitan statistical areas (MSA) and nonfarm areas. The differences are not large. This provides evidence that for many domains disproportionate coverage is not problematic.

Table 2 shows the coverage of the teenage population by household income. As expected, the proportion of the teen population covered increases with household income. Less than 80 percent of the teens in households whose household income is \$15,000 or less are covered. These households contain about 7 percent of the teens.

Table 3 shows the frame coverage of the teenage population for age, income, and race domains. This table shows that even controlling for household income, coverage is lower for Black teens than for other teens. While such analysis provides a clue on coverage gaps, in such analyses one should consider both family size (i.e., number of individuals in the family) and income.

Over half of the teens in households without a telephone live in the South. Geographic region is, thus, seen as important variable in coverage analysis.

Table 4 shows telephone availability coverage rates for domains defined by region and household poverty status. Based on its family size and income, we classify households as above or below the poverty line. The poverty line criterion is annually revised.

While 93 percent of the teens in households above the poverty line were estimated to have a reported telephone number, only 74 percent of the teens in households below the poverty line were estimated not to have a reported telephone. In the South and the East, the frame contained less than 65 percent of the teens in households below the poverty line. By region, controlling for poverty level, Black teens were slightly less likely to be in the frame.

Some caution is appropriate when using these figures, since some statistics are based on small sample sizes. The

patterns, however, are clear. For many domains, however, disproportionate coverage is not a problem.

### SECTION 3. SURVEY DATA BY TELEPHONE STATUS

No information on smoking for teens is available by telephone status. In the Cancer supplement to the 1987 NHIS, however, information on smoking by telephone status is available for the population 18 years of age or older. Self response was used in the Cancer supplement. For individuals exactly 18 years of age, according to estimates based on the Cancer supplement, individuals in households with an available telephone were slightly less likely to smoke than other 18 year old individuals.

These differences are not generalizable to all teens and the estimate for teens in nontelephone households is subject to a large sample error. The Cancer Supplement sample included only 84 persons aged 18 years who lived in a household without an available telephone number.

We needed to identify other cognate variables on teens health status, as reported by a parent or guardian, which would be available in the core 1988 NHIS. Doctor visits, hospital visits, and conditions did not seem to make sense as a proxy for smoking. We thought health status might. The response categories for health status are excellent, very good, good, fair, and poor.

Table 5 and table 6 show the proportion of teens reporting excellent or very good health status by telephone status. By region and poverty level, the proportion of teens reporting excellent/very good health status is substantially lower among teens residing in households without an available telephone number. The sample sizes are small for many of these domains.

While not definitive, this provides a clue that we may be slightly undercovering teenage smokers. This suggests analysts may need to be careful when analyzing data from this survey for teens from low income households.

This analysis also suggests that we introduce a TAPS post-survey weighting adjustment for incomplete frame coverage. One approach would be to use region, race, and possibly poverty status. This adjustment would attempt to compensate for the nonrandom nature of frame coverage. This is similar to the approach described by Massey and Botman in "Weighting Adjustments in RDD Surveys" in Groves 1988.

### SECTION 3. SUMMARY

This analysis identifies domains of study disproportionately covered by a frame of teens limited to households with available telephone numbers. Also by looking at health status, we see that a smaller proportion of the teens not covered has excellent/very good health status than the teen population covered.

This provides some indirect evidence that this survey approach slightly under represents teens with lower health status, which may include proportionally more smokers. For most domains, we did not identify disproportionate coverage rates.

#### Bibliography

James T. Massey and Steven L. Botman, Weighting Adjustments in Random Digit Dialed Surveys, in Telephone Survey Methodology, Edited by Robert M. Groves, Paul Biemer, Lars E. Lyberg, James T. Massey, William L. Nichols II, and Joseph Waksberg. Wiley, 1988.

Robert M. Groves, Survey Errors and Survey Costs. Wiley, 1989.

Table 1. Coverage of the Population Aged 12–18 with Available Telephone Number by Assorted Socio–Demographic Characteristics

| Socio–Demographic Characteristic | Region |      |       |      | U.S. |
|----------------------------------|--------|------|-------|------|------|
|                                  | East   | NC   | South | West |      |
| All                              | 0.88   | 0.93 | 0.86  | 0.89 | 0.89 |
| <u>Urban/Rural</u>               |        |      |       |      |      |
| <250,000                         | 0.84   | 0.91 | 0.89  | 0.89 | 0.89 |
| 250,000–999,999                  | 0.89   | 0.95 | 0.88  | 0.91 | 0.90 |
| 100,000–249,000                  | 0.96   | 0.95 | 0.85  | 0.84 | 0.89 |
| <100,000                         | 0.98   | 0.95 | 0.77  | 0.95 | 0.89 |
| Other urban                      | 0.98   | 0.94 | 0.82  | 0.93 | 0.89 |
| Rural                            | 0.94   | 0.92 | 0.84  | 0.82 | 0.87 |
| <u>Metro Status</u>              |        |      |       |      |      |
| MSA center city                  | 0.80   | 0.91 | 0.86  | 0.88 | 0.86 |
| MSA not center city              | 0.91   | 0.94 | 0.89  | 0.90 | 0.91 |
| Non MSA–non farm                 | 0.95   | 0.92 | 0.83  | 0.88 | 0.87 |
| Non msa–farm                     | 1.00   | 0.96 | 0.96  | 0.73 | 0.95 |
| <u>Family size</u>               |        |      |       |      |      |
| 1                                | 0.84   | 0.94 | 0.92  | 0.86 | 0.90 |
| 2                                | 0.80   | 0.87 | 0.84  | 0.85 | 0.84 |
| 3                                | 0.89   | 0.91 | 0.87  | 0.89 | 0.89 |
| 4                                | 0.91   | 0.95 | 0.89  | 0.93 | 0.92 |
| 5                                | 0.87   | 0.94 | 0.88  | 0.91 | 0.90 |
| 6                                | 0.90   | 0.92 | 0.82  | 0.83 | 0.87 |
| 7                                | 0.71   | 0.96 | 0.77  | 0.91 | 0.84 |
| 8                                | 0.97   | 0.92 | 0.78  | 0.82 | 0.86 |
| 9+                               | 0.74   | 0.82 | 0.64  | 0.76 | 0.71 |

Table 4. Coverage of the Population Aged 12–18 with Available Telephone Number by Region, Race and Poverty Status.

| Race and Poverty Status | Region |      |      |       |      |
|-------------------------|--------|------|------|-------|------|
|                         | U.S.   | East | NC   | South | West |
| Population Coverage     |        |      |      |       |      |
| U.S.                    | 0.89   | 0.88 | 0.93 | 0.86  | 0.89 |
| Above poverty line      | 0.93   | 0.91 | 0.96 | 0.92  | 0.93 |
| Below poverty line      | 0.74   | 0.71 | 0.81 | 0.70  | 0.78 |
| UN                      | 0.76   | 0.75 | 0.81 | 0.75  | 0.71 |
| Black                   |        |      |      |       |      |
| Above poverty line      | 0.90   | 0.90 | 0.93 | 0.90  | 0.85 |
| Below poverty line      | 0.71   | 0.63 | 0.78 | 0.69  | 0.73 |
| UN                      | 0.67   | 0.43 | 0.90 | 0.68  | 0.61 |
| NonBlack                |        |      |      |       |      |
| Above poverty line      | 0.93   | 0.91 | 0.96 | 0.92  | 0.93 |
| Below poverty line      | 0.76   | 0.76 | 0.82 | 0.70  | 0.78 |
| UN                      | 0.79   | 0.84 | 0.80 | 0.80  | 0.72 |

Table 2. Telephone Status of the Population 12–18 By Age and Income

| Phone Status                     | All Income | Household income in Thousands |        |         |         |         |        | UN    |
|----------------------------------|------------|-------------------------------|--------|---------|---------|---------|--------|-------|
|                                  |            | <5                            | 5< <10 | 10< <15 | 15< <20 | 20< <24 | 25+    |       |
| Estimated Population (Thousands) |            |                               |        |         |         |         |        |       |
| TOTAL 12–18                      | 24,288     | 1,319                         | 1,592  | 1,616   | 2,197   | 1,821   | 12,233 | 3,511 |
| Telephone Status                 |            |                               |        |         |         |         |        |       |
| TOTAL                            | 1.00       | 1.00                          | 1.00   | 1.00    | 1.00    | 1.00    | 1.00   | 1.00  |
| Reported                         | 0.89       | 0.72                          | 0.74   | 0.80    | 0.86    | 0.92    | 0.96   | 0.83  |
| Refused                          | 0.04       | 0.01                          | 0.03   | 0.03    | 0.03    | 0.04    | 0.03   | 0.08  |
| No phone                         | 0.07       | 0.26                          | 0.22   | 0.15    | 0.09    | 0.04    | 0.01   | 0.08  |
| Unknown                          | 0.01       | 0.01                          | 0.01   | 0.01    | 0.01    | 0.00    | 0.01   | 0.02  |

Table 3. Coverage of the Population 12–18 in Households with Available Telephones by Age and Race and Family Income

| Age                         | Family Income in Thousands |        |         |         |         |      | Unknown | Total |
|-----------------------------|----------------------------|--------|---------|---------|---------|------|---------|-------|
|                             | <5                         | 5< <10 | 10< <15 | 15< <20 | 20< <25 | 25+  |         |       |
| Coverage                    |                            |        |         |         |         |      |         |       |
| Population 12–18            |                            |        |         |         |         |      |         |       |
| 12–13                       | 0.58                       | 0.71   | 0.84    | 0.85    | 0.91    | 0.96 | 0.82    | 0.88  |
| 14–16                       | 0.62                       | 0.74   | 0.80    | 0.87    | 0.92    | 0.95 | 0.83    | 0.89  |
| 17–18                       | 0.83                       | 0.76   | 0.78    | 0.86    | 0.92    | 0.96 | 0.84    | 0.89  |
| TOTAL                       | 0.72                       | 0.74   | 0.80    | 0.86    | 0.92    | 0.96 | 0.83    | 0.89  |
| Black Population Aged 12–18 |                            |        |         |         |         |      |         |       |
| 12–13                       | 0.62                       | 0.68   | 0.75    | 0.82    | 0.92    | 0.91 | 0.69    | 0.78  |
| 14–16                       | 0.66                       | 0.72   | 0.82    | 0.91    | 0.89    | 0.95 | 0.73    | 0.82  |
| 17–18                       | 0.68                       | 0.72   | 0.86    | 0.88    | 0.96    | 0.91 | 0.75    | 0.82  |
| TOTAL                       | 0.65                       | 0.71   | 0.81    | 0.88    | 0.92    | 0.93 | 0.72    | 0.81  |
| NonBlack Population 12–18   |                            |        |         |         |         |      |         |       |
| 12–13                       | 0.55                       | 0.72   | 0.87    | 0.86    | 0.90    | 0.96 | 0.84    | 0.90  |
| 14–16                       | 0.57                       | 0.75   | 0.80    | 0.86    | 0.93    | 0.96 | 0.85    | 0.90  |
| 17–18                       | 0.86                       | 0.78   | 0.75    | 0.86    | 0.92    | 0.96 | 0.85    | 0.90  |
| TOTAL                       | 0.76                       | 0.75   | 0.80    | 0.86    | 0.92    | 0.96 | 0.85    | 0.90  |

Table 5. Proportion of the Population Aged 12–18 in Households with an Available Telephones by Region, Race, and Poverty Status Reported with Excellent/Very Good Health

| Race and Poverty Status | U.S. | Region  |      |       |      |
|-------------------------|------|---|------|-------|------|
|                         |      | East  | NC   | South | West |
|                         |      | <u>Proportion in<br/>Excellent/Very Good Health</u> |      |       |      |
| U.S.                    | 0.78 | 0.80  | 0.80 | 0.74  | 0.80 |
| Above poverty line      | 0.81 | 0.83  | 0.83 | 0.79  | 0.82 |
| Below poverty line      | 0.64 | 0.62  | 0.68 | 0.59  | 0.68 |
| UN                      | 0.67 | 0.76  | 0.72 | 0.60  | 0.68 |
| Black                   |      |   |      |       |      |
| Above poverty line      | 0.68 | 0.74  | 0.72 | 0.63  | 0.69 |
| Below poverty line      | 0.55 | 0.37  | 0.61 | 0.55  | 0.66 |
| UN                      | 0.60 | 0.81  | 0.64 | 0.57  | 0.39 |
| NonBlack                |      |   |      |       |      |
| Above poverty line      | 0.83 | 0.84  | 0.84 | 0.82  | 0.83 |
| Below poverty line      | 0.67 | 0.73  | 0.71 | 0.62  | 0.68 |
| UN                      | 0.69 | 0.75  | 0.74 | 0.62  | 0.70 |

Table 6. Proportion of the Population Aged 12–18 in Households without an Available Telephones by Region, Race, and Poverty Status Reported with Excellent/Very Good Health

| Race and Poverty Status | U.S. | Region  |      |       |      |
|-------------------------|------|---|------|-------|------|
|                         |      | East  | NC   | South | West |
|                         |      | <u>Proportion in<br/>Excellent/Very Good Health</u> |      |       |      |
| U.S.                    | 0.66 | 0.73  | 0.71 | 0.58  | 0.70 |
| Above poverty line      | 0.77 | 0.80  | 0.79 | 0.76  | 0.75 |
| Below poverty line      | 0.52 | 0.58  | 0.62 | 0.40  | 0.68 |
| UN                      | 0.61 | 0.72  | 0.71 | 0.54  | 0.58 |
| Black                   |      |   |      |       |      |
| Above poverty line      | 0.63 | 0.84  | 0.42 | 0.61  | 0.68 |
| Below poverty line      | 0.48 | 0.55  | 0.53 | 0.42  | 0.76 |
| UN                      | 0.66 | 0.84  | 0.78 | 0.58  | 0.53 |
| NonBlack                |      |   |      |       |      |
| Above poverty line      | 0.80 | 0.80  | 0.85 | 0.81  | 0.76 |
| Below poverty line      | 0.54 | 0.61  | 0.66 | 0.39  | 0.67 |
| UN                      | 0.59 | 0.59  | 0.70 | 0.50  | 0.59 |