



ANALYSIS OF LATE RESPONDERS TO PROBABILITY-BASED WEB PANEL RECRUITMENT AND PANEL SURVEYS

Yelena Kruse¹, Charles DiSogra², Erlina Hendarwan³, J. Michael Dennis⁴
^{1,2,3,4}Knowledge Networks, 1350 Willow Road, Ste 102, Menlo Park, CA 94025

Abstract

Transitioning from a ten-year legacy of RDD sampling, the probability-based online KnowledgePanel® uses an address-based sample (ABS) frame for mail-based recruitment of panel members. This paper analyzes nonresponse through a statistical analysis of “late responders” at two stages of the panel survey process: panel recruitment and the online surveys themselves. The characteristics of “late responders” provide a statistical, however imperfect, window into the characteristics of non-responders at these two stages. Furthermore, any differences between early and late responders would have implications for nonresponse bias in the event the field period is shortened for either the panel recruitment or the online surveys. First, to examine late responders to the ABS-sourced panel recruitment, we compared full demographic and other profile-based information that had been collected among both early and late responders to the mail-based recruitment. After these mail-based recruitment efforts, non-responders are contacted using a telephone-based recruitment whenever a landline phone number can be matched to a residential address. Second, to examine late responders to KnowledgePanel web surveys themselves, we analyzed the samples for three large-sample online surveys. Preliminary research found that late responders to the mail recruitment and panel surveys are disproportionately non-white ($p\text{-value} < 0.05$) and younger adults ($p\text{-value} < 0.01$). The findings will be useful in designing more effective recruitment and online methods encourage response (such as using a targeted segmentation of messages and incentives) in order to minimize response bias.

Key Words: KnowledgePanel, address-based sampling, non-response

1. Introduction

In this paper we examine the characteristics of early and late respondents during the probability-based panel recruitment stage and during the panel survey-taking stage, in which panel members routinely participate in online surveys.

The paper is organized in the following sections: we will start with a brief literature review of studies examining late or hard-to-reach survey respondents, describe Knowledge Networks' recruitment methodology and our research questions. Next we will describe the design analysis and conclude the paper with a discussion of our findings and their implications for an online panel recruitment and survey participation.

1.1 Brief Literature review

The ever declining response rates have lead many research organizations to increase their refusal conversion efforts in order to get the late respondents (also called hard-to-reach) to complete their surveys, with the assumption that late respondents' characteristics of interest resemble individuals who refuse to take the surveys. The majority of studies comparing late and non-late respondents have been conducted in the Federal surveys using face-to-face, phone or mail modes of data collection (Duhart et al, 2001; Kennickell, 2008; Diffendal, 2001; Bates & Creighton, 2000; Kennickel, 2008). To our knowledge, only a few studies have compared early and late respondents in online surveys (Borg, 2000; Kruse et al, 2009). Several studies with multiple modes of data collection found differences in age, ethnicity, household size, home ownership, and children in the household between the early and late respondents (Duhart, 2001; Bates, 2000; Diffendal, 2001; Kruse, 2009). When examining these difference, one has to take mode into account, because many dissimilarities may dissapear when the mode of data collection is face-to-face (Diffendal, 2001). Besides looking at the demographic characteristics at the person and household level, a number of researchers examined and found differences between early and late survey participants in their responses to key survey questions (Duhart, 2001; Diffendal, 2001; Kennickel, 2008). We review a few of the findings on demographic differences only from mail, phone, and online modes of data collection below.

Age.

In his study, Kennickel found that a higher proportion of 35-44 year olds and a lower proportion of 75 or more year olds were late responders to the Survey of Consumer Finances (Kennickel, 2008). Duhart et all examined four government studies and saw the age difference in all surveys: late interviews were less likely to contain persons aged 55+ (Duhart, 2001). A different study by Diffendal examined late respondents characteristics in the American Community Survey; he too found that 55+ respondents were less likely to be late in the mail and phone modes (Diffendal, 2001). In an online longitudinal survey Kruse et al found a higher proportion of younger age groups among the late respondents (Kruse et al, 2009).

Ethnicity.

Several studies we examined had shown that minority respondents were more likely to be late; there were some mixed results, however. Hispanic origin was found to be a significant predictor of being late in the 3 out of 6 surveys we examined. The majority of studies found that African-Americans were more likely to be late respondents with the

exception of the Survey of Consumer Finances, in which more African-Americans were early respondents (Kennickell, 2008).

Household size.

Four out of five studies that had information about household size showed that there was a relationship between the number of people living in the same household and being a late respondent. Generally, multiple person households were more late compared to households with a single person. The pattern was reversed in the personal visit mode: it was harder to reach respondents in 1-person households (Diffendal, 2001).

Children in household.

An unpublished study done by Knowledge Networks found that it is more difficult to recruit households with kids, and, from another study, they also tend to be late respondents to the online surveys, compared to the households without children (Kruse, 2009). None of the studies we reviewed examined this characteristic in relation to the late respondent status.

In our study, we attempt to compare early and late respondents during the multimodal recruitment to the online Knowledge Networks panel and during the web surveys data collection by examining their demographic characteristics.

1.2 Knowledge Panel recruitment process summary

KnowledgePanel®, created by Knowledge Networks, is an online Non-volunteer panel. Potential panel members are chosen via a probability-based sampling method, using known published sampling frames that cover 99% of the U.S. population. Sampled non-Internet households are provided a laptop computer and free Internet service. KnowledgePanel consists of about 50,000 U.S. residents, age 18 and older, including cell phone-only households and those who are of Hispanic origin. In addition, KnowledgePanel includes approximately 3,000 teens, ages 13 to 17, whose parents or legal guardians have provided consent. The panel size fluctuates because of the addition of panelists from the on-going recruitment and because of voluntary withdrawals and retirements of panelists reaching the end of their panel tenure.

Until recently, KnowledgePanel's probability-based recruitment had been based exclusively with a national RDD frame. In 2009, KN added address-based sample (ABS) frame to first supplement and later replace the RDD frame in response to the growing number of cell phone-only households that are outside of the RDD frame, and in response to declining RDD response rate. ABS involves probability-based sampling of addresses from the U.S. Postal Service's Delivery Sequence File. Randomly sampled addresses are invited to join KnowledgePanel through a series of mailings (in English and Spanish) and by telephone follow-up of non-responders when a telephone number can be matched to the sampled address. Invited households can join the panel by one of several means: by completing and mailing back a paper form in a postage-paid envelope; by calling a toll-free hotline maintained by KN; or by going to a designated KN Web site and completing the recruitment form at the Web site (Knowledge Networks, 2010).

1.3 Research questions

- To characterize early and late respondents at the recruitment stage of an online non-volunteer survey panel.
- To characterize early and late respondents at the survey taking stage of an online non-volunteer survey panel.

2. Data and Methods

Recruitment Stage Analysis

Classification of respondents into early and late responder groups depends on the study design: data collection mode, field period, and refusal conversion efforts. Respondents in our study were recruited via mail, online and phone. To be classified as late respondents in the mail mode, respondents had to respond after 27 or more days since the initial mailing. Respondents in the online and phone modes were classified as late respondents if they were recruited after 25 or more days. Twenty-seven percent of respondents (n=436) were classified as late respondents. The rest, 73 percent, were the early respondents.

To examine characteristics of late/early respondents, we ran a logit model using the recruitment data. The response variable was coded 1, if respondent was identified as late and 0, if respondent was identified as early (i.e., anyone not late) respondent. We included phone match, cell phone only household, region, and recruitment mode as covariates and respondents' age, gender, race/ethnicity, household size, and having internet at home as the independent variables of interest (Table 1).

Table 1. Variables in the recruitment stage analysis.

Variable	Description
Age	18-24, 25-34, 35-54, 55+ (reference group)
Gender	Male as a reference group
Race/Ethnicity	White (reference group), African-American, Hispanic, other racial group
Household size	1-person household (reference group), 2-person household, 3+ person household
Internet in the household	no Internet as a reference group
Match of residential address to landline phone number	no phone match as a reference group
Cell phone only household	not a cell phone only household as a reference group
Region	Northeast, Midwest, South (reference group), West
Recruitment mode	Mail (reference group), online, call-in

Initially we included children in the household as a predictor variable, but due to its high correlation with other independent variables, we removed it from the model. The data for all variables, except the household size, were collected during recruitment. About forty percent of cases had missing household data at the recruitment stage. To avoid imputing the data, we used the responses about the household size from the KnowledgePanel Core Adult profile survey for all respondents. Table 2 describes the model. Controlling for

other variables, 35-54 year old respondents were 1.6 times as likely to be late respondents as those who were 55 or older, but younger respondents were not more likely to be late respondents, compared to the older persons. In addition, minorities (African-American and Hispanics) and respondents living alone were more likely to be late, compared to Whites and those living in 2 or more person households. Gender and having Internet at home were not significant predictors of being late recruits.

Table 2. Probability of Being a Late Respondent during recruitment.

	β	<i>SE</i>	<i>Wald</i>	<i>df</i>	<i>p</i>	<i>Exp</i> (β)
Age (Ref. 55 or older)			10.31	3	0.016	
18-24	0.068	0.249	0.08	1	0.784	1.071
25-34	0.309	0.190	2.66	1	0.103	1.362
35-54	0.471	0.164	8.23	1	0.004	1.601
Gender (Ref. Male)	0.187	0.129	2.10	1	0.147	1.206
Race/ethnicity (Ref. White)			12.62	4	0.013	
African-American (Non-Hispanic)	0.414	0.167	6.12	1	0.013	1.513
Hispanics	0.499	0.186	7.21	1	0.007	1.646
2+Race, (Non-Hispanic)	-0.197	0.475	0.17	1	0.678	0.821
Other, (Non-Hispanic)	0.401	0.265	2.29	1	0.130	1.494
Household size (Ref. 1-person household)			7.28	2	0.026	
2-person household	-0.417	0.162	6.59	1	0.010	0.659
3+ person household	-0.354	0.158	5.04	1	0.025	0.702
Internet household	0.137	0.222	0.38	1	0.536	1.147
Region (Ref. South)			1.19	3	0.755	
Midwest	0.018	0.151	0.01	1	0.907	1.018
Northeast	0.182	0.171	1.13	1	0.289	1.199
West	0.038	0.157	0.06	1	0.809	1.039
Cell phone only household (Ref. not a cell phone only household)	-0.105	0.150	0.49	1	0.484	0.900
Recruitment mode (Ref. Mail)			20.49	2	<.0001	
Call-in	0.945	0.212	19.98	1	<.0001	2.573
Online	0.495	0.167	8.79	1	0.003	1.640
Phone match	-0.126	0.122	1.06	1	0.303	0.882
Constant	-1.812	0.310	34.09	1	<.0001	0.163

Panel surveys Stage Analysis

To examine early and late responders to panel surveys, we chose three KnowledgePanel studies conducted in the fall of 2009 with general population respondents. To be classified as an early responder, a person had to complete the survey within 3 days after the survey started (before the standard reminder email). Late responders were those who completed the survey after 3 days since the survey start. Thirty two percent of respondents (n=1865) were classified as late respondents; the rest 68 percent were early respondents.

To examine characteristics of late/early respondents at the survey taking stage, we ran a logit model using respondents' demographic data. The response variable was coded 1, if respondent was identified as late and 0, if respondent was identified as early respondent. We included cell phone only household as a covariate and respondents' age, gender, race/ethnicity, household size, and having internet at home as the independent variables of interest (Table 3).

Table 3. Variables in the panel survey stage analysis.

Variable	Description
Age	18-24, 25-34, 35-54, 55+ (reference group)
Gender	Male as a reference group
Race/Ethnicity	White (reference group), African-American, Hispanic, other racial group
Household size	1-person household (reference group), 2-person household, 3+ person household
Internet in the household	no Internet as a reference group
Cell phone only household	not a cell phone only household as a reference group

According to the data analysis, younger respondent, minorities, and people living in 3 or more person households were more likely to be late respondents than 55+ year olds, whites, and respondents living alone. In addition, those with internet connection at home were 18 percent less likely to be late respondents than the non-internet households. Gender and being a cell phone only household were not significant predictors of being late at taking surveys.

Table 4. Probability of Being a Late Respondent during survey taking stage.

	β	SE	Wald	df	p	Exp(β)
Age (Ref. 55 or older)			19.77	3	<.001	
18-24	0.351	0.117	8.99	1	0.003	1.420
25-34	0.344	0.098	12.43	1	<.001	1.411
35-54	0.280	0.072	15.10	1	<.001	1.323
Gender (Ref. Male)	0.063	0.057	1.22	1	0.269	1.065
Race/ethnicity (Ref. White)			11.72	4	0.020	1.000
African-American (Non-Hispanic)	0.274	0.101	7.31	1	0.007	1.315
Hispanics	0.225	0.105	4.61	1	0.032	1.253
2+Race, (Non-Hispanic)	0.070	0.171	0.17	1	0.683	1.072
Other, (Non-Hispanic)	-0.097	0.161	0.36	1	0.547	0.907
Household size (Ref. 1-person household)			10.79	2	0.005	
2-person household	0.120	0.086	1.92	1	0.166	1.127
3+ person household	0.280	0.090	9.76	1	0.002	1.323
Internet household	-0.196	0.068	8.27	1	0.004	0.822
Cell phone only household (Ref. not a cell phone only household)	-0.053	0.096	0.31	1	0.576	0.948
Constant	-1.025	0.088	135.33	1	<.0001	0.359

3. Discussion

Similar to other studies, we found that younger age groups are more likely to be late respondents than those 55 years or older. This effect was more pronounced at the survey-taking stage (all groups 18-54 years old were more likely to be late than 55+ year olds) than at the recruitment stage, in which only the 35-54 group differed from the 55+ group in their lateness, while younger respondents (18-34) were no more or less likely to be late respondents than 55+ year olds.

Race/ethnicity was also a significant predictor of being a late respondent at both, recruitment and survey taking stages. While controlling for other variables, we found that minorities (African American and Hispanic) were at least 1.5 times as likely to be late respondents as non-Hispanic whites during the recruitment to Knowledge Networks' panel and at least 1.2 times as likely when taking surveys online.

Household size had a mixed relationship with being a late responder in each stage. Persons living in 2+ people households were less likely to be late respondents than those living in 1-person households during the recruitment stage, but panel members from households with 3 or more people were more likely to be late respondents to the surveys than panel members from 1-person households.

Being an Internet/non-Internet household did not have a significant relationship with being a late responder during recruitment, but made a difference during the survey-taking stage. Those who had Internet connection at the time of joining the Knowledge Networks panel were less likely to be late respondents for panel surveys than those who did not have Internet connection prior to joining the panel.

Finally, gender was not a significant predictor of being a late respondent when controlling for other variables.

To conclude, we hope that these findings demonstrate the necessity to employ effective strategies to encourage early response among minorities and persons under 55 to minimize possible response bias—especially when a short fielding period is required.

References

- Bates, N., & Creighton, K. (2000). The Last Five Percent: What Can We Learn From Difficult/Late Interviews? *Proceedings of the Annual Meeting of the American Statistical Association*.
- Diffendal, G. (2001). The Hard-To-Interview in the American Community Survey. *Proceedings of the Annual Meeting of the American Statistical Association*.
- Duhart, D., Bates, N., Williams, B. T., Diffendal, G., & Chiu, P. L. (2001). *Are Late/Difficult Cases in Demographic Survey Interviews Worth the Effort? A Review of Several Federal Surveys*. Retrieved from <http://www.fcsm.gov/01papers/Duhart.pdf>
- Kennickell, A. (2008). The Bitter End? The Close of the 2007 SCF Field Period. *Proceedings of the Annual Meeting of the American Statistical Association*, (pp. 490-497).
- Knowledge Networks. (2010). *KnowledgePanel® Design Summary*. Knowledge Networks, Inc.
- Kruse, Y., Callegaro, M., Dennis, J. M., Subias, S., Lawrence, M., & Tompson, T. (2009). Panel Conditioning and Attrition in the AP-Yahoo! News Election Panel Study. *64th Annual Conference of the American Association for Public Opinion Research*, (pp. 5742-56). Hollywood, FL.