Estimating Nonresponse Bias in the Omnibus Household Survey

Promod Chandhok
Department of Statistics, The George Washington University, 2140 Pennsylvania Avenue NW, Washington, DC 20052

Abstract

The Omnibus Household Survey is a random-digit-dial survey of slightly more than 1,000 non-institutionalized residents in telephone households in the U.S. Information is collected to monitor expectations of and satisfaction with the transportation system. Since the response rate for the survey is approximately 50%, there has been considerable interest in gauging the nonresponse bias in this survey. There are mainly three options for assessing nonresponse bias. The first option, used frequently, and hence widely accepted, is an intensive follow-up of nonrespondents. This was not an option here because of cost. The second option is external validation -- benchmarking against some other data, and this was done here for commonly used demographic variables. Another option is to use data on survey operations. This presentation discusses results from using call history data to estimate nonresponse bias in several key variables for the 2007 Omnibus Household Survey.

Key Words: travel surveys, call history, nonresponse error

1. Introduction

The Omnibus Household Survey (OHS) is a random-digit-dial (RDD) survey of slightly more than 1,000 non-institutionalized residents in telephone households in the U.S. From a target population of non-institutionalized persons 18 and above, information is collected to monitor expectations of and satisfaction with the transportation system. Also, event, issue, and mode-specific information is gathered. Examples of information collected is average days in a week the respondent drove in a car and the number of times s/he flew in a plane in a year, In general, these surveys support the collection of information on a wide range of transportation-related topics. The OHS was recently conducted in November 2006 and November 2007 and for these two instances the response rate has been around 50%. This has raised concerns whether nonresponse bias in important transportation related variables is large. For this research, call history data were used to estimate nonresponse bias in five key survey variables; three of these variables involve travel-related behavior while two measure satisfaction with air travel.

With declining response rates in household surveys, nonresponse bias is a growing concern for many survey practitioners. Several nonresponse bias studies have been conducted with mixed findings; some are discussed in the next section.

2. Review of Literature

Siemiatycki and Campbell (1984) compared nonresponse bias in early versus all responders in a mail and telephone health survey. The initial mail survey using 1,555 adults achieved 68.5% response and a follow up by telephone and home interview increased the response rate to 80.9%. The responders to the mail survey were designated as early responders and they were similar to all responders on age, sex, education and reported health status. In a parallel study, an initial survey of 1,595 adults was done by telephone. This achieved a 72.7% response rate. A follow-up was done by mail and personal interview and the response rate increased to 88.2%. Comparisons between respondents and the entire sample and between initial stage respondents and all respondents found no difference in terms of age, sex, education and reported health status. They concluded that in both survey strategies, there was no evidence of substantial non-response bias and estimates of morbidity and health care would not have differed much if follow-up were not conducted.

O'Neil (1979) reported results from a telephone survey of 1,392 eligible Chicago households in which there were 1,037 completed interviews after the first contact and 1,209 after the second. He called the 1,037 cases "amenables" and the

172 households as "resistors" and compared the characteristics of the two groups. He found that resistors are older, are more commonly white, have less education and lower incomes than is characteristic of the entire sample.

Thompson (1979) investigated 1,500 calls from the 15,234 "no answer/busy" calls in a telephone interview survey conducted by the US Coast Guard in 1977. In this survey, the population of interest was boating households. He found that under the assumption that these 1,500 represent 15,234, using this information, the resulting estimates showed 19.4% fewer boats, 25.0% fewer boat operators and 20.3% fewer operator households – a seemingly large nonresponse bias.

Triplett et al. (2002) reported results from a 2,000 finalized telephone number follow-up of the National Survey of America's Families (NSAF). They found NSAF refusals to be more likely from larger households, black, to have graduated from high school or have received their GED. Those who completed NSAF are more likely to be white, own their own home and be unemployed. Cohen and Duffy (2002) concluded that nonrespondents to health surveys are generally less well than respondents.

Friedman et al. (2003) from a study of beneficiaries of the Military Health System conclude that "late respondents" differ from early respondents and nonrespondents in terms of enrollment status, beneficiary status, age, marital status, and whether or not they reside in the U.S. However, for the most part, they do not appear to differ from early respondents in terms of satisfaction with TRICARE, health usage, and self-reported mental and physical health. Late respondents were also found to have significantly poorer data quality, which may indicate that many of their other responses are less reliable, as well. They remarked that data quality would be improved if a different mode were used as the follow-up rather than a second mail questionnaire.

McFarlane et al. (2006) found that male physicians are more likely to respond than female physicians. Male physicians were also significantly more likely to be early responders. Also, survey length had a significant impact on response rates. Physicians who received a short questionnaire were significantly more likely to respond than physicians receiving the slightly longer questionnaire.

Billiet (2007) classifies respondents into "cooperative respondents" and "reluctant respondents" and assumes that the reluctant respondents have much in common with the refusals. He notes that this viewpoint is supported by evidence and postal surveys and cites a 2003 paper written in Dutch by him. He looked at data from several European countries but analyzed data from only Germany and Netherlands because the numbers of reluctant respondents in other countries were too small for this type of analysis. The findings were mixed. For some variables, the effect of "type of respondent" was insignificant, while for others, it was significant, but small. Also, the relationship between the type of respondent and the attitudinal and background variables was not all in the same direction in all countries. In the next section, we analyze the determinants of nonresponse bias.

3. Methodology

Suppose the population is divided into two strata: a stratum of respondents and a stratum of nonrespondents. Suppose N is the population size, and N_{nr} the number in the nonresponse stratum, $\bar{Y}_r$ the population mean of the y-values in the response stratum and $\bar{Y}_{nr}$ the population mean of the y-values in the nonresponse stratum. Then,

$$\text{Bias } (\bar{y}_r) = \frac{N_{nr}}{N}\left(\bar{Y}_r - \bar{Y}_{nr}\right)$$

We see that the nonresponse bias is the product of the nonresponse rate and the difference in means between the respondents and the nonrespondents. If the means of the 2 strata are the same then the bias is zero, irrespective of the nonresponse rate. A small nonresponse rate (or a large response rate) does not ensure small bias; all it ensures is that the multiplier to the difference in means is small. See, for example, Chandhok (2006).

There are mainly three approaches to estimating the nonresponse bias. The first one is more commonly used and hence widely acceptable, and this is a follow-up of nonrespondents. Here, we assume that the respondents to the follow-up represent the nonrespondents and for the results of a follow-up to be viewed as valid, a large response rate in the

follow-up is necessary. The second is benchmarking using external data. For example, if we know the proportion of drivers in the population and if we ask each respondent whether or not she or he is a driver, the sample proportion of drivers can be benchmarked to the population proportion. The difference in the two gives you nonresponse bias. The third approach is to use information from survey operations such as call history records.

Since the follow-up of nonrespondents was not possible for lack of funds and we did benchmark to known totals for age, gender, race and national origin, this research estimates the nonresponse bias using call history records. We assume that there is a continuum of resistance to responding to the survey and respondents with similar resistance are similar in transportation characteristics. In this research, we use call attempt as a proxy for resistance to response. Due to the small overall OHS sample size (for this type of research), we saw that for later call attempts, the number of responses was either very small or zero and hence we decided to assume that response is a dichotomous variable indicating early versus late response. Since I was looking at only five variables, I decided to examine the distribution of means by call attempt and look for a cutoff for the tail of the distribution, as the nonrespondents would be represented by the extreme right of the distribution of the mean. I wanted similar and enough observations in each of the early and late response strata. This approach takes into account both unit and item nonresponse. There is no existing theory for cutoff determination and for a multipurpose survey with many questions it may not be practical to look at, for each characteristic, the distribution of mean by call attempt. Survey practitioners would prefer one cutoff for the whole survey and if the shape of the distribution of the mean is the same for all characteristics, then you would get one cutoff. If the shape is different for different characteristics, then you may end up with different cutoffs for different characteristics.

4. Analysis and Results

In the 2007 OHS, data were collected in November 2007 and at most 37 call attempts were made to a phone number in the sample. The first key characteristic analyzed for this research is the number of days, in a typical week, a person drives or rides in a car, van, SUV, pickup truck, RV or motorcycle. We see that the response to this question ranges between 0 and 7 and so the nonresponse bias is bounded by 7. Table 1 gives the number of responses by call attempt, the mean response by call attempt and the coefficient of variation (CV) of the estimated mean.

Figure 1 is a graph of the average number of days in a car, van, SUV, pickup truck, RV or motorcycle against call attempt. We observe large fluctuations after call attempt 22. This is expected due to small sample sizes. For a cutoff between early and late responders, call attempt 9 was selected. There were 203 observations in the late response stratum. While the average number of days based on early responders is 5.4567, it is 5.8126 for late responders. Assuming that the weighted difference between early and late responders is the same as that between responders and nonresponders, the nonresponse bias is - 0.176 and the percentage nonresponse bias is 3.28%.

The second key variable is the proportion that flew on a commercial airline in the past year. Table 2 presents the proportion that flew on a commercial airline in the past year by call attempt and an estimate of the CV of the proportion. Figure 2 shows the estimated proportion that flew on a commercial airline in the past year by call attempt. It can be seen that the proportion lies between 0.35 and 0.41 in the first few call attempts and then fluctuates outside this band. Due to small sample sizes, this was not unexpected. Here, call attempt 9 was a good cutoff. The proportion for early responders is 0.3571and for late responders is 0.4529. The nonresponse bias is -0.0474 and the percentage nonresponse bias is 14.71%

The next key characteristic analyzed is waiting time for a screening checkpoint. Figure 3 gives the average waiting time for screening checkpoint (in decimal hours) by call attempt. Since the total number of responses for this question was 371 -- about a third of the number of responses to the previous two questions, more volatility in the average by call attempt is expected and observed. Here, call attempt 10 was a good cutoff. The average waiting time for screening checkpoint for early responders is 0.2671 hours, for late responders is 0.2721 hours. The nonresponse bias is –0.0025 and the percentage nonresponse bias is 0,93%

Table 1: Mean Number of Days in a Car, Van, SUV, Pickup Truck, RV or Motorcycle

Call Attempt	*n*	*Mean*	*CV of Mean*
1	226	5.3272	2.88%
2	138	5.4150	4.20%
3	106	5.3589	5.01%
4	104	5.6746	3.68%
5	63	5.5804	5.53%
6	50	5.5879	5.59%
7	49	5.0505	10.80%
8	52	5.9710	4.08%
9	38	5.4972	6.32%
10	34	5.5437	7.15%
11	22	5.8904	6.02%
12	22	6.4356	3.27%
13	23	6.0693	6.76%
14	12	6.1219	8.06%
15	10	5.8334	12.23%
16	10	5.4408	12.59%
17	9	6.1057	8.17%
18	7	6.6503	3.68%
19	9	5.9741	7.67%
20	7	6.0219	10.71%
21	3	5.9018	12.82%
22	9	5.8204	8.60%
23	5	7.0000	0.00%
24	2	1.4009	75.42%
25	2	5.9874	11.81%
27	1	2.0000	
37	1	7.0000	
Total	1014		

Table 2: Proportion that Flew on a Commercial Airline in the Past Year

Call Attempt	*n*	*Proportion*	*CV of proportion*
1	227	0.3137	11.28%
2	137	0.3313	14.44%
3	106	0.4152	14.00%
4	104	0.3949	14.70%
5	63	0.3143	20.96%
6	50	0.3612	21.18%
7	50	0.2960	27.67%
8	52	0.3335	21.28%
9	38	0.6233	14.22%
10	34	0.4741	21.36%
11	22	0.2338	39.72%
12	22	0.2325	41.94%
13	23	0.4787	23.07%
14	12	0.2729	60.53%
15	10	0.4298	40.18%
16	10	0.4103	41.83%
17	9	0.2992	57.18%
18	7	0.7720	19.33%
19	9	0.5046	37.62%
20	7	0.5825	36.33%
21	3	0.8583	17.39%
22	9	0.5908	32.67%
23	5	0.4014	64.59%
24	2	0.0000	0.00%
25	2	1.0000	0.00%
27	1	0.0000	
37	1	1.0000	
Total	1015		

In addition to the three travel-behavior characteristics analyzed, we also examined nonresponse bias in two 'satisfaction' items. One key 'satisfaction' characteristic is the overall satisfaction with the most recent security checkpoint experience. This variable is at four levels: (1) very satisfied, (2) satisfied, (3) dissatisfied, and (4) very dissatisfied. Figure 4 gives the average overall satisfaction with the most recent security checkpoint experience by call attempt. We see that the higher the number, the greater the dissatisfaction. Here, call attempt 10 was selected as the cutoff. The average overall satisfaction with the most recent security checkpoint experience for early responders is 1.8149, for late responders it's 1.8830. The nonresponse bias is -0.0337 and the percentage nonresponse bias is 1.88%

For the last key characteristic researched for this paper -- experience versus expectation of the most recent security screening wait time, Figure 5 gives the average value of this characteristic by call attempt. Here, call attempt 10 was selected as the cutoff. The average experience versus expectation of the most recent screening wait time for early responders is 2.7722; for late responders it's 2.7517. The nonresponse bias is - 0.0101 and the percentage nonresponse bias is 0.37%

Figure 1. Mean number of days per week in a car, van, SUV, pickup truck, RV or motorcycle

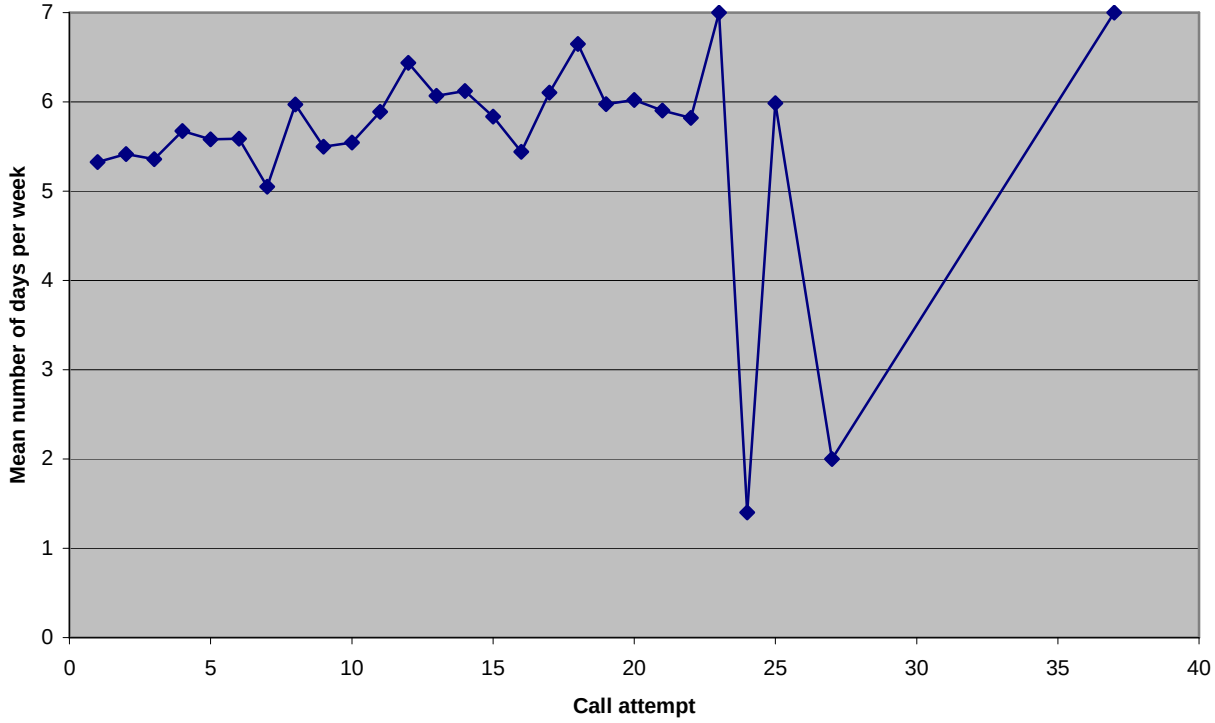


Figure 2. Proportion that flew on a commercial airline in the past year

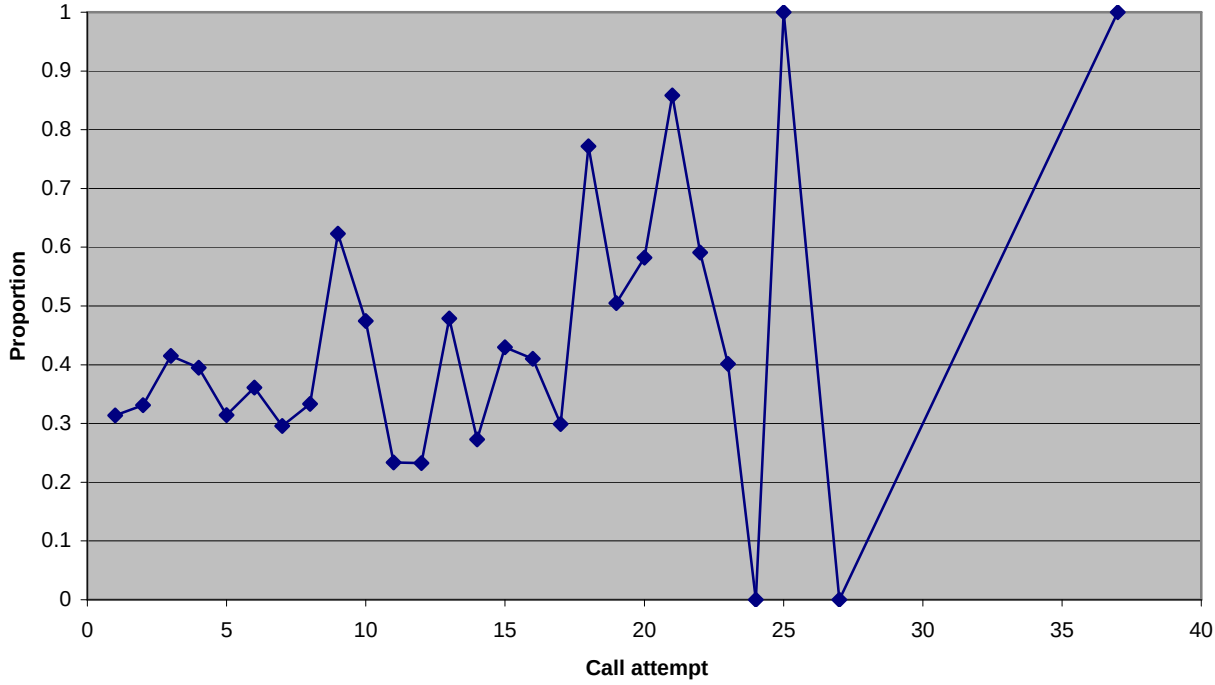


Figure 3. Waiting time for a screening checkpoint (in decimal hours)

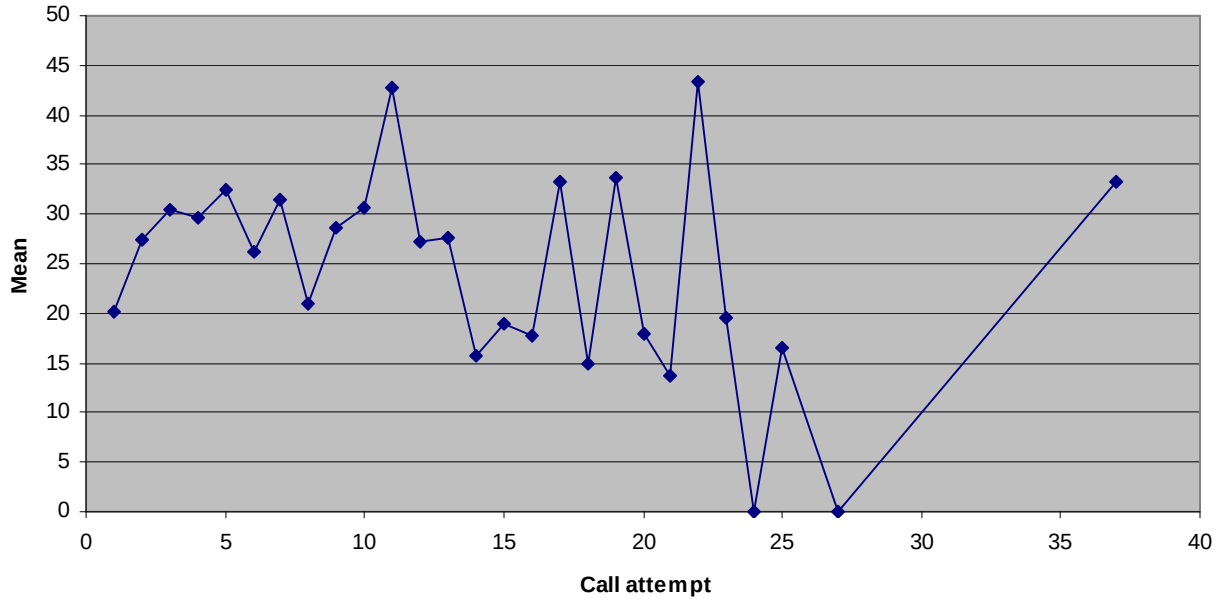


Figure 4. Overall satisfaction with the most recent security checkpoint experience

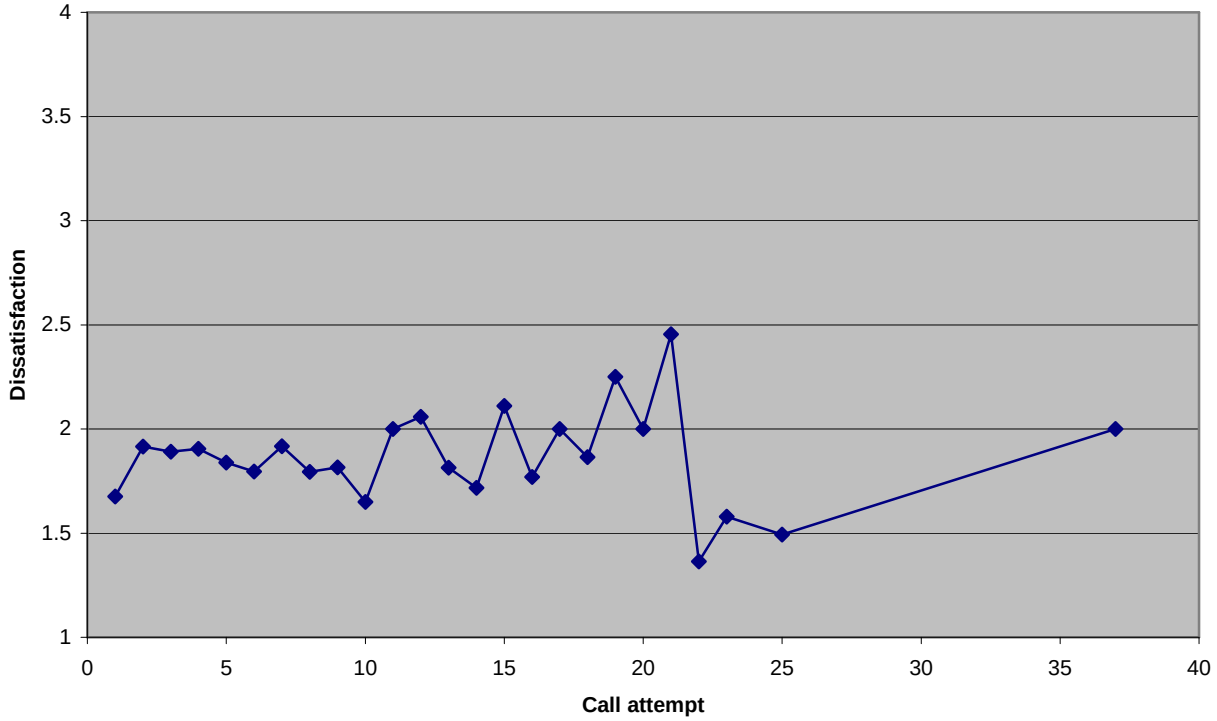
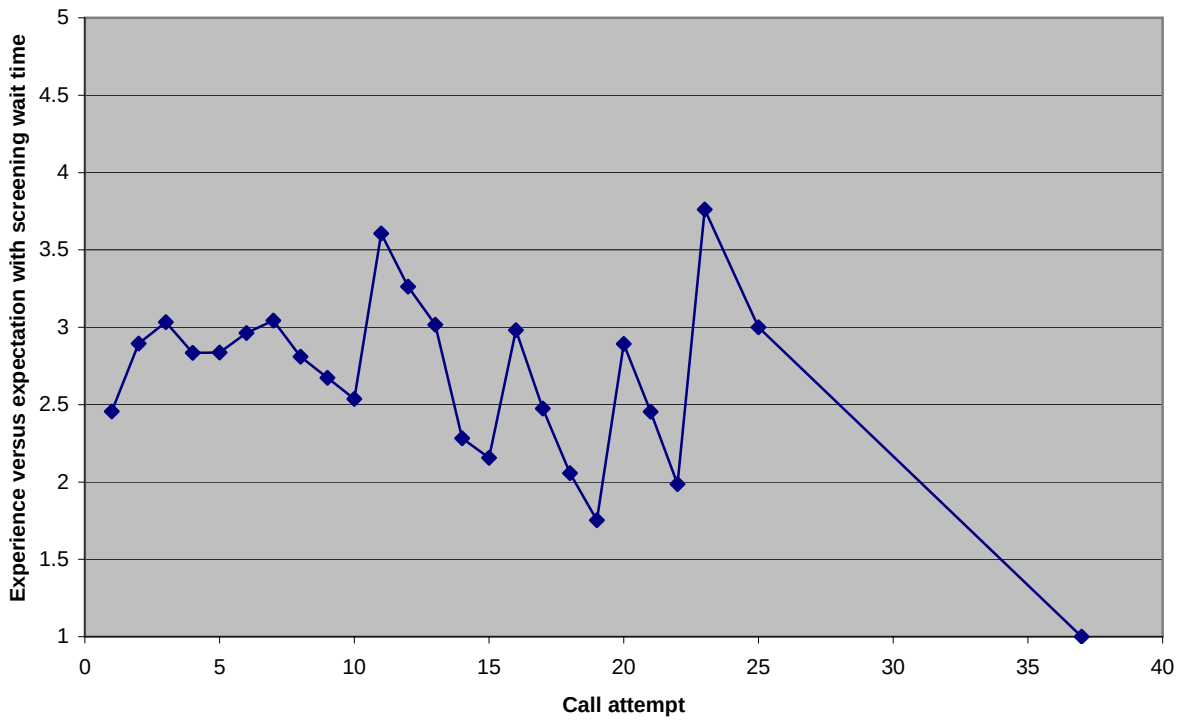


Figure 5. Experience versus expectation with the most recent screening wait time



5. Conclusion and Discussion

Table 6: Nonresponse Bias in Key Variables

Characteristic	*Estimate from Early Responders*	*Estimate from Late Responders*	*Nonresponse Bias*	*Nonresponse Bias (%)*
Mean number of days per week in a car, van, SUV, pickup truck, RV or motorcycle	5.4567	5.8126	-0.1760	3.28%
Proportion that flew on a commercial airline in the past year	0.3571	0.4529	-0.0474	14.71%
Waiting time for a screening checkpoint	0.2671	0.2721	-0.0025	0.93%
Overall satisfaction with the most recent security checkpoint experience	1.8149	1.8830	-0.0337	1.88%
Experience versus expectation with the most recent screening wait time	2.7722	2.7517	0.0101	0.37%

Table 6 provides estimates of key variables using early responders and late responders. It also gives the nonresponse bias and the percentage nonresponse bias in the five key variables.

The first two variables are related to travel behavior. For the first characteristic, the more a person drives or rides in a car, van, SUV, pickup truck, RV or motorcycle, the more likely the person is a late respondent. Nonresponse bias for the first characteristic, though small, is negative.

We see that nonresponse bias is relatively small for all the key variables except for the second -- proportion that flew on a commercial airline in the past year. An explanation is that late respondents may have a greater propensity to fly in a commercial airline than early respondents. A reason for being a late respondent could be that the person was traveling when called earlier.

The propensity to respond is related to travel characteristics and travel characteristics are also related to age, gender, ethnic origin and race. Poststratification for age, gender, race and ethnicity reduced nonresponse bias. Nonresponse bias would have been higher without poststratification.

Generally, the results suggest that late responders travel more than early responders and so for characteristics that are positively related to travel, the more a person travels, the higher the value of the characteristic. Nonrespondents are expected to have a higher mean and thus nonresponse bias is expected to be negative for those characteristics.

Acknowledgements

The author would like to thank Michael P. Cohen and reviewers for their comments. However, the views expressed in this paper are attributable only to the author.

References

Billiet, Jaak, Philippens, Michel, Fitzgerald, Rory and Stoop, Ineke (2007) Estimation of nonresponse bias in the European Social Survey: Using information from reluctant respondents, *Journal of Official Statistics*, 23, 135-162

Chandhok, P. (2006) Nonresponse Bias in the Omnibus Household Survey, *ASA Proceedings of the Joint Statistical Meetings*, American Statistical Association, Alexandria, VA.

Friedman, Esther M., Clusen, Nancy A., and Hartzell, Michael (2003) Better late: Characteristics of late respondents to a health care survey, *ASA Proceedings of the Joint Statistical Meetings*, 992-998, American Statistical Association, Alexandria, VA

Cohen, G. and Duffy, J. C. (2002) Are nonrespondents to health surveys less healthy than respondents? *Journal of Official Statistics*, 18, 13-23

McFarlane, Emily, Olmsted, Murrey G., Murphy, Joe and Hill, Craig A. (2006) Nonresponse bias in a mail survey of physicians, *ASA Proceedings of the Joint Statistical Meetings*, 4175-4180, American Statistical Association, Alexandria, VA

Siemiatycki, Jack and Campbell, Sally (1984) Nonresponse bias and early versus all responders in mail and telephone surveys, *American Journal of Epidemiology*, 120, 291-301

O'Neil, Michael J. (1979) Estimating the nonresponse bias due to refusals in telephone surveys, *Public Opinion Quarterly*, 43, 218-232

Thompson, N. R. (1979) Nonresponse bias from “no answer/busy" calls in a telephone survey, *ASA Proceedings of the Section on Survey Research Methods*, 250-251, American Statistical Association, Alexandria, VA

Triplett, Timothy, Wang, Kevin, Safir, Adam, Steinbach, Rebecca and Pratt, Simon (2002) Using a short follow-up survey to compare respondents and nonrespondents, *ASA Proceedings of the Joint Statistical Meetings*, 3496-3501, American Statistical Association, Alexandria, VA