# A Comparison of Level of Effort and Benchmarking Approaches for Nonresponse Bias Analysis of an RDD Survey

Daifeng Han and David Cantor
Westat, 1650 Research Blvd., Rockville, MD 20850

## Abstract

One method for studying nonresponse bias is to analyze estimates by level of effort (LOE) needed to complete an interview. LOE is typically measured using the number of times needed to contact a respondent and/or whether the respondent had initially refused to do the interview. LOE analyses are dependent on the assumption that non-respondents are similar to those who respond with more effort. A second method to assess nonresponse bias is to compare estimates to an external benchmark. In this paper, we compare the results of a nonresponse bias analysis for an random digit dialing (RDD) survey using a LOE approach to one that compares estimates to an external benchmark of a survey with a higher response rate. Important differences are found in the results between the two approaches. One conclusion is that the assumptions behind the level of effort model may not be correct for this survey.

**Keywords:** Response rate, refusal conversion, nonresponse adjustment.

## 1. Introduction

Assessing nonresponse bias requires making inferences about the characteristics of those persons for whom no observations have been made. The methods to make this assessment are based on inferences using assumptions about the non-respondents that are difficult to validate. In standard nonresponse bias analysis, two common methods are: (1) comparing estimates to a "benchmark" – an external data source with better and/or different error properties; and (2) examining variation in measures by the level of effort (LOE) used to complete an interview, where LOE is typically measured in two ways: the difficulty to contact a potential respondent, and the respondent's willingness to cooperate, once contacted.

Each of these methods has strengths and weaknesses. In this paper we compare the results of a nonresponse bias analysis for an RDD survey using a LOE approach to one that compares estimates to an external benchmark of a survey with a higher response rate. The findings from each method will be evaluated and discussed in more detail in the next two sections. The final section will compare the results across different methods in an effort to emphasize the necessity to take multi-method approach for comprehensive nonresponse bias study.

### 1.1 Three Analyses for Evaluating Nonresponse Bias

We conducted three analyses to assess nonresponse bias. One was a benchmark analysis. The other two were based on LOE: 1) level of effort to contact potential respondents, and 2) level of effort to earn cooperation.

Benchmarking analysis compares survey estimates to an external data source with different, preferably better, error properties, such as higher coverage and higher response rate. The usefulness of benchmarking depends on the availability of external data source. It is usually difficult to find directly comparable data on the key measures of the survey. Quite often, the only available external data is demographic information – in such cases benchmarking analysis only provides an indication of nonresponse bias to the extent the auxiliary demographic variables are related to the outcome statistics of the survey.

Another method for assessing bias due to nonresponse is through an analysis by the LOE spent to complete an interview. The basis of this type of analysis is the assumption that those respondents that are interviewed with the most effort resemble those that are not interviewed at all. Using this logic, LOE analyses compare respondents who are interviewed with less effort to those interviewed with a greater amount of effort. If there is a difference, then under the above assumption, there is evidence of bias in the estimates.

There are two sources of nonresponse, each of which is a function of different types of effort. One source is whether or not the respondent was contacted by the interviewer. An interview cannot occur if the respondent is not approached by an interviewer. In

RDD surveys, for example, respondents who are home when the interviewer calls will be less subject to this type of nonresponse. The difficulty to contact the target respondents is also affected by the extent that respondents screen their telephone calls and/or primarily rely on mobile telephones to communicate. The second source of nonresponse is the respondent's willingness to cooperate, once contacted. Some respondents are more willing to participate in a voluntary survey than others. For RDD surveys, one of the major points where this type of nonresponse occurs is at the initial household contact to complete the screening interview. This is the point of the interview when the person answering the telephone is asked to answer a short battery of questions so that an adult within the household can be selected for the longer interview.

There are different respondent characteristics and survey design features that affect non-contacts and cooperation in RDD surveys. One important respondent characteristic affecting non-contact, for example, is how often there is someone at home to answer the phone. Survey design features that affect making contact are such things as the number of times the household is called and the extent that those calls are made at different times or days of the week. According to the Leverage-Salience Theory (Groves, Singer, and Corning, 2000), cooperation is the result of the interaction of respondent-specific priorities and the survey design. Individuals vary in their priorities of what are the most important reasons to consider when deciding on whether to cooperate. Some respondents may put the topic of the survey as most important when deciding on whether to participate, while others may look for signs that the survey is being done by a credible organization (e.g., federal government). The survey design influences cooperation through the presentation of features that may (or may not) coincide with concerns by the respondent.

Given these different nonresponse mechanisms, LOE analyses typically employ two measures. To assess the result of not contacting the respondent, analyses compare outcomes by the number of calls it takes to complete an interview. Those needing the most calls to complete an interview are assumed to resemble the non-respondents to a greater extent than those needing fewer calls. To assess the effects of cooperation, respondents are compared by whether they initially refused to do the interview to those that had to be called back and "converted" after initially refusing.

As stated above, the basis of LOE analysis is the assumption that nonrespondents are similar to those who respond with more effort. However, this assumption has been questioned by several studies (Lin and Schaeffer, 1995; Olson, 2006). The effectiveness of the LOE approach depends largely on the strategies used in the late data collection effort, particularly on whether additional effort has been made to gain hard-to-reach cases.

Considering the strength and weakness of each method, we took a multi-method approach in our study by first using benchmarking to obtain direct information on the nonrespondents on a limited set of variables, and then applying the LOE approach to draw inferences on the key outcome variables of the survey.

## 1.2 Data Source: Health Information National Trends Survey (HINTS)

The data used for conducting this study was from Health Information National Trends Survey (HINTS), which is an RDD survey to estimate the prevalence of cancer-related knowledge, attitudes and behaviors in the US adult (age 18+) population. The HINTS survey consisted of a short screener interview to determine household eligibility and a longer extended interview to collect the primary data of interest. The screening interview included information on the subject matter (i.e., nationwide survey about health issues related to cancer) and questions designed to assess eligibility and select the person that should be administered the extended interview. The extended interview was approximately 30 minutes in length.

The continuing erosion of response rates associated with RDD surveys has also been reflected in HINTS. The response rate for HINTS was approximately 20 percent in 2005, mainly due to a very low screener response rate. Although it has become increasingly clear that response rate is not a direct indicator of nonresponse bias (Biemer and Lyberg, 2003), such a low response rate has raised enough concerns about the quality of the data and therefore the necessity to assess bias in the important outcome measures of the survey.

## 2. Benchmarking Method Using the National Health Interview Survey (NHIS)

In benchmarking analysis, we compared selected estimates from HINTS with the counterparts from the National Health Interview Survey (NHIS). NHIS is sponsored by National Center for Health Statistics and

has a response rate of 69 percent for the sample adult questionnaire. Since the NHIS data are collected through a personal household interview, the survey also has better coverage relative to HINTS, especially of younger age group, minority group, and those without landline telephones.

The common variables between the two surveys are mainly related to general health and health services. The NHIS results were used as the "gold standard" to examine the quality of the HINTS data.

Table 1 shows the major differences between HINTS and NHIS estimates on health and health service measures. The HINTS standard errors on listed in the parentheses. The estimates from the two surveys are significantly different in all the variables except "% having access to health insurance". Compared to NHIS, the proportion of people who self-reported to have good or excellent health in HINTS is approximately 11 percent points lower. In all the 6 variables that measure the feelings of negative emotions, the HINTS estimates are significantly higher; Table 1 only listed sadness or nervousness. The HINTS survey also seems to have included more ever-smokers and more cancer patients.

Table 1.        Comparison of HINTS and NHIS estimates on health and health service measures[*]

| Health and health service measures | HINTS estimates | NHIS estimates | HINTS - NHIS |
|---|---|---|---|
| % in good or excellent health | 76.3 (0.81) | 87.6 | −11.3 |
| % feeling sad some, most, or all of the time | 19.4 (0.91) | 11.4 | 8.0 |
| % feeling nervous some, most, or all the time | 30.3 (0.93) | 15.6 | 14.7 |
| % smoked 100+ cigarettes in life time | 47.2 (0.94) | 42.1 | 5.1 |
| % ever had cancer | 11.3 (0.51) | 7.3 (0.17) | 4.0 |
| % having access to health insurance | 83.5 (0.73) | 84.9 (0.29) | −1.4 |
| % never visited a doctor in the past 12 months | 16.4 (0.85) | 18.6 (0.29) | −2.2 |

* All the numerical values in the table are in percentage points.

To interpret the differences, we should keep in mind that (1) such difference is a combined effect of noncoverage and nonresponse; and (2) mode effect may be a potential confounding factor. Existing literature does not suggest consistent patterns of a mode effect for health-related variables. The difference of such magnitude suggests that respondents on HINTS tend to be less healthy both physically and emotionally.

Benchmarking analysis has shed some light on the characteristics of nonrespondents. However, among the approximately 200 HINTS variables, only less than 30 were also measured in NHIS and these variables were mainly on demographics and general health and health services. The key outcome variables in HINTS were cancer-related information seeking measures, for which no external data can be used as reliable comparable base. Due to the limitation of benchmarking method, we used LOE analysis to evaluate key outcome measures.

**3. LOE Analysis – Effort to Contact Respondent**

During the field period of HINTS 2005, the cases were released by "batch groups". Each batch group can be considered a cross section of the total sample. As the field period progressed, yield from the initial sample turned out significantly lower than expected. This resulted in the release of additional batch groups relatively late in the field period in order to increase the number of completes for the survey. However, batch groups released late in the field period were not subject to as many callbacks as those released early. This had a significant effect on the screener level response rate.

Three batch groups were used for the purpose of nonresponse bias analysis, referred to as long, middle, and short groups, corresponding to longest, medium, and shortest length of period respectively.

Table 2 shows that a case in the short field period group received an average of 5 calls at the screener level, while the long field period group received

approximately 7 calls on average. As a result, the screener level response rates are positively associated with the length of field period, as indicated by the difference of 8.6 percentage points between the long and the short groups. On the other hand, at the extended interview level, the response rates across groups were not noticeably different (data not shown).

Table 2.     Three batch groups for LOE analysis – effort to contact target respondents

|  | Full sample | Length of field period | | |
|---|---|---|---|---|
|  |  | Long | Median | Short |
| Size of screener sample | 35,802 | 10,951 | 15,065 | 9,875 |
| Average number of call attempts at the screener level | 6.15 | 7.18 | 6.05 | 5.14 |
| Unweighted screener response rate | 34.8% | 39.7% | 33.7% | 31.1% |
| Final completed cases | 5,586 | 1,823 | 2,319 | 1,444 |

In the analysis of LOE to contact respondents, estimates from three groups were compared, first using base weights, and then using final adjusted weights.

No substantial differences were found between base weight and final weight estimates except on several demographics variables. In this paper we present base weight results because it gives better information on who the nonrespondents are.

A list of 49 variables were selected for the analysis, including demographics, health and health service measures, and communication measures. For simplicity we focused on examining differences between the two extreme groups – the longest and the shortest field period groups displayed statistically significant differences in 8 variables.

Table 3 shows the estimates by batch groups using base weights. It indicates that greater effort increases the yield on those who are typically difficult to contact on RDD surveys, such as male, low income population, and immigrants. The differences observed in health and communication variables may be due to the demographic differences across the groups.

When final adjusted weights were applied, the differences in demographic characteristics across groups were almost eliminated. This is mainly due to the calibration mechanism which adjusted for demographics. Applying final weights did not eliminate the differences in health and communication variables. This suggests that there were other factors that affected response propensity and were not captured through weighting adjustment.

Table 3.     Results of LOE analysis – effort to contact target respondents[*]

|  | Length of field period | | | Longest – shortest field period |
|---|---|---|---|---|
|  | Long | Median | Short |  |
| % male | 38.6 | 35.6 | 34.9 | 3.7 |
| % household income <=20,000 | 15.7 | 16.0 | 13.0 | 2.7 |
| % born outside US | 14.8 | 12.3 | 10.6 | 4.2 |
| % feeling nervous | 31.8 | 29.8 | 25.0 | 6.8 |
| % had mammogram in the past 12 months | 66.9 | 65.8 | 73.3 | -6.4 |
| % read newspaper every day | 36.3 | 41.3 | 43.0 | -6.7 |
| % found cancer info too hard to understand | 27.2 | 23.0 | 20.3 | 6.9 |
| % use health provider as primary source of cancer information | 58.4 | 55.2 | 55.0 | 3.4 |

* All the numerical values in the table are in percentage points. All differences between short and long are statistically significant at p<.05.

### 4. LOE Analysis – Effort to Convert Initial Refusers

In this section a second measure of LOE is used representing the level of cooperation of the respondents. This analysis assesses nonresponse bias by comparing estimates from initial cooperators and to those that completed the interview after initially refusing but then agreeing to participate after being called back.

As noted above, the HINTS survey consisted of a short screener interview to determine household eligibility and a longer extended interview to collect the primary data of interest. Table 4 shows the distribution of final respondents based on whether they had initially refused to complete the screener or extended interview.

Approximately 30 percent of the completed cases had initially refused the interview. Similar to other national telephone surveys the HINTS also experienced its highest rates of nonresponse when attempting to complete the screener survey. Among the converted refusals, approximately two-thirds of the cases had initially refused
.

the screener – once they were persuaded to complete the screening, they continued to cooperate in the extended interview. The rest of the converted cases had initially refused extended interviews, only a small proportion of which had also refused screener initially.

In 15 out of the 49 measures we examined, significant differences were found between the initial cooperators and converted refusers. When breaking down refusers by the nature of refusal, we saw more substantial differences for extended refusers. Among all the significant measures, three variables stood out for more discussion – age, ever had cancer, and likelihood to seek cancer information.

Table 5 indicates different characteristics of the initial cooperators and the converted refusers in terms of their age and cancer history. The p-value for "% ever had cancer" is approximately 7 percent; all the other differences were all significant at the 5 percent level.

Compared to the initial cooperators, the converted refusers were more likely to be age 65 and older, had ever had cancer, and have access to health insurance.

Table 4.      Distribution of respondents based on refusal history

| Refusal history | Cases among final respondents |
|---|---|
| Never had refused screener or extended interview | 69.3% |
| Only had refused screener (but never refused extended interview) | 20.5% |
| Ever had refused extended interview and never had refused screener | 7.8% |
| Ever had refused extended interview and ever had refused screener | 2.4% |
| Total number of completed cases | 5586 |

Table 5.      Results of LOE analysis – effort to covert initial refusers: demographic and health variables[*]

| | | | Among converted refusers | |
|---|---|---|---|---|
| | Initial cooperators | Converted refusers | Only refused screener | Ever refused extended interview |
| % age 18 – 34 | 23.8 | 20.0 | 22.5 | 14.9 |
| % age 65+ | 20.2 | 27.0 | 24.8 | 31.3 |
| % ever had cancer | 13.4 | 15.4 | 15.3 | 15.6 |
| % with health insurance | 85.4 | 87.7 | 88.2 | 86.5 |

* All the numerical values in the table are in percentage points.

By breaking down the refusers into two subgroups, we were able to see that most of the age difference was due to the extended refusers. This is probably because at the extended interview level, the target respondent could obtain more information about the survey and

their refusal or cooperation decision was less random than at the screener level.

As stated in Section 1.1, many LOE analysis assumes that converted refusers are more similar to final non-

respondents. If we applied this assumption to our age and cancer variables, we would conclude that the survey under-represented older population and under-estimated the cancer population. Is this a correct conclusion?

Reexamining the results from benchmarking analysis will help answer this question. HINTS estimate for "% ever had cancer" was about 4 percent point higher than NHIS, which suggests that HINTS has overrepresented cancer population (see Table 2).

The assumption also runs counter to the calibration that was done to create the final weights. In Table 6, by comparing the estimates using base weights and final weights, we see that the calibration brought younger age group up and older age group down, which is an indication that the survey overestimated the proportion of people in the older age group, rather than underestimated it. The same is true with the estimate of cancer population.

One hypothesis for this pattern may be related to the way the topic of the survey is presented. At the screener, the survey is presented as a nationwide survey about health issues related to cancer. The mentioning of cancer may be more salient to older individuals, and thus helped bring in more older or even cancer people during refusal conversion. In this sense, converted refusers increased bias of demographics and health related characteristics.

This result poses a methodological question: with all the time and money it incurs, is it actually worth to do refusal conversion? Will refusal conversion in any way help us?

To gain more comprehensive understanding, we compared converted refusers to initial cooperators in terms of communication variables. As shown in Table 7, the initial refusers seem less likely to use Internet and more likely to read newspaper and use health provider as primary source of cancer info. These differences can be attributed mainly to demographic differences described above.

What cannot be explained by the demographic difference is the likelihood for seeking cancer information. Since the converted refusers consisted more older people and cancer patients, we would expect them to be more likely to seek cancer information. On the contrary, we see a negative difference between the converted refusers and the initial cooperators, and the result holds once controlling for age and ever having cancer. This reflect some characteristics of the refusers that are independent of demographic and health variables. We suspect that the estimate for seeking cancer info was biased in an upward direction.

Since the major interest of the HINTS survey was to measure health or cancer related information seeking, we suspect that the converted refusers were more similar to the final refusers in this aspect, so the refusal conversion effort has helped reduce nonresponse bias in key survey measures.

Table 6.       Calibration results

| Overall estimates | Using base weights | Using final adjusted weights |
|---|---|---|
| % age 18 – 34 | 22.6 | 31 |
| % age 65+ | 22.3 | 16.1 |

Table 7.       Results of LOE analysis – effort to covert initial refusers: communication variables[*]

| | Initial cooperators | Converted refusers | Converted refusers – initial cooperators |
|---|---|---|---|
| % read newspaper everyday | 38.8 | 43.0 | 4.2 |
| % use internet | 61.6 | 57.1 | -4.5 |
| % use health provider as primary source of cancer information | 54.5 | 59.9 | 5.4 |
| % seek cancer information by self | 54.2 | 47.8 | -6.4 |

* All the numerical values in the table are in percentage points.

## 5. Conclusion

In conclusion, HINTS 2005 has over-represented (1) those most concerned about their health, including elderly and less healthy (more likely to have cancer; negative self-assessment) – refusal conversion tends to increase this bias; and (2) those most likely to seek information about cancer – refusal conversion tends to reduce this bias.

Table 8 is a summary of comparison between LOE and benchmarking methods. Important differences are

found in the results between the two approaches. Since the common assumption behind the LOE model does not apply to some measures of the survey, we could have drawn a wrong conclusion in nonresponse bias if we had used LOE as the single approach.

The lessons we learned through this study is that LOE methods can provide useful information on how survey operations affect sample composition. However, to assess bias, it is important to use LOE in conjunction with other indicators of the non-respondent population such as calibration and benchmarking to other surveys.

Table 8.       Comparison between benchmarking method and LOE analysis

| Examples of common variables on between HINTS and NHIS | Benchmarking: significant difference between HINTS and NHIS? | LOE analysis: p-value for difference between two groups | Two methods reaching consistent conclusions? |
|---|---|---|---|
| % in good or excellent health | Y | 0.99 | N |
| % feeling sad some, most, or all of the time | Y | 0.85 | N |
| % feeling nervous some, most, or all the time | Y | 0.44 | N |
| % smoked 100+ cigarettes | Y | 0.54 | N |
| % ever had cancer | Y | 0.07 | Y/N |
| % having health insurance | N | 0.02 | N |
| % never visited a doctor in the past 12 months | Y | 0.28 | N |

## References

Biemer, P. and Lyberg, L. (2003). *Introduction to Survey Quality*. New York: Wiley.

Groves, R.M., Singer, E., and Corning, A. (2000). Leverage-Saliency Theory of Survey Participation. *Public Opinion Quarterly*, 64, 299-308.

Lin, I-F., and Schaeffer, N. (1995). Using Survey Participants to Estimate the Impact of Nonparticipation. *Public Opinion Quarterly*, 59: 236–58.

Olson, K. (2006). Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias. *Public Opinion Quarterly*, 70, 737-758.