



Non-Response Bias: Recent Findings from Address-based Panel Recruitment

Joe Garrett, J. Michael Dennis, and Charles DiSogra

Introduction. Achieving reasonable levels of response to surveys continues to be problematic as response rates have generally declined over the past several years. While this trend is not desirable, it is important to recall that nonresponse bias is not solely a function of nonresponse rate, but actually is a multiplicative function between the nonresponse rate and the difference between respondents and nonrespondents with respect to the survey characteristics being measured.

Recently, Knowledge Networks (KN) started to replace its traditional RDD-based panel recruitment for KnowledgePanel® with address-based sampling (ABS) using the U.S. Postal Service Delivery Sequence File as the source of sample addresses. This transition was done primarily to provide better statistical coverage of cell phone-only households that now represent over 20% of all U.S. households as well as to improve coverage of young adults and minorities.

One research effort, however, arising from KN's strategic shift to ABS recruitment was that KN was able to keep a record of those persons who joined the KN panel and those persons who did not. Using this information, KN could link the address of cases to databases that contained data, either actual or modeled, for address-based characteristics in order to compare invited cases to recruited cases. These comparisons could be viewed as a proxy comparison of the differences between responders and nonresponders

This paper presents the results of the various comparisons that were made between recruited and sampled cases. Results of these recruited/non-recruited comparisons across several demographic and socio-economic variables (as well as Obama/McCain vote tallies) show householders having a similar propensity to join KnowledgePanel. These findings help reviewers of KnowledgePanel survey data to better understand the potential effect on bias due to nonresponse stemming from differences between respondents and nonrespondents at the KN panel recruitment stage.

Bias Due to Nonresponse: The Basic Setting. There are at least two models of nonresponse in sample surveys. One is the stochastic model in which every element in the population has a distinct probability of responding (or not) if selected into the survey. A second and older model is the fixed-response model. This model views the *population* as being composed of two strata – those that would respond if selected and those that would not. The fixed-response model basically assumes that for a fixed set of survey conditions (e.g., survey budget, saliency of the survey content, planned nonresponse follow-up methods, etc), the survey's target population can be viewed as being composed of those people who will respond to the survey if invited and those that will not, and that this effect is *fixed* unless the survey conditions are changed.

One advantage of the fixed-response model is its rather simple form for stating the composition of bias due to nonresponse. In the fixed-response model using simple random sampling, people

in both “strata” (responders and nonresponders) will be selected into the sample, with the nonresponders becoming the sample’s nonrespondents.

In the fixed-effects model, the bias of a survey estimate for the population mean can be expressed as the following:

$$\text{Bias } (\bar{y}_R) = ((N_{NR})/N) \times (\bar{Y}_R - \bar{Y}_{NR}) ,$$

where the terms in this formula are defined as follows:

\bar{y}_R = the estimate of a population mean based on using just the cases responding in the sample

N_{NR} = the total number of *nonresponders* in the population

N = the total number (*responders and nonresponders*) in the population

\bar{Y}_R = the population average (for characteristic *y*) for *responders* in the population

\bar{Y}_{NR} = the population average (for characteristic *y*) for *nonresponders* in the population

Under the fixed-response model and a little algebra, it is easy to see the two components of bias due to nonresponse. However, it is quite often the case that the nonresponse rate is used almost solely as the primary indicator of potential bias due to nonresponse. There are probably two reasons for this. First, the nonresponse rate is part of the formulation of bias due to nonresponse as shown in the above formula. Second, this component is measurable from the sample itself (i.e., using the sample ratio n_{NR}/n as an estimate of population-based ratio N_{NR}/N). Accordingly, it often gets quantified. However, due to its two components and the multiplicative effect implied, assessing data quality on the non-response rate alone is risky because it ignores the second component. Moreover, measuring the second component often requires another survey effort using different methods in order to obtain data for nonresponders in the population so as to be able to make comparisons to the responding sample cases.

It is not the purpose of this paper to argue the relative merits of which model of nonresponse (stochastic or fixed-effects) is more useful, nor for debating which of the two components in the bias formulation above are more important. Rather, it is to share some unique insights on the second component based on Knowledge Networks’ efforts in implementing the ABS frame for recruiting members to its panel. Because of the ABS frame, we have addresses and also know which people joined the panel. So, we can compare characteristics of those who joined the panel with those who were recruited and those who did not join the panel.

KN Comparisons of Recruited Panelists and Non-recruited Panelists using Address-Based Sampling. Comparison data were obtained using information linked to addresses that were sampled and fielded an invitation to join KnowledgePanel in April 2009 (approximately 41,000 addresses). While the primary purpose for KN using ABS was to ensure better statistical coverage of cell phone-only households, the large number of addresses used in ABS permitted us to compare recruited cases to known benchmarks from the Census Bureau and to compare the invited sample cases to the total recruited sample using auxiliary variables linked to addresses from sources such as Experian and InfoUSA. Using actual addresses, we were able to use both actual and modeled data to research the possibility of systemic nonresponse bias associated with the ABS recruiting process. In the sections below we report our findings.

Cell Phone-only Households. As mentioned previously, improving coverage of cell phone-only households was our primary purpose for using ABS. Based on these data, that goal was greatly aided. ABS recruits were 17% of the recruited cases, whereas the National Health Interview Survey was indicating about 20% of U.S. households were cell phone-only households.

Demographics of Recruited Cases. The table below shows the comparison between ABS recruits on an unweighted basis to similar data from the Census Bureau's Current Population Survey during a similar time period. In particular, we were pleased to see successful recruitment at younger age ranges.

Table 1: Comparing Recruited Cases to the Current Population Survey

| Demographics | ABS Recruits (unweighted) | Current Population Survey (June 2009) |
|--------------|------------------------------|--|
| Gender | Percentage | Percentage |
| Male | 42.8 | 48.3 |
| Female | 57.2 | 51.7 |
| Age | | |
| 18-24 | 12.7 | 12.6 |
| 25-34 | 17.9 | 17.8 |
| 35-44 | 19.4 | 18.1 |
| 45-54 | 21.1 | 19.6 |
| 55-64 | 16.7 | 15.3 |
| 65+ | 12.2 | 16.7 |

Race/Ethnicity Comparison. Based on our comparisons, the recruited cases were similar to the invited cases based on race.

Table 2: Comparing Race/Ethnicity between Invited and Recruited Cases

| Race/Ethnicity | Invited Sample Cases | Recruited Cases |
|------------------------|----------------------|-----------------|
| | Percentage | Percentage |
| Caucasian | 75.4 | 74.9 |
| Black/African American | 9.2 | 10.2 |
| Hispanic | 11.2 | 10.9 |
| Other | 4.1 | 4.1 |

Educational Attainment. Based on our comparisons, the recruited cases were similar to the invited cases based on educational attainment.

Table 3: Comparing Educational Attainment between Invited and Recruited Cases

| Educational Attainment | Invited Sample Cases | Recruited Cases |
|------------------------|----------------------|-----------------|
| | Percentage | Percentage |
| Less than high school | 20.3 | 19.9 |
| High school | 31.3 | 29.8 |
| Some college | 25.6 | 26.4 |
| Bachelor | 14.3 | 14.8 |
| Graduate school | 8.5 | 9.1 |

Home Ownership versus Renting. Based on our comparisons, the recruited cases were similar to the invited cases based on owning versus renting.

Table 4: Comparing Home Ownership between Invited and Recruited Cases

| Owning Versus Renting | Invited Sample Cases | Recruited Cases |
|-----------------------|----------------------|-----------------|
| | Percentage | Percentage |
| Owning | 56.6 | 58.9 |
| Renting | 43.4 | 41.1 |

Household Income. Based on our comparisons, the recruited cases were similar to the invited cases based on household income.

Table 5: Comparing Household Income between Invited and Recruited Cases

| Household Income | Invited Sample Cases | Recruited Cases |
|--------------------|----------------------|-----------------|
| | Percentage | Percentage |
| \$1,000-14,999 | 9.8 | 9.2 |
| \$15,000-24,999 | 12.6 | 11.2 |
| \$25,000-34,999 | 11.1 | 10.7 |
| \$35,000-49,999 | 15.5 | 16.1 |
| \$50,000-74,999 | 19.2 | 19.7 |
| \$75,000-99,999 | 12.0 | 12.5 |
| \$100,000-124,999 | 7.4 | 8.8 |
| \$125,000-149,999 | 4.4 | 3.9 |
| \$150,000 and over | 8.0 | 7.9 |

Number of Adults in Household. Based on our comparisons, the recruited cases were similar to the invited cases based on adult household size.

Table 6: Comparing Household Size between Invited and Recruited Cases

| Adult Household Size | Invited Sample Cases | Recruited Cases |
|----------------------|----------------------|-----------------|
| | Percentage | Percentage |
| 1 adult | 53.1 | 49.3 |
| 2 adults | 33.0 | 35.5 |
| 3 adults | 10.1 | 11.6 |
| 4 or more adults | 3.8 | 3.6 |

Region of Residence. Based on our comparisons, the recruited cases were similar to the invited cases based on region of residence (using census regions).

Table 7: Comparing Region of Residence between Invited and Recruited Cases

| Region of Residence | Invited Sample Cases | Recruited Cases |
|---------------------|----------------------|-----------------|
| | Percentage | Percentage |
| Pacific | 15.3 | 13.6 |
| Mountain | 6.8 | 7.0 |
| West-South Central | 11.7 | 11.7 |
| East-South Central | 6.5 | 6.8 |
| South Atlantic | 20.1 | 19.8 |
| West-North Central | 6.9 | 7.9 |
| East-North Central | 15.2 | 17.1 |
| Mid-Atlantic | 12.6 | 11.9 |
| New England | 4.8 | 4.2 |

2008 Presidential Vote. We also decided to look at a variable somewhat apart from the usual demographic and socio-economic variables illustrated above. We looked at data aggregated at the county level by actual vote share for Obama.

Table 8: Comparing Presidential County Vote between Invited and Recruited Cases

| County Vote for Obama | Nonresponders | Recruited Cases |
|------------------------------|----------------------|------------------------|
| | Percentage | Percentage |
| Less than 35% | 11.8 | 12.0 |
| 35% to less than 50% | 29.3 | 31.2 |
| 50% to less than 65% | 39.3 | 39.7 |
| 65% or more | 19.6 | 17.2 |
| Total | 100 | 100 |

Conclusions. Comparing responders and nonresponders using respondent-specific data can be a challenging exercise because sample data may exist for responders, but actual data on nonresponders are often difficult to obtain for comparative purposes. ABS recruitment data provides a basis for a nonresponse bias assessment using available data linked to home addresses. It is these addresses that enable a comparative view of responders and nonresponders, even if using data that are modeled based on the physical address. Thus, one can obtain comparisons between recruited cases and invited cases through these data.

As noted previously, the bias associated with nonresponse is a multiplicative function of the nonresponse rate and the difference between responders and nonresponders. Based on our comparative research using the above variables, ABS-sourced recruitment supports KN panel representativeness on age, gender, race/ethnicity and other core demographic and socioeconomic variables. These comparisons provide evidence that responders and nonresponders in KnowledgePanel may not be all that different, at least for the variables researched in this paper. Although the bias due to nonresponse in any particular survey will be a function of the nonresponse rate and the actual differences between nonresponders and responders *for the particular survey items in that survey*, the extent to which the variables studied here are reasonably well correlated with survey content variables across a wide variety of subject areas, these results are useful and supportive of KN’s panel representativeness.