An Evaluation of Nonresponse Bias in Internet Surveys Conducted Using the Knowledge Networks Panel

Vicki J. Huggins, Michael J. Dennis and Katya Seryakova
Knowledge Networks
1360 Willow Road, Menlo Park, CA 94025

KEY WORDS: Internet Surveys, Nonresponse Bias

Abstract

Bias due to nonresponse in survey estimates is a function of the level of nonresponse and the magnitude of differences between nonrespondents and respondents to key questions of interest. In construction of the Knowledge Networks Panel, a great deal of data is collected to profile the demographic, economic and behavioral characteristics of panel members. Some elements of profile data are available for up to 94% of panel members. Samples for Internet surveys are selected from the panel and suffer from differing levels of nonresponse – 20% up to 65% depending on the survey topic, length of time in the field and other factors. The richness of the profile data on Knowledge Networks panel members allows us to evaluate extensively whether nonrespondents and respondents are different and whether survey outcomes are seriously affected by nonresponse bias.

I. Introduction

Nonresponse exists in practically any sample survey and needs to be understood, documented, and attenuated in the final study results to the extent possible. Otherwise, study conclusions may be distorted. In most cases, very little is known about nonrespondents, so it is difficult to assess the impact of their missingness on final data analyses. In construction of the Knowledge Networks Panel however, a great deal of data is collected to profile the demographic, economic and behavioral characteristics of all panel members that are contacted for Internet surveys. Some elements of profile data are available for up to 94% of panel members. The type of profile data available includes the following:

- **Person and Household Demographics:** Age, race, sex, ethnicity, education level, number in household, employment, income, computer and Internet experience,
- **Computer and Internet Use**: Computer usage for email, type of Internet service, Internet on-line activities, ownership of digital equipment,
- **TV and Cable:** information such as cable or satellite stations received and watched,
- **Health and Ailments**: Includes height, weight, exercise activities, self and doctor diagnosed ailments, pregnancy and menopause
- **Political Profile:** Political behavior, party affiliation, religious affiliation,
- **Magazine and Newspaper Readership**,
- **Financial Profile**: Includes detailed asset information,
- **Lifestyle Profile**: Data on games and sports, gardening, reading, participation in civic organizations, etc.

Samples selected from the Knowledge Networks panel for Internet surveys suffer from different levels of nonresponse – 20% up to 65% depending on the survey topic, length of time in the field and other factors. The availability of the profile data collected on panel members allows us to examine the differences between nonrespondents and respondents and assess whether final analyses and study conclusions are affected by bias due to nonresponse.

This paper will provide a brief description of the sample design of the Knowledge Networks panel, provide a brief synopsis about the difference between respondents and nonrespondents for several studies using profile data, and determine whether nonresponse bias would have significantly affected outcome distributions.

II. Overview of KN Panel Design and Sample Weighting for Individual Surveys

Successfully targeting a nationally representative panel sample over the Internet has been intractable, primarily because a large proportion of U.S. households do not have Internet access. An approach created and implemented at Knowledge Networks overcomes this inherent shortcoming. The methodology begins with selection of a representative sample of households using RDD telephone methods. By phone, the sampled households are asked to participate in the Knowledge Networks research panel sample. Once recruited, the households are then equipped with simple Internet access devices attached to their televisions that are used to field multi-media based surveys. To improve the efficiency of sampling, panel members are sent profile surveys that collect information on their demographic, economic, political and social characteristics. Once panel members complete the core profile survey, they are available for assignment to specific surveys according to specified sampling criteria.

The sample design for the Knowledge Networks Panel Sample begins as an equal probability design that is self-weighting with several known deviations from this guiding principle to make the sample more flexible and efficient. Adjustments are calculated and applied to base sampling weights to account for these known

deviations. There are also several other sources of survey error that are an inherent part of any survey process such as nonresponse, non-coverage and response error. We address these sources of sampling and nonsampling survey error using multiple adjustments to the weights, which we describe below.

More detail about the differential sampling in the Knowledge Networks panel and the effects on the mean square error of key estimates can be found in the paper titled by Huggins and Tang, 2002.

B. Preparation of Final Weights for Individual Internet Studies

Once the samples are drawn, assigned, and the data returned from the field, we subject the final respondent data to a poststratification process to adjust for variable nonresponse and noncoverage. Demographic distributions from the most recent Current Population Survey data are used as benchmarks in this adjustment. A separate nonresponse adjustment to reduce the effects of differential nonresponse for the individual survey is applied on a survey-by-survey basis prior to poststratification to independent benchmarks.

The purpose of implementing a separate nonresponse adjustment on sample weights of completed cases is to reduce bias associated with the fact that nonresponders to the survey may have differenet characteristics than responders to the survey. Nonresponse adjustment is implemented using data known about those initially selected to receive the survey.

The final steps to produce the weights include an examination of the distribution of the final weights to identify outliers, truncation of outliers at the tails, and a ratio adjustment of the weights back to the completed sample size.

The weight after all adjustments -- panel design, nonresponse, post-stratification, trimming, and scaling -- is called the final survey post-stratification weight.

Each profile survey is treated like an individual Internet survey and final weighting procedures are applied to each. The final weight for profile surveys is called the profile weight.

III. Research Plan for Evaluating Nonresponse Bias on Survey Outcomes

We selected data from 3 different surveys conducted using the Knowledge Networks Panel for the nonresponse investigation: a Health Study, a Computer Use Study and an Investment Study. We examined the data at several points of fielding, looking at results when the response rate was at 30%, 50% and 70%:

Figure 1 below summarizes the sample universe, sample design and final response rates for the 3 studies examined in this paper:

Figure 1. Summary of Sample Criteria

Study Name	Sample Universe	Sample Size	Final Response Rate
Health	21+	12,868	69%
Computer Use	18+ w/ specific job functions & use office software	30,527	77%
Investments	18+	2,370	84%

For each of the three surveys, we retrieved the profile data, prepared weighted demographic and economic distributions for the respondent and nonrespondent subgroups to the Field Survey. In particular, we looked at age, race, sex, ethnicity, income, education and region level estimates. We also prepared the same distributions for the combined group of respondents and nonrespondents. The goal here was to summarize the demographic differences between respondents and nonrespondents and the impact of the differences at varying nonresponse levels.

In the second step we identified profile data that is highly correlated with the survey's key outcomes. For example, we used "Do you have access to a computer at home?", "Do you have access to a computer at work?", "Do you have access to a computer at another place?" as the highly correlated profile questions with questions asked in the fielded Internet survey.

The third step consisted of comparing the differences in outcome distributions of respondents, nonrespondents and respondents combined with nonrespondents using the profile information highly correlated with the survey outcomes. We examined the data for different field cut-offs dates, which provides data sets at different levels of response.

The fourth step was a simulation where we used the final survey weight for the individual study and estimated the distributions for the correlated profile outcome variables. We did this to see how well weighting adjustments compensated for sampling and nonsampling error.

IV. Results

Results from the analyses presented below on nonresponse error in Knowledge Networks Internet surveys are consistent with assumptions and findings in most other nationally representative sample panels. Nonrespondents are certainly different than respondents with respect to demographic distributions. And the existence and degree of nonresponse bias depends greatly on the nonresponse levels and the key characterictics under examination. We describe the results in detail below.

A. Demographic Differences Between Respondents and Nonrespondents

We examined demographic distributions for members assigned to the Health Study. The estimates are calculated using the profile weights of all assigned members and the profile data we have on all members. We tabulated the profile data for only Respondents to the Health Survey and for Respondents and Nonrespondents together. The response rate to the study was 69%. What is clear and consistent in the results with what is known about nonresponse in general is that it skews to younger people, Blacks, Hispanics, people with lower incomes and those with lower education levels. With that scenario, its no surprise that nonresponse is lower in the Midwest region of the U.S. as compared to the other three regions.

Similar analyses were conducted for the Computer Use Study. The response rate for the computer use Study was 77% and recall that the study was for the 21+ population that works in certain industries and uses office software. The same skewness to nonresponse patterns exist though: Nonrespondents are more likely to be young, Black and less educated.

For the Asset Study, the response rate at the time the data were aggregated was 56% and the population is 18+. Only one difference between Respondents and Nonrespondents is statistically different (age 65-74), but the differences in the distributions between respondents and nonrespondents are in the same directions as the other two studies.

B. Differences in Outcome Statistics

With the profile data collected on all panel members, we can simulate the effect of nonresponse on field survey data that is highly correlated with the profile data. Table 1 present health outcomes by response group. The table was prepared using data from different field cut-offs where the response rate was at 28% (2 days), when response was at 50% (5 days) and finally when response was at 69% (7 weeks). Significant differences between responders only and responders and nonresponders jointly are highlighted in the tables in bold. The key question we would like to answer is whether having responses from nonrespondents would statistically affect the outcome measures.

Comparing columns 2 where the response rate was 28% and 8 in table 1, we see that estimates for the general state of health are not statistically different between respondents only and respondents and nonrespondents jointly. Statistical differences do show up for reports of Hypertension and Heart Problem/Disease.

With the response rate at 50% for 5 days of fielding, we observe the same statistically significant differences for Hypertension and Heart Problem/Disease. We also find a significant difference for reporting Cancer.

With the response rate at 69%, the statistical differences disappear except for report of Hypertension.

Table 2 reports on the outcomes from the Computer Use Study with a response rate of 77%. No statistical differences are found when comparing Responders only to Responders and Nonresponders together.

Table 3 reports Investment outcomes from the profile data for the respondents and nonrespondents in the Asset Study. The response rate was at 56% when the analysis was conducted. No statistical differences are observed when comparing Responders only to Responders and Nonresponders together.

C. Impact of Weighting to Reduce Bias on Outcome Estimates

After the each Internet survey closes, we apply final weighting adjustments as described earlier to reduce sampling and nonsampling error in the outcome estimates. Table 4 presents the outcomes for the Health Study using the final survey weight for the Responders column and the profile weight for the joint Responders and Nonresponders columns.

For the Health study data, we did not apply a separate nonresponse adjustment, only a post-stratification adjustment to CPS population benchmarks. Looking at data for the Health study in table 4, we find that the statistical differences we noted for hypertension between Responders and joint Responders and Nonresponders in table 1 still exist after final post-stratification weighting adjustments are applied to respondent data for the survey. There is also a statistical difference for the estimate on depression between respondents and joint Respondents and

Nonrespondents. The other estimates: General state of health, estimates for diabetes, cancer and Heart Problem/Disease are not statistically different.

These results suggest that sample weighting was not as effective as desired for the total sample. To investigate the statistical differences identified from table 4 in more detail, we evaluated the data for several subpopulations: Men/Women, 18-34/35+, Black/NonBlack. The goal is to determine if the sample weighting was effective for subpopulations or whether the overall sample results was consistent for all subpopulations. Our analyses suggest that sample weighting was effective for women; persons aged 18-34, and Blacks. However, statistical differences exist for more characteristics for men, persons aged 35+ and NonBlacks. Estimates for men and NonBlacks in general were substantially affected.

We should consider a separate nonresponse adjustment prior to post-stratification adjustment that utilizes the health profile data cross-classified by gender, age and race.

We found comparable results for the computer use study where the final survey post-stratification weight was used to tabulate results for Respondents and the Profile weight was used for the Joint Respondent/Nonrespondent group. It appears that the final post-stratification introduced variability in a several outcomes. We should definitely consider a separate nonresponse adjustment for this dataset as well, taking advantage of more of our profile data.

No significant differences between respondents using the final survey post-stratification weight and the benchmark for joint respondents/nonrespondents were found for the Asset Study.

V. Conclusions and Recommendations

At what level of nonresponse do we see significant bias in profile outcome variables? Response rates around 70-80% appear to be high enough to minimize nonresponse biases for the key characteristics examined in the 3 studies selected for this paper. The studies represent a range of characteristics, from health, income and technology use.

For the health study, the number of statistical differences between responders and joint responders/nonresponders were reduced going from 28% up to 69% response. The absolute differences between responders and joint responders/nonresponders went from 1.2 percentage points, to .93 percentage points to .73 percentage points.

For the Asset study, no significant differences were identified at 32% response, 52% response and 84% response. This is in large part due to the small sample size for the study – 919. Absolute differences between responders and joint responders/nonresponders averaged 3.45 percentage points at 32% response, 1.8 percentage points at 52% response and .36 percentage points at 84% response. The response rate increased by 62% in the last two readings, and the average bias decreased by 80%.

Bias for studies with less than 50% response can be important substantively even if they do not show up as statistically different. In the case of health characteristics for nonBlack men

Which outcome variables are affected the most? The one characteristic that seemed to be most affected by nonresponse bias out of all the characteristics examined is the report of Hypertension. Higher response rates or sample weighting did not mitigate the differences. The subpopulation of nonBlack men appears to be a subpopulation that is very affected by nonresponse bias with multiple health outcomes being significantly different and substantively from the profile benchmarks.

How effective was final survey sample weighting at reducing nonresponse bias? In each of these studies, a separate nonresponse adjustment to profile data was not implemented. Only a post-stratification adjustment to CPS population benchmarks was applied. This one size fits all post-stratification adjustment appears to have a positive impact for characteristics in the asset study. However, it is clear that a separate nonresponse adjustment using profile data for the health study and the computer use study should be considered

In general, separate nonresponse adjustments that take advantage of the profile data should be implemented as part of the final survey weighting methodology. Knowledge Networks has a tremendous amount of information on panel members that can be utilized to improve the mean square error of individual surveys.

References

Couper, Mick P., “Web Surveys: a Review of Issues and Approaches”, Public Opinion Quarterly, 64:464-494, 2000.

Huggins, V. and Tang, X., “ Evaluation of Differential Sampling in the Knowledge Networks panel: What is the Effect on Coverage, Variance and Bias for Key Statistics?”, Proceedings from the Section on Survey Research Methods, American Statistical Association, Forthcoming.

Table 1. Health Outcomes by Subgroup After 2 Days, 5 Days and 7 Weeks of Fielding

	Responders: Profile			Nonresponders: Profile			Responders + Nonresponders: Profile
	2 Days	5 Days	7 Weeks	2 Days	5 Days	7 Weeks	
General state of health							
Excellent	13.9%	13.9%	14.5%	14.2%	14.4%	13.4%	14.1%
Very good	33.8%	35.1%	34.8%	34.8%	34.0%	33.9%	34.5%
Good	34.6%	34.9%	34.5%	34.5%	34.1%	34.4%	34.5%
Fair	14.2%	13.2%	13.2%	13.2%	13.8%	14.2%	13.5%
Poor	3.5%	3.0%	3.0%	3.3%	3.7%	4.0%	3.3%
Hypertension	**24.9%**	**23.8%**	**23.1%**	19.0%	17.5%	15.7%	20.7%
Diabetes	10.5%	10.5%	10.8%	10.3%	10.3%	9.6%	10.4%
Cancer	11.0%	**11.0%**	10.2%	8.5%	7.4%	7.0%	9.2%
Heart Problem/Disease	**20.4%**	**19.0%**	18.2%	14.8%	13.9%	12.7%	16.5%
Depression	9.4%	9.2%	9.7%	9.1%	9.2%	8.2%	9.2%

Statistically significant differences are highlighted in **bold** ($p < 0.05$)
Response rate -- 28% for 2 Days, 50% for 5 days, 69% for 7 Weeks

Table 2. Computer Use by Subgroup

	Responders: Profile	Nonresponders: Profile	Responders + Nonresponders: Profile
Computer access at home	58.1%	62.4%	59.0%
Computer access at work	39.8%	40.6%	40.0%
Computer access at other place	17.5%	19.8%	18.0%
R has no computer access	19.8%	15.5%	18.8%

Statistically significant differences are highlighted in **bold** ($p < 0.05$)
Response rate -- 77%

Table 3. Types of Investments by Subgroup

	Responders: Profile	Nonresponders: Profile	Responders + Nonresponders: Profile
Which of the following types of investments, if any, do you have?			
401K or 403B	26.1%	23.6%	25.2%
IRA	32.4%	23.6%	29.1%
Real Estate	17.3%	12.4%	15.4%
None of the above	14.3%	16.0%	14.9%
Individual stocks	26.3%	23.5%	25.2%
U.S. Savings Bonds	15.2%	16.0%	15.5%
Money Market Funds	21.3%	17.0%	19.7%
Mutual Funds	26.7%	19.1%	23.9%
Other Investments	4.9%	5.1%	5.0%
Respondent has no investments	31.7%	42.7%	35.8%

Response rate 56%

Statistically significant differences are highlighted in **bold** ($p < 0.05$)

Table 4. Health Outcomes by Subgroup After 7 Weeks of Fielding, Weighted by the Final Survey Weight for Responders

	Subgroup		
	Responders: Profile	**Nonresponders: Profile**	**Responders + Nonresponders: Profile**
General state of health			
Excellent	15.3%	13.4%	14.1%
Very good	35.5%	33.9%	34.5%
Good	33.5%	34.4%	34.5%
Fair	12.9%	14.2%	13.5%
Poor	2.9%	4.0%	3.3%
Hypertension	18.0%	15.7%	20.7%
Diabetes	8.9%	9.6%	10.4%
Cancer	6.8%	7.0%	9.2%
Heart Problem/Disease	12.6%	12.7%	16.5%
Depression	11.1%	8.2%	9.2%

Statistically significant differences are highlighted in **bold** ($p < 0.05$)
Response rate -- 69%