A Non-Response Bias Analysis to Inform the Use of Incentives in Multistage RDD Telephone Surveys

Benjamin Skalland, NORC
Kirk Wolter, NORC
Hee-Choon Shin, NORC
Stephen J. Blumberg, NCHS

Abstract

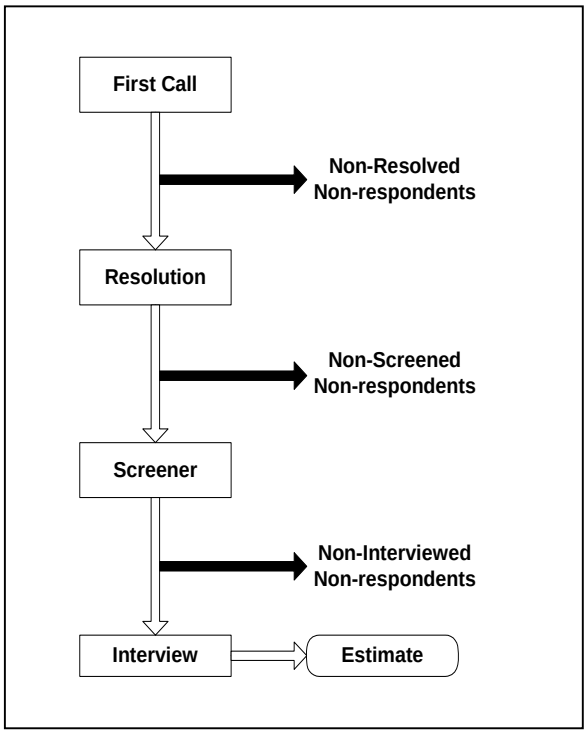
To raise telephone survey response rates, often the level of effort devoted to data collection is increased, either through the use of incentives or through an increase in the number of callback attempts. In a multistage random-digit-dial (RDD) telephone survey, this increase in effort may be implemented at different stages of the interview - e.g. before resolution, after resolution but before screening, or after screening. In this paper, we discuss how a non-response bias analysis can be used to identify the stage of the interview at which the increased effort might have the greatest effect on bias reduction. We also discuss the assumptions necessary for this approach and their plausibility. An example is given using data from round two of the National Survey of Children with Special Health Care Needs (NS-CSHCN), a multistage RDD survey conducted as part of the State and Local Area Integrated Telephone Survey (SLAITS) mechanism. This survey is conducted by the National Opinion Research Center on behalf of the National Center for Health Statistics.

Keywords: Survey non-response, Bias, Incentives, NS-CSHCN

1. Sources of Bias in Multistage RDD Surveys

Figure 1 shows a schematic of a typical multistage, random-digit-dial telephone survey. In order for an interview to be completed, a case must pass through three stages. First, the telephone number must be resolved; that is, it must be determined whether the telephone number belongs to a household, is non-residential, or is a non-working number. Next, if the telephone number is of the appropriate type (for most surveys, this will be a household), the household must be screened; that is, it must be determined whether or not the household is eligible for the interview. Finally, if the household is eligible, it must be interviewed, during which the data are collected that are used to produce the survey estimate. (For simplification, we assume the survey has a single key survey estimate.)

Figure 1: Multistage RDD Survey



If incentives are to be offered or the number of callback attempts increased in such a survey, there is a choice about the stage at which to implement the incentives or to increase the callbacks: it could be done before resolution, after resolution but before the screener, or after the screener. Regardless of which stage is chosen, response rates should increase; however, non-response bias in the estimate will not necessarily be reduced. It is desirable, then, to offer incentives or increase the number of callback attempts at the stage that will lead to the greatest reduction in the bias of the estimate.

The types of non-respondent in a multistage RDD survey are also shown in Figure 1. Non-response occurs when we are unable to determine the type of phone number we are calling (resolution non-response), when we are able to identify a household

but are unable to screen that household (screener non-response) or when we have screened a household as eligible for the survey but are unable to complete an interview (interview non-response).

Non-response bias in the estimate could then be due to one or more of these types of non-respondent. If it were known which type of non-respondent contributed most to the bias in the estimate, that type of non-respondent could be targeted with incentives or increased callback attempts.

2. Determining the Greatest Source of Bias

We cannot directly measure the bias in the estimate because we do not know the response values for the non-respondents. However, bias is often assessed by way of an indirect method, such as a level-of-effort analysis. For some respondents, very little interviewing effort is required before the interview is completed, whereas other respondents require greater interviewing efforts. We will refer to these respondents as "low-effort respondents" and "high-effort respondents," respectively. In a level-of-effort analysis, it is assumed that as the number of necessary callbacks increases, the respondents tend to resemble non-respondents to a greater and greater extent; that is, to use the language of Lin and Schaeffer (1995), it is assumed that there is a "continuum of resistance," with low-effort respondents placed at one end of the continuum and high-effort respondents and non-respondents placed at the other. By comparing the low-effort respondents to the high-effort respondents, then, the non-response bias in the estimate can be measured. Dunkelberg and Day (1973) and Traugott (1987) use this approach to gauge the effect of callbacks on bias reduction.

Using this same idea but applying the analysis separately at each stage of the interview, it is possible to identify the stage or stages at which non-response bias is being generated. That is, for all interviewed households, we can examine the estimate by the number of calls before resolution, the number of calls between resolution and the screener, and the number of calls after the screener. Differences between estimates attributable to low- and to high-effort respondents at a particular stage would then indicate that non-response bias is being generated at that stage, which could then be targeted with incentives or increased callback attempts.

3. Example from the NS-CSHCN II

To illustrate this approach we use preliminary data from round two of the National Survey of Children with Special Health Care Needs (NS-CSHCN), a multistage telephone survey sponsored by the Maternal and Child Health Bureau and conducted by the National Opinion Research Center on behalf of the National Center for Health Statistics. Data collection for this survey began in April 2005 and will continue through 2006. The goal of the survey is to produce national and state estimates of the prevalence and impact of children with special health care needs (CSHCN), and the target population is U.S. households with at least one child less than eighteen years of age.

Many estimates might be produced from the data collected in this survey, but our focus here will be on the household-level prevalence estimate – that is, of households with children, the percentage that have at least one child with special health care needs.

Figure 2 shows how this survey fits into the Figure 1 schematic.[1]

Based on data collected in 2005, the raw national estimate of the household-level prevalence rate is 23.8%. This is a "raw" estimate because, while it has been weighted to reflect the survey design, it has not been adjusted for non-response, nor does it reflect post-stratification. Therefore it is not a final population estimate, and when we discuss the potential for bias in this section and in section 5, we are talking about the potential for bias in the raw estimates and not necessarily in the final population estimates. However, it is desirable to minimize the bias in the raw estimates, because the less biased the raw estimates are, the smaller the subsequent weighting adjustments would

[1] The description of the NS-CSHCN given in this paper is a simplification. For example, once it has been determined that a household has a child with special health care needs, a detailed interview is conducted about one of the CSHCN in the household; therefore the questions identifying special needs serve as a second screener. In this paper, for simplicity, we treat the survey as if there is only one screener that determines whether the household contains any children, and we treat the special-needs screener as if it were the interview. Furthermore, the NS-CSHCN is not a stand-alone survey; before a household is screened for the NS-CSHCN, it is first screened, and possibly interviewed, for the National Immunization Survey. For a full description of the NS-CSHCN, see Blumberg et al., 2003.

need to be, and therefore there would be less variance in the final estimates.

Figures 3 through 6 give the results of applying the level-of-effort analysis discussed in section 2 to the household-level prevalence estimate from the NS-CSHCN. First, a plot of prevalence versus the total number of calls until the interview is given in Figure 3, and shows that the prevalence estimate declines significantly as the total number of calls until the interview increases. Given our assumption that high-effort respondents resemble non-respondents, this indicates that there is upward bias in the raw prevalence estimate.

In figures 4, 5, and 6, we attempt to locate the source of this bias by plotting the prevalence estimate by the number of calls before resolution, the number of calls between resolution and the screener, and the number of calls after the screener. Only figure 5 shows a significant decline in prevalence as the number of calls at the stage increases. Therefore, if our assumptions are correct, we would conclude that the bias in the raw prevalence estimate is due to non-response after resolution but before the screener.

Figure 2: The NS-CSHCN

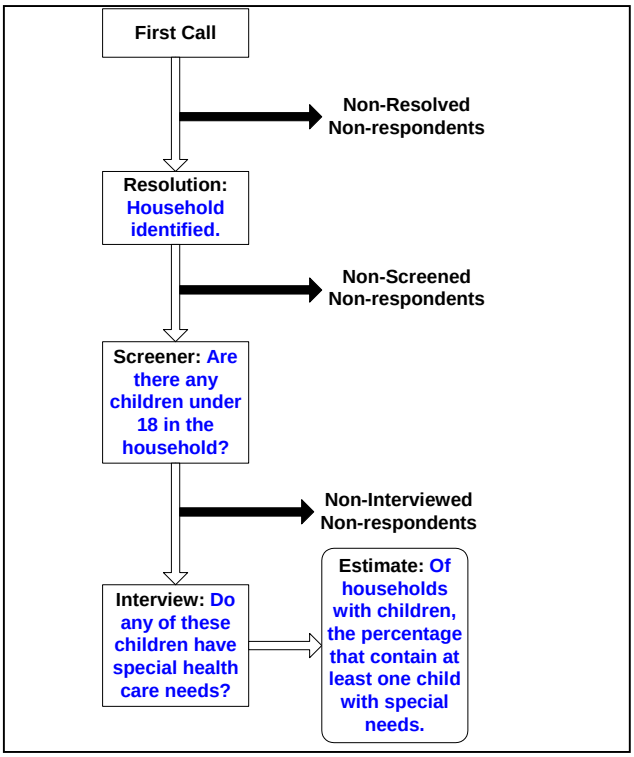


Figure 3: Prevalence by Total Calls

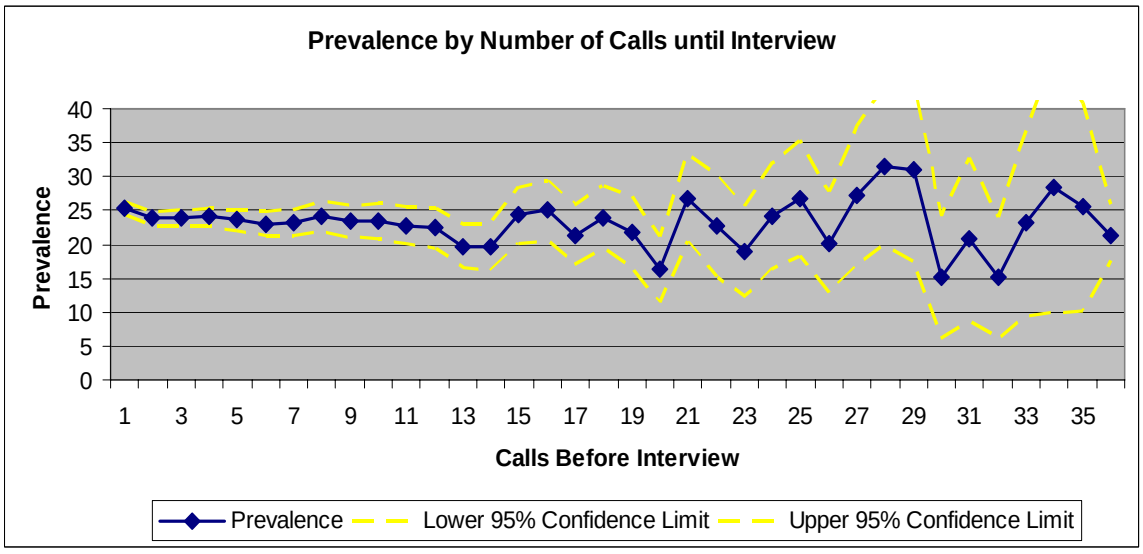


Figure 4: Prevalence by Resolution Calls

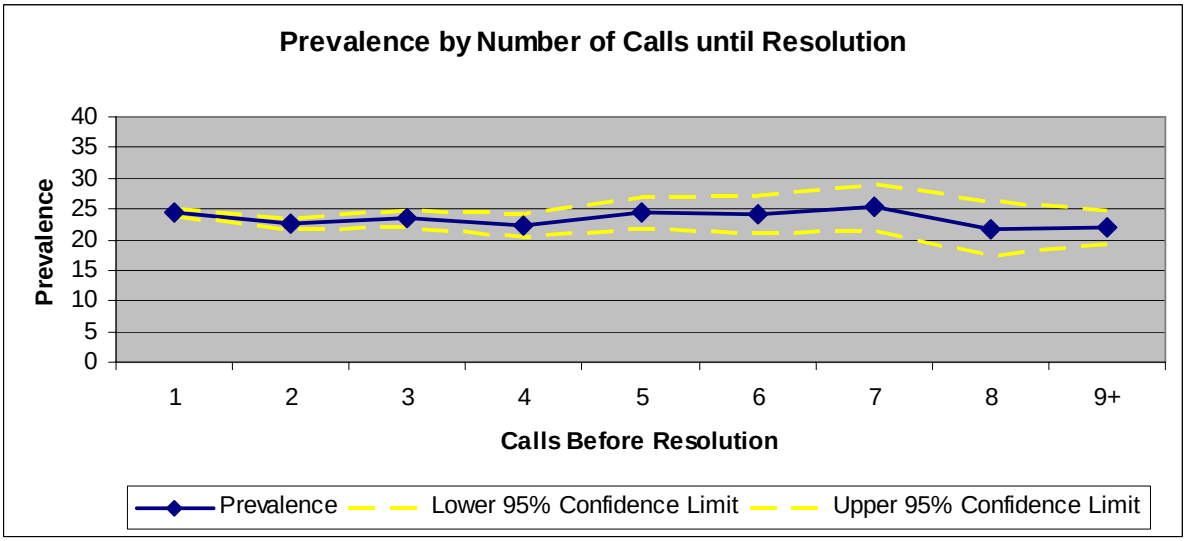


Figure 5: Prevalence by Screening Calls

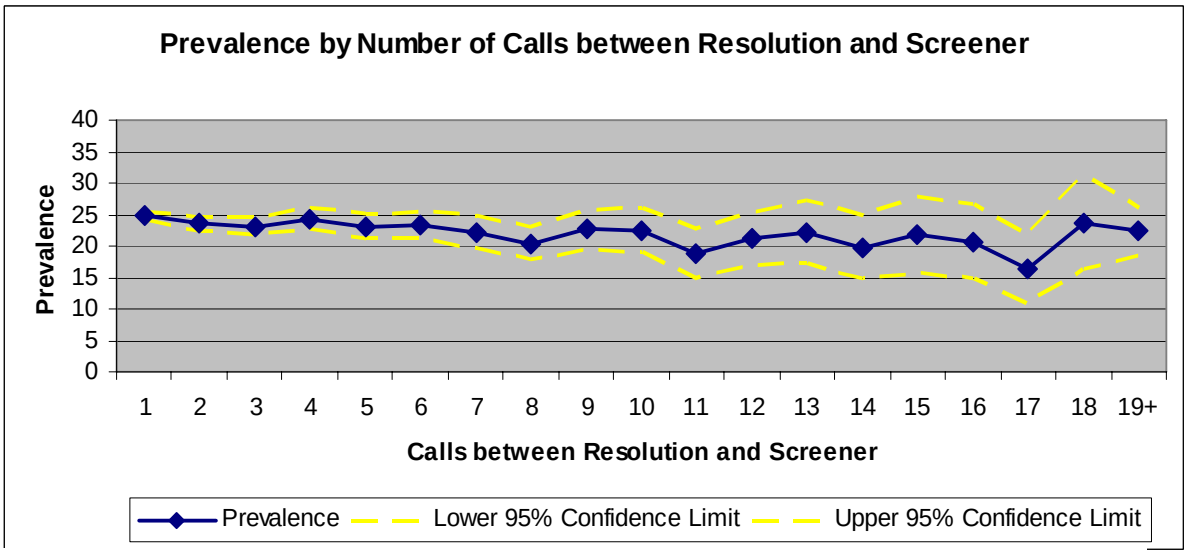


Figure 6: Prevalence by Interviewing Calls

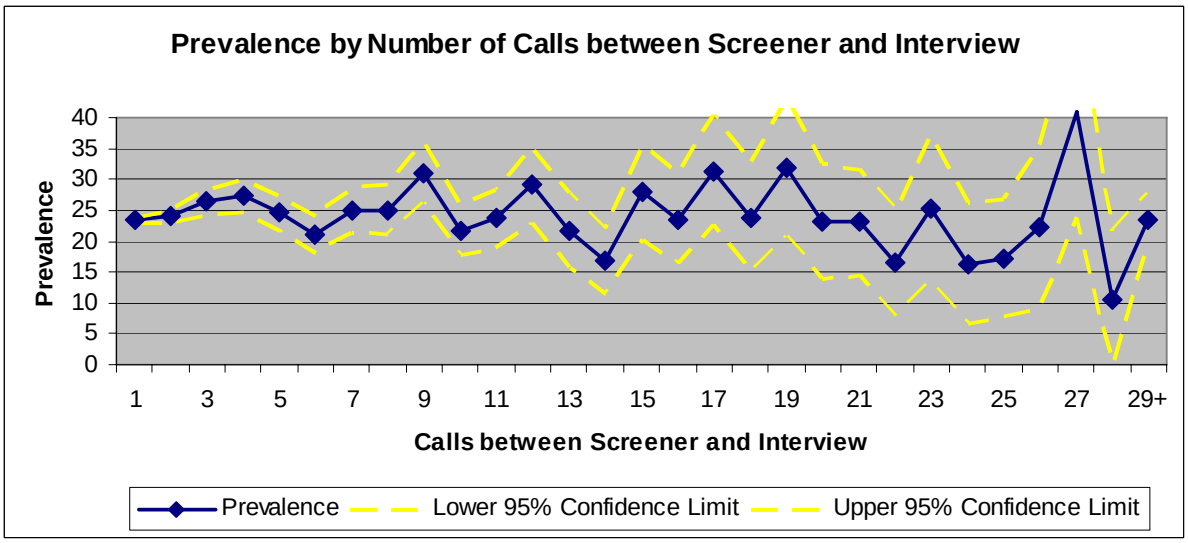


Figure 7: Example Population Under Assumption #1

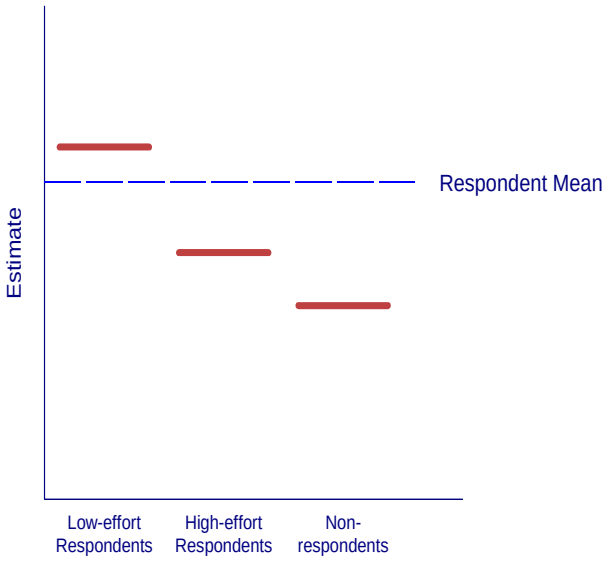
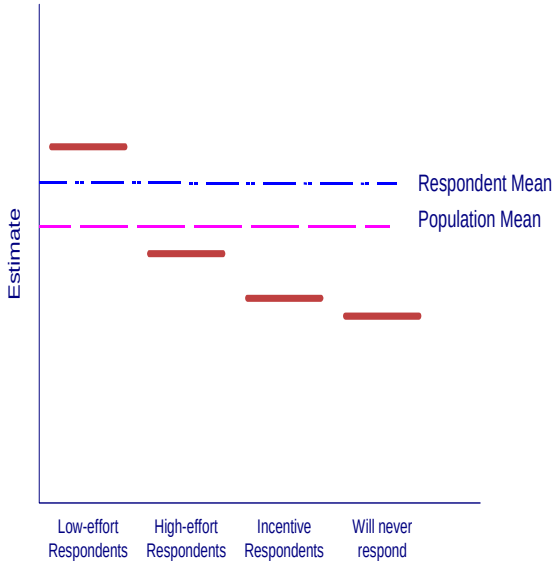


Figure 9: Example Population Under Assumption #3



4. The Assumptions

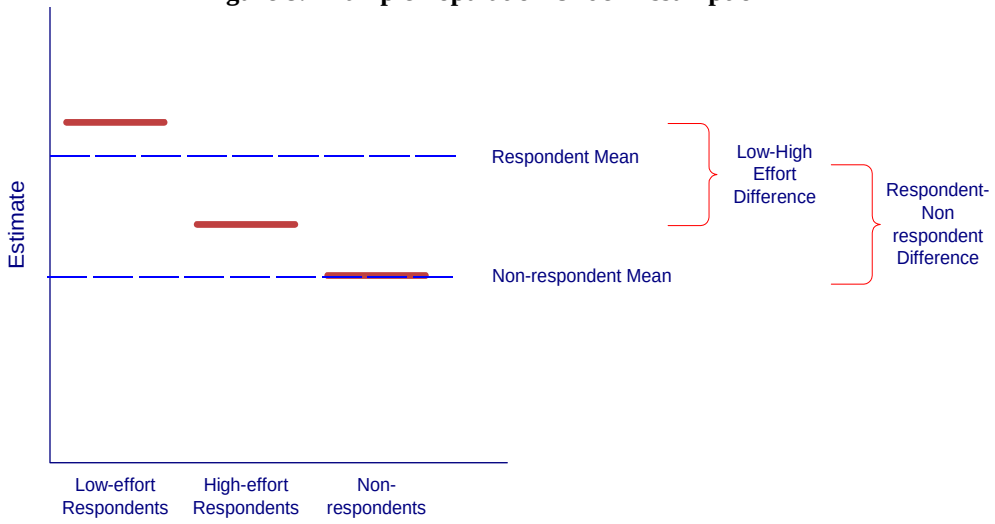
In conducting this analysis, we are making three assumptions:

1. As mentioned in section 2, we are assuming that within each stage, high-effort respondents resemble non-respondents more than do low-effort respondents. That is, respondents who require more call attempts before they complete the stage resemble non-respondents more than do respondents who require fewer call attempts. Figure 7 shows an example of what we expect the population at each stage to look like under this assumption.

2. The second assumption is an extension of the first: we are assuming that at each stage the difference between low- and high-effort respondents is indicative of the difference between all respondents and non-respondents. See figure 8. We need this assumption because, by comparing low- and high-effort respondents at each stage, we are purporting not only to identify the stage or stages at which bias is being generated, but we are also trying to identify the stage that is contributing *most* to the bias in the estimate.

3. We are assuming that if incentives are implemented or the number of callback attempts increased, the new respondents obtained through the added effort will actually reduce the bias. That is, we're assuming that by converting some of the non-respondents into

Figure 8: Example Population Under Assumption #2



respondents, the respondent mean will be pulled closer to the population mean. Figure 9 shows a hypothetical population under this assumption.

5. Testing the Assumptions

In this section, we examine more closely the first and second assumptions. (At the time of this paper, the results of an incentive effort for the NS-CSHCN were not yet available, and so we do not discuss assumption 3 here.) Some previous studies have suggested that assumptions 1 and 2 do not always hold. Lin and Schaeffer (1995), in a telephone study of child-support payments in Wisconsin, found that non-respondents did not resemble high-effort respondents, but instead resembled low-effort respondents. Fitzgerald and Fuller (1982) found that difficult-to-reach respondents generally did not resemble non-respondents in an in-person community survey in Northern California. Because these two studies concerned somewhat specialized populations, and because neither was an RDD multistage telephone survey (and therefore did not have to address the issues of resolution and screening), we felt it would be useful to test our first and second assumptions using data from the NS-CSHCN.

Of course we cannot use the prevalence estimate to test the assumptions because we don't know the prevalence among the non-respondents. However, we do have some information available for both respondents and non-respondents. The sample for this survey was generated by the GENESYS system (Marketing Systems Group, 2005), which supplies, along with each sampled telephone number, auxiliary information specific to the number's telephone exchange, such as the median household income for the telephone exchange and the percent of the population that is Hispanic in the telephone exchange.[2] Since this information is available for both respondents and non-respondents at each stage, we can use it to compare low-effort and high-effort respondents to non-respondents. Note, though, that this information is not collected during the interview itself and is not case-specific. This is frame information about the telephone exchange in which each case is located.

Ideally, we would conduct the same analysis as in section 3, but instead of using the prevalence estimate (which we don't have for non-respondents), we would use the telephone-exchange auxiliary information (which we have for both respondents and non-respondents). That is, for all interviewed households, we would compare low-effort and high-effort respondents at each stage to the non-respondents at the stage. However, "non-respondent" must be defined in terms of the definition of "respondent." If the respondents are defined as all interviewed cases (as they were in section 3), then, by the very fact that they were interviewed, we know that they are households with children. If we are to compare them fairly to non-respondents, then non-respondents would have to be defined in the same way; that is, non-resolved non-respondents would have to be defined as households with children whose telephone number was never resolved, non-screened non-respondents would have to be defined as households with children who were never screened, and non-interviewed non-respondents would have to be defined as households with children who where never interviewed. Yet if the telephone number was never resolved and/or never screened, we have no way of knowing whether the telephone number belongs to a household with children. Therefore, if respondents are defined as all interviewed households, we cannot identify the corresponding non-respondents at the resolution and screener stages.

In testing the assumptions, we therefore define respondents and non-respondents at each stage separately; that is, at the resolution stage, respondents are all resolved telephone numbers, and non-respondents are all non-resolved telephone numbers; at the screening stage, respondents are all screened households, and non-respondents are telephone numbers that have been resolved as households but that have not been screened; and at the interviewing stage, respondents are all interviewed households, and non-respondents are all screened households that were not interviewed. Our test of the assumptions, then, is not a full test of the non-response bias analysis we described in section 2 and carried out in section 3. Nevertheless, in defining non-respondents and respondents differently at each stage, we are still able to test the assumption that high-effort respondents resemble non-respondents within each stage.

In conducting the test of the assumptions, we define low-effort respondents at each stage as those cases completing the stage in fewer than five calls; high-effort respondents are defined as cases completing the stage in five or more calls. Because this choice of five calls as the divide between low- and high-effort respondents is somewhat arbitrary, we also look at the estimated slope of a regression line fitted to a plot of the analysis variable versus the number of calls to complete the stage.

[2] The ultimate source of the GENESYS system's exchange-level information is the tract data from the 2000 U.S. Census.

Table 1 shows, for nine exchange-level analysis variables, the percentage difference between non-respondents and respondents at each stage, the percentage difference between high- and low-effort respondents at each stage, and the slope of the regression line of a plot of the response variable versus the number of calls at the stage. The table also indicates which of the differences and slopes are significant at the 0.05, 0.01, and 0.001 levels.

For 20 of the 27 differences presented in Table 1, the direction of the difference between non-respondents and respondents is the same as the direction of the difference between high- and low-effort respondents (i.e., assumption 1). Moreover, the magnitudes of the percentage differences are similar (i.e., assumption 2); the correlation between the non-respondent/respondent differences and the high-effort/low-effort respondent differences is 0.94. There are instances, however, where the high-effort/low-effort respondent comparison is misleading. For example, there is almost no difference between the high- and low-effort respondents at the resolution stage for the "Percent Hispanic" variable, but respondents have a significantly higher value than do non-respondents.

For 8 of the 9 analysis variables, the stage with the largest percentage difference between high- and low-effort respondents is also the stage with the largest percentage difference between non-respondents and respondents. Only for the "Percent Non-Hispanic Black" variable is this not true: the greatest difference between high- and low-effort respondents is at the interview stage (11.55%), whereas the greatest difference between non-respondents and respondents is at the screener stage (17.72%).

A regression line fitted to a plot of the analysis variable versus the number of calls at each stage is also a fairly good indicator of non-response bias. For 18 of the 27 rows, the sign of the slope of the line matches the sign of the non-respondent/respondent difference, and for 8 of the 9 variables, the stage with the largest slope is also the stage with the largest bias. However, for "Percent Owner-Occupied Homes," the largest slope is at the resolution stage (0.24), and so we would expect non-resolved non-respondents to be the greatest source of bias; however, the largest percentage difference between non-respondents and respondents occurs at the screener stage (-2.65%).

Therefore, our first two assumptions seem to hold fairly well: high-effort respondents tend to resemble non-respondents more than do low-effort respondents, and the magnitude of the difference between high- and low-effort respondents is highly correlated with the magnitude of the difference between all non-respondents and all respondents. However, as Table 1 shows, there are instances where the assumptions do not hold, and our non-response bias analysis could produce misleading results.

6. Summary

If incentives are to be implemented or the number of callback attempts increased in an RDD multistage telephone survey, these increases in effort should be done in a way that reduces non-response bias, if feasible. A comparison of low- and high-effort respondents may be able to indicate the type of non-respondent that contributes most to non-response bias, and this type of non-respondent could then be targeted with increased callbacks or incentives. However, there is evidence that the assumptions necessary for this approach may not always hold and that the results of the analysis could therefore be misleading.

References

Blumberg, Stephen J., Lorayn Olson, Martin Frankel, Larry Osborn, Christopher J. Becker, K. P. Srinath, and Pamela Giambo. 2003. "Design and Operation of the National Survey of Children with Special Health Care Needs, 2001." *Vital and Health Statistics* 1(41).

Dunkelberg, William C., and George S. Day. 1973. "Nonresponse Bias and Callbacks in Sample Surveys." *Journal of Marketing Research* 10: 160-168.

Fitzgerald, Robert, and Linda Fuller. 1982. "I Hear You Knocking but You Can't Come In: The Effects of Reluctant Respondents and Refusers on Sample Survey Estimates." *Sociological Methods and Research* 11: 3-32.

Lin, I-Fen, and Nora Cate Schaeffer. 1995. "Using Survey Participants to Estimate the Impact of Nonparticipation." *Public Opinion Quarterly* 59: 236-258.

Marketing Systems Group. Genesys Sampling Systems Methodology, Philadelphia, PA. MSG 2005.

Traugott, Michael W. 1987. "The Importance of Persistence in Respondent Selection for Preelection Surveys." *Public Opinion Quarterly* 51: 48-57.

Table 1: Comparison of Nonrespondents, Respondents, High-Effort Respondents and Low-Effort Respondents

Analysis Variable[a]	Stage	Non-respondent/ Respondent % Difference[b]		High-effort Respondent/ Low-effort Respondent % Difference[c]		Slope[d]	
Median HH Income	Resolution	3.57%	***[e]	3.83%	***[f]	371.2	***
	Screener	1.06%	***	0.91%	***	55.9	***
	Interview	-0.23%		0.34%		12.6	
Percent Hispanic	Resolution	9.85%	***	-0.08%		0.073	
	Screener	31.02%	***	27.93%	***	0.345	***
	Interview	5.55%	***	8.98%	***	0.029	
Percent Non-Hispanic White	Resolution	-2.47%	***	-0.01%		-0.032	
	Screener	-8.87%	***	-6.76%	***	-0.547	***
	Interview	-2.82%	***	-3.74%	***	-0.111	***
Percent Non-Hispanic Black	Resolution	-0.87%	**	0.52%		0.002	
	Screener	17.72%	***	10.92%	***	0.138	***
	Interview	9.88%	***	11.55%	***	0.067	***
Median Home Value	Resolution	8.79%	***	5.38%	***	1207.6	**
	Screener	4.70%	***	4.74%	***	926.4	***
	Interview	2.77%	***	1.15%		-20.3	
Median Years Education	Resolution	0.46%	***	0.53%	***	0.009	***
	Screener	-0.38%	***	-0.30%	***	-0.003	***
	Interview	-0.23%	***	-0.23%		0.000	
Percent Owner Occupied	Resolution	-1.00%	***	0.94%	***	0.240	*
	Screener	-2.65%	***	-2.52%	***	-0.192	***
	Interview	-0.84%	***	-0.85%	**	-0.014	
Percent College Graduate	Resolution	3.62%	***	3.85%	***	0.139	***
	Screener	-1.27%	***	0.04%		0.007	
	Interview	-1.49%	***	-0.38%		0.004	
Median Age	Resolution	0.24%	***	0.14%	**	-0.008	
	Screener	-1.76%	***	-1.24%	***	-0.051	***
	Interview	0.33%	***	-0.19%		-0.004	

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

a. Telephone-exchange level information.

b. (Non-respondent mean - respondent mean)/respondent mean.

c. (High-effort respondent mean - low-effort respondent mean)/low-effort respondent mean.

d. Slope of a regression line fit to the plot of the analysis variable versus the number of calls at the stage.

e. Significance test of whether non-respondent mean – respondent mean = 0.

f. Significance test of whether high-effort respondent mean – low-effort respondent mean = 0.