Nonresponse Bias in a Travel Survey of Nontelephone Households*

J. Neil Russell, Jonaki Bose and Lee H. Giesbrecht
U.S. Department of Transportation, Bureau of Transportation Statistics, Washington, DC 20590

Abstract

The Bureau of Transportation Statistics (BTS) co-sponsored two national random digit dial (RDD) telephone surveys during the summer of 2002. Stakeholders expressed concern about the potential bias in travel-related behavior estimates because the sample designs of these surveys excluded the four percent of nontelephone (non land line) households. Nontelephone households may have different travel-related behaviors compared to telephone households. While there have been other studies of nontelephone households, they have not focused on travel-related behaviors. BTS responded to this concern by conducting the Supplemental Survey, a travel-behavior-related nontelephone household survey designed to investigate possible differences. However, the survey only yielded an unweighted CASRO response rate of 21 percent. Because data for both respondents and nonrespondents were available from the sample frame, we conducted a nonresponse bias analysis for this unique survey. One of the purposes of this work was to identify whether nonrespondents in travel surveys are different compared to nonrespondents in other surveys. The analysis includes both bivariate and multivariate analyses to ascertain predictors of nonresponse to the survey. Multivariate analyses indicate that in this nontelephone household mail survey, nonrespondents were more likely to be under 35, male, have less than a high school education, and live in a household with more than one person, children, and no vehicle.

Keywords: Nontelephone Households, Nonresponse Bias, Mail Survey, Recontact

Background on the Supplemental Survey

BTS conducted the Supplemental Survey, a nontelephone household mail survey designed to investigate possible travel-behavior-related differences between telephone and nontelephone households. Prior to this, BTS co-sponsored two travel-related surveys that used telephone number-based samples. Random digit dial (RDD) telephone samples typically exclude households that have no land-line telephone service and most cell phone exchanges. This exclusion is considered sample frame bias (coverage bias). Even though estimates from BTS surveys attempt to account for this bias through post-stratification of weights, stakeholders were concerned that nontelephone households differed from telephone households on travel-related behavior. BTS RDD surveys would not capture differences in travel behavior between telephone and nontelephone households.

About four percent of U.S. households[1] do not have telephone service. These nontelephone households are of three types:

- No telephone service at all,
- Temporary, interrupted telephone service, or
- Cell phone only households.

BTS was interested in collecting travel data from households who had no telephone. Collecting travel data from nontelephone households to supplement data collected via RDD sampling methods would help BTS assess the statistics from existing surveys and serve to improve the reliability of those statistics.

However, the Supplemental Survey only had a 21 percent unweighted CASRO response rate. Thus, before data from this survey could be used to supplement travel estimates from the other RDD surveys, an evaluation of the quality of the data was needed. The paper examines characteristics of respondents and nonrespondents to the Supplemental Survey and is based on sample frame characteristics. The next sections describe the survey, method of analysis, and the results of the nonresponse bias analysis.

Overview of Recent BTS RDD Surveys

The two RDD surveys sponsored by BTS during the summer of 2002 were the:

- 2002 National Survey of Pedestrian and Bicyclist Attitudes and Behaviors (Bike/Ped Survey), and
- 2002 National Transportation Availability and Use Survey (Disability Survey).

The Bike/Ped Survey was designed to collect general information on bicycling and walking. The main objectives of the survey were to collect information on:

- The scope and magnitude of bicycle and pedestrian activity,
- Origin and destination and trip purpose,
- Attitudes and opinions on bicycling and walking, and
- Pedestrian and bicycle safety practices.

Telephone interviews were conducted with 9,616 persons 16 years or older between June 11 and August 20, 2002. The overall unweighted CASRO response rate for the survey was 27 percent. The results of the survey are only applicable to the 2002 summer season (summer defined as May through August).

* Paper presented at the 2004 AAPOR annual meetings, Phoenix, Arizona.

[1] December 2002 Current Population Survey.

The Disability Survey was initiated because of the President's 2000 "New Freedom Initiative" that included an objective to expand transportation options for persons with disabilities. In response, BTS planned a national survey to collect data about how persons with disabilities (PWDs) use transportation, what barriers they face, and their overall satisfaction with the transportation system. The objectives of the survey were to:

- Benchmark PWD transportation experiences,
- Provide comparison data between PWD and persons without disabilities relative to their transportation experiences, and
- Provide data on use of adaptive equipment.

The survey was fielded from July 12 to September 29, 2002 using telephone interviewing. The goal was to complete equal numbers of interviews with disabled and non-disabled persons. The actual number of completed interviews was 2,321 persons with disabilities and 2,698 non-disabled persons for a total of 5,019. Proxy respondents were used for those persons who could not respond for themselves or who were less than 18 years of age. The overall unweighted CASRO response rate for the survey was 56 percent.

BTS Supplemental Survey Design

BTS identified the need for the Supplemental Survey in February 2002. The survey had to be planned, designed, sample procured, and contract awarded in a very short time period. BTS wanted the Supplemental Survey to be fielded during the same time period as the other RDD surveys so that comparison of statistics between surveys would not be complicated by different data collection periods.

Identifying Nontelephone Household Sample Frame - BTS undertook market research to identify potential sources of a probability sample of nontelephone households for the Supplemental Survey.

The first phase of the market research eliminated all federally funded surveys that included nontelephone households because of confidentiality issues related to restricted access. BTS had only four months to procure a sample of nontelephone households, and this expedited schedule hindered the necessary planning needed to work out the confidentiality issues in using a Federal agency's sample.

The second phase identified privately funded surveys (which would not have restricted access) that included nontelephone households. Four privately funded surveys were identified. However, before using any sample of nontelephone households, BTS set three conditions that had to be met. The sample had to:

- Use probability methods of selection, down to the final unit of analysis (household or person),
- Use random methods of selection, and
- Be no more than 2 years old.

These conditions would ensure that reliable data were collected, inferences could be made from the data, and the sample frame would be relatively complete, thus reducing the level of noncontact.

Four private surveys having samples of nontelephone households were assessed on how they met the three conditions. The Gallup Organization and RoperASW "Omnibus Survey" samples did not meet these conditions. They were either too old or did not use random selection methods down to the level of the unit of analysis.

The Urban Institute's National Survey of American Families (NSAF) and Mediamark Research's (MRI) Survey of American Consumers (SAC) met these conditions. The sponsors of these two surveys were contacted about access to their samples of nontelephone households. The Urban Institute declined access to their sample. However, MRI indicated a willingness to allow BTS access to their sample under the condition they conduct the survey for BTS. BTS entered into a sole source contract with MRI for access to their nontelephone household sample and to conduct the survey.

Sample Design of Supplemental Survey - The Supplemental Survey relies on the sample design of the SAC. SAC collects information from 25,000 adults twice per year on various commercial product preferences and buying habits. SAC uses a commercial file of listed residential telephone numbers augmented by other sources of household information, but excludes telephone numbers from Hawaii and Alaska. The sample frame has more than 90 million listings. SAC's sample design consists of three strata:

- Ten major media markets, each of which is a self-representing primary sampling unit (PSU),
- Other metropolitan statistical areas (MSAs) outside these ten markets, and
- Non-metropolitan statistical areas (non-MSAs).

Within the MSA and non-MSA strata, a set number of PSUs are selected based on probability proportionate to a weighted household count within each of these strata. Within each PSU, a set number of clusters are selected using a random start and a sampling interval. Each of these clusters contains a starting address and the next 14 listings.

All households listed within each cluster are confirmed in-person by a MRI interviewer. Listing sheets with selected addresses are assigned to interviewers, who then canvass the area (cluster) and record any additional dwellings (new construction, non-telephone households and those with unlisted numbers; group quarters are excluded). This method accounts for all dwelling units within each cluster. Hence, nontelephone and unlisted households are captured in the sample.

Within each cluster, the final sample of housing units is selected through a systematic, random selection process. The selected households are randomly predesignated to select either a male or female adult respondent (18 years or older). When the interviewer contacts the sample household, they record the names and ages of all adults of the predesignated sex on a grid. A random number is assigned for selecting the household adult of the specified sex. If the household has no adult member of

the predesignated sex, then all adult names are listed and a sample respondent is randomly selected.

Nontelephone Sample Selection - Using four waves of SAC sample, MRI identified 1,087 adults from nontelephone households for the Supplemental Survey[2]. The sample for this study consisted of all interviewed households that reported having no landline telephone when the SAC interview was conducted. These households constituted the sample for the Supplemental Survey.

Data Collection - Once MRI identified all nontelephone households for the Supplemental Survey, the contact information for these households was updated using the National Change of Address (NCOA) to reduce noncontact. As a result, 136 addresses were updated and 79 were determined to have address or name problems or the respondent moved. Advance letters were sent out to all valid addresses, followed by a mailout of the questionnaire a week later. Based on the initial mailing, 149 respondents sent back questionnaires. Reminder postcards were sent to the nonrespondents, followed by a second questionnaire mailing about two weeks later. The field period was from October 8 to December 31, 2002. In the end, a total of 230 completed questionnaires out of 1086 (1 deceased) were received.

MRI processed and keyed the data from returned questionnaires, and merged SAC respondent data with BTS survey data to facilitate analysis of nonresponse.

Original Analysis Plan - The survey design and analysis plan were based on making travel-related behavior estimates for nontelephone households. To do this, survey design differences between the two BTS RDD surveys and the Supplemental Survey were minimized. Every effort was made to keep question wording, reference periods, skip patterns, and fielding periods the same across the surveys. This would enable definitive comparisons of estimates between the Supplemental Survey (nontelephone households) and the BTS RDD surveys (telephone households). However, as the analysis work progressed, complicating issues arose which impacted the analysis and the results.

Low Response Rates and Nonresponse Bias

As mentioned before, the Supplemental Survey resulted in an unweighted CASRO response rate of 21 percent. With such a low response rate, confidence in any estimate is difficult. The Supplemental Survey did have some follow-up measures that helped improved the response rates but it was not possible to get an adequate response because of out-of-date recontact information due to the time lapse between the original SAC interview and the Supplemental Survey fielding. Despite using the NCOA service, 292 of the 1,087 addresses were unusable. Hence, before BTS could field the survey, the noncontact rate was already at 27 percent. Survey budget restrictions made the survey mode mail only. No in-person interviews were planned because of the extreme cost. If costs were not a consideration, the survey would have been designed differently, with a focus on in-person contact to mitigate a high noncontract rate.

Characteristics of Nonrespondents - We examined the literature on characteristics of survey nonrespondents to help develop our analysis. Groves and Couper (1998) looked at data from six federal demographic surveys that were matched to the 1990 decennial census to fill in information about survey nonrespondents that completed the census. They found higher response rates among those households with all members under 30 and households with all members over 69. However, household size interacted with age. While older persons tended to be more willing to cooperate overall, there was lower response from elderly persons living alone. Generally, larger households tended to cooperate more. They found lower response rates from those living in high-cost housing in high-density urban areas. They also found little difference in response rates between non-Hispanic whites and other groups when controlling for socioeconomic class. In a similar match study using the American Community Survey and the 2000 Decennial Census, Leslie, et al (2003) found similar results for age and household size and replicated the results of other studies in finding lower response rates among males.

Black and Safir (2000) looked at nonresponse bias in the first round of the National Survey of America's Families using information about nonrespondents collected in the second round. They found results similar to those found by Groves and Couper. In addition, they found that nonorespondents tended to have experienced interruptions in telephone service.

One characteristic of mail surveys compared with other modes is the inability to distinguish between those that refuse to participate and those that are not contacted. Kojetin (1994) examined nonresponse in the Current Population Survey and the Consumer Expenditures Surveys. He found that characteristics of refusers were more similar to respondents than those of noncontacts. Dixon (2002) also examined the Consumer Expenditures Surveys and found a result of particular importance to this study. Nonrespondents to the Consumer Expenditures Quarterly Interview Survey had higher relative expenditure estimates for transportation. This does not necessarily mean that nonrespondents travel more (particularly for our sample of nontelephone households), but it does indicate the potential for nonresponse bias in transportation-related data.

While the survey research literature is rich with work on nonresponse in general, the topic of mail survey nonresponse is a small component. One of the few texts on this specific topic is Moore and Tarnai (2002). They explore the impact of monetary incentives on nonresponse bias and find some evidence that the incentive works to increase response rates among those already well represented by respondents, thus increasing the potential for bias.

Based on this prior research we expected nonrespondents to:

- Be single,

[2] The nontelephone households identified for the Supplemental Survey were in SAC during one of four waves (43 to 46) of data collection that occurred between March 2000 and April 2002. MRI reported the response rates for SAC waves 43 and 44 as 67 percent and waves 45 and 46 as 67.5 percent.

- Be younger,
- Be male,
- Have lower education levels,
- Have lower household income,
- Come from an urban household, and
- Live in a household with very few persons.

In addition, since the units of the sampling frame were interviewed at any one of four points in time from across two years, we included a time-based variable. Our hypothesis was that cases that were included in the sample from earlier interviews would be more likely to be nonrespondents.

Methodology

There were methodological problems that had to be addressed prior to conducting this study. Some of these problems in this study were highlighted in Bose et al (2003). In addition, Bose et al (2003) presented some preliminary findings that compared estimates between the Supplemental Survey and the two BTS RDD surveys. However, the methodological problems hampered any definitive conclusions about differences between the estimates. In this study, we addressed the weighting and variance estimation problems identified in Bose et al (2003).

Reweighting - MRI provided BTS with weights that came from SAC, but these weights were created for their own analytic purposes. The SAC weights were not post-stratified by the presence or absence of a telephone, so the weights did not sum to the total number of persons living in nontelephone households. These weights also represented four different time periods (waves), each with their own unique population control totals. For these reasons, BTS created revised weights by post-stratifying the original SAC weights to March 2001 CPS population control totals by age, race/ethnicity, and gender for nontelephone households.

Variance Estimation - MRI did not provide explicit sample design information or replicate weights for variance estimation. However, certain geographic information (MSA status) on clustering in the sample was used to construct pseudo strata and primary sampling units (PSUs) for variance estimation consistent with how the SAC sample was designed.

Sample Frame Analytic Variables - A number of sample frame variables were available for this study. Guided by previous research, we restricted the variables we examined to the following:

- Demographic: age, sex, race, ethnicity, marital status, employment status,
- Household: income, size (adults and children), presence of cell phones, presence of vehicle, urbanicity, language spoken at home,
- Behavior: watched TV, computer use, and
- Other: sample wave.

Statistical Analysis - Bivariate analysis of frame variables by response status was conducted to assess initial significant predictors. Then, logistic regression was used to evaluate factors affecting response status in a multivariate context. The dependent variable was whether the person was a respondent or not, and the independent variables were a selection of frame variables. Models were tested for significance at an alpha level of 0.05 percent. Both the bivariate and multivariate analyses were conducted using SUDAAN (using linear approximation).

Results

Based on the final logistic regression model, it appears that persons in nontelephone households are more likely to be nonrespondents if they:

- Have less than high school education,
- Have children in the household,
- Younger (less then 35),
- Male,
- Live in a household without a vehicle, and
- Live in a household having very few persons.

More specifically, persons with less than high school education were 2.5 times more likely to be a nonrespondent compared to those with a high school degree or some college. Those with no children in the household were 1.6 times less likely to be a nonrespondent. Adults between 18 and 24 were 4.0 times and adults between 25 and 34 were 2.5 times more likely to be a nonrespondents compared to adults over 35 years of age. Males were 1.6 times more likely to be a nonrespondent compared to females. Persons in a household with no vehicle were 1.8 times more likely to be a nonrespondent compared to households with one or more vehicles. Finally, each additional person in the household would decrease the likelihood of nonresponse by 1.2 times.

While we expected education, age, sex, and the presence of vehicles to affect response status, we did not expect persons from households with children to be more likely to be nonrespondents. However, this could result from multicollinearity with household size that was also significant.

Contrary to our expectations, sample wave and household income was not significant. Household income was skewed which reflected the nature of the nontelephone household sample. There may not have been enough variation in the household income distribution to detect a significant effect. In addition, persons from urban areas were no more likely to be nonrespondents compared to persons from rural areas.

Conclusion

It is not completely clear what we can conclude from these results given the limitations of our study:

- By revising the weights, we tried to eliminate possible bias in the data due to sampling and non-sampling errors. Ideally, the weights should have been built from the original selection probabilities.
- Households in the Supplemental Survey represented nontelephone households. However, because as time went by, about 64 percent of the respondents indicated

they had a telephone in their households at the time of the BTS interview. While respondents and nonrespondents to this survey appear different, it is possible that a large percentage of the nonrespondents to the Supplemental Survey also had a telephone in their households.

- From the original 1087 households, 292 households could not be contacted due to wrong or missing address information. Hence, for this study, 292 households did not respond because of noncontact. It is possible our results and the response rate to the survey would be different if we had good addresses for these households or conducted the interviews in person.

Some of the six characteristics that were identified as differing between respondents and nonrespondents are related to travel behavior. For example, from the 2001 National Household Travel Survey, we know that males spend about 67 minutes a day behind the wheel compared to 44 minutes for females. Similarly, adults between 20 and 24 make on average 4.1 trips daily compared to 3.4 trips for adults 65 and older (USDOT/BTS 2003).

It is difficult to quantify how the findings from this study affect national estimates from RDD surveys. Nontelephone households form a very small percent of households, and nonrespondents in nontelephone households are an even smaller percent. While the difference between telephone and nontelephone households may not affect national estimates significantly or substantially, the non-inclusion of nontelephone households or the low response rates of nontelephone households in mails surveys could adversely affect estimates from subgroups with larger proportions of nontelephone households.

References

Black, T., and Safir, A. (2000), "Assessing Nonresponse Bias in the National Survey of America's Families," *Proceedings of the Survey Research Methods Section of the American Statistical Association*, Alexandria, VA.

Bose, J., Russell, JN., Giesbrecht, LH., and Cohen, MP. (2003) "Characteristics of Nontelephone Households: A Supplemental Survey to Two Travel Related Surveys," *Proceedings of the American Statistical Association*, [CD-ROM], Alexandria, VA.

Dixon, J. (2002) "Nonresponse Bias in the Consumer Expenditure Quarterly Survey," *Proceedings of the Survey Research Methods Section of the American Statistical Association*, Alexandria, VA.

Groves, RM. and Couper, MP. (1998) *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons, Inc.

Kojetin, BA. (1994) "Characteristics of Nonrespondents to the Current Population Survey (CPS) and Consumer Expenditure Interview Survey (CEIS)," *Proceedings of the Survey Research Methods Section of the American Statistical Association*, Alexandria, VA.

Leslie, TF., Raglin, DA, and Braker, EM. (2003) "Can the American Community Survey Trust Using Respondent Data to Impute Data for Survey Nonrespondents? Are Nonrespondents to the ACS Different from Respondents?" *Proceedings of the Survey Research Methods Section of the American Statistical Association*, Alexandria, VA.

Moore, DL. and Tarnai, J. (2002) "Evaluating Nonresponse Error in Mail Surveys," in RM Groves, DA Dillman, JL Eltinge, and RJA Little, (eds.) *Survey Nonresponse*, New York: John Wiley & Sons, Inc.

U.S. Department of Transportation, Bureau of Transportation Statistics. (2003) *NHTS 2001 Highlights Report*, Washington, DC.