2017

Within-Household Selection in Mail Surveys: Explicit Questions Are Better Than Cover Letter Instructions

Kristen Olson
University of Nebraska-Lincoln, kolson5@unl.edu

Jolene Smyth
University of Nebraska-Lincoln, jsmyth2@unl.edu

Within-Household Selection in Mail Surveys: Explicit Questions Are Better Than Cover Letter Instructions

Kristen Olson and Jolene D. Smyth

University of Nebraska–Lincoln

Corresponding author — Kristen Olson, University of Nebraska–Lincoln, Department of Sociology, 703 Oldfather Hall, Lincoln, NE 68588-0324, USA. *email* kolson5@unl.edu

Abstract

Randomly selecting a single adult within a household is one of the biggest challenges facing mail surveys. Yet obtaining a probability sample of adults within households is critical to having a probability sample of the US adult population. In this paper, we experimentally test three alternative placements of the within-household selection instructions in the National Health, Wellbeing, and Perspectives study (sample n = 6,000; respondent n = 998): (1) a standard cover letter informing the household to ask the person with the next birthday to complete the survey (control); (2) the control cover letter plus an instruction on the front cover of the questionnaire itself to have the adult with the next birthday complete the survey; and (3) the control cover letter plus an explicit yes/ no question asking whether the individual is the adult in the household who will have the next birthday. Although the version with an explicit question had a two-point decrease in response rates relative to not having any instruction, the explicit question significantly improves selection accuracy relative to the other two designs, yields a sample composition closer to national benchmarks, and does not affect item nonresponse rates. Accurately selected respondents also differ from inaccurately selected respondents on questions related to household tasks. Survey practitioners are encouraged to use active tasks such as explicit questions rather than passive tasks such as embedded instructions as part of the within-household selection process.

Introduction

Selecting one adult randomly from a household is central to obtaining a probability sample of adults in households. Although this is easily

accomplished in interviewer-administered surveys, it is much more difficult to complete successfully in mail surveys. At least 30 percent of all households, and about half of households with at least two adults, end up with an incorrectly selected respondent (Battaglia et al. 2008; Olson and Smyth 2014; Olson, Stange, and Smyth 2014; Stange, Smyth, and Olson 2016). This means that the assumptions underlying a probability sample of adults in mail surveys are violated at rates roughly equivalent to a coin flip.

Within-household selection procedures have the potential to affect total survey error through both coverage (i.e., inaccurate selections) and nonresponse (i.e., people included in the sample by the selection procedure who do not do the survey). For either coverage problems or increased nonresponse rates to bias estimates, those who are inaccurately selected or who do not respond need to be different on the construct of interest from those who should have been selected or should have responded. For survey constructs where household members are expected to have similar responses, we would not expect coverage error due to inaccurate selections. This was the case in work by Stange, Smyth, and Olson (2016), where estimates produced from only correctly selected respondents did not differ from those produced from all respondents. However, for constructs on which household members do vary (e.g., household tasks that differ across household members), we would expect inaccurate selections to lead to biased estimates. In addition, the within-household selection process may discourage some households from responding at all, which has the potential to further bias survey estimates through nonresponse. Finding ways to improve the accuracy of selection of people within households without drastically increasing nonresponse error is thus critical for survey inference and validity of mail surveys.

In this paper, we examine three alternative methods of presenting the within-household selection task to households in a mail survey with the intent of improving accuracy of selection. The three methods vary in the degree of emphasis and active participation in the selection task for the household informant—placing the instructions in the cover letter alone (i.e., a very passive method), adding a sentence to the cover of the questionnaire with the within-household selection instructions (i.e., a moderately passive method), and adding a sentence and an explicit question asking whether the person completing the questionnaire is the person matching the within-household selection instructions (i.e., an active method).

Research on within-household selection in mail surveys has demonstrated that three main mechanisms are at work to undermine accurate selections— confusion about the selection task, concealment of household members, and commitment to being a survey participant (Tourangeau et al. 1997; Martin 1999; Olson and Smyth 2014). One primary issue related to confusion about the task is a separation of the instructions for selecting a

respondent within a household from the questionnaire itself. The instructions are typically placed in the cover letter, not the questionnaire. Thus, moving the placement of the within-household selection instructions to the front of the questionnaire should increase the chance that the household actually sees the instructions and has them accessible when needed, increasing the chance that they are followed.

Another possibility is that households do not have a problem reading the instructions, but may lack the commitment from the correct adult in the household to complete the survey. In this case, another member of the household may complete the survey instead. Asking the survey respondent to actively confirm that they are the correct person reinforces the importance of this part of the survey selection task, compared to passively reading the cover instructions. To the extent that commitment is the primary mechanism driving inaccurate selections, we would expect that an explicit question on the cover of a questionnaire will decrease response rates because it will more directly target less committed (i.e., lower response propensity) individuals in the household, but will increase accuracy of selection because it will discourage response from well-intentioned household members who are not selected. Thus, we may expect a trade-off between coverage and nonresponse errors. In addition, to the extent that these less committed individuals are not as attentive to the task of being a survey respondent, we may expect data quality to suffer as well (e.g., higher rates of item nonresponse). That is, we may see higher item nonresponse rates for the active condition compared to the more passive conditions.

To evaluate these three methods of presenting within-household selection instructions, we examine four outcomes: response rates; composition of the completed sample on demographics and other covariates anticipated to be related to potential inaccuracy; accuracy of selection; and item nonresponse. We also examine the effects of inaccurate selections on survey estimates. Thus, we answer the following five questions:

1. What is the effect of placement of within-household selection instructions on response rates?
2. What is the effect of placement of within-household selection instructions on composition of the respondent pool?
3. What is the effect of placement of within-household selection instructions on accuracy of selection?
4. What is the effect of placement of within-household selection instructions on item nonresponse rates?
5. Do survey estimates differ for accurately and inaccurately selected respondents?

Data and Methods

The National Health, Wellbeing, and Perspectives study (NHWPS) was used to test the effects of placement of within-household selection instructions. The NHWPS used a simple random sample of 6,000 addresses in the United States selected from the Computerized Delivery Sequence File by Survey Sampling International (SSI). The topics of the NHWPS included health and mental health, political and social attitudes, experiences of victimization, and demographics. The mail survey was fielded by the research team and the Bureau of Sociological Research at the University of Nebraska–Lincoln between April 10 and August 12, 2015, with an initial cover letter and questionnaire sent to all households, followed by a reminder postcard and up to two replacement questionnaires to nonrespondents. The overall response rate was 16.7 percent (AAPOR RR1), with 1,002 completed questionnaires.

The next-birthday method was used for selecting adults within a household. The next-birthday method is a quasi-probability selection method that is commonly used in telephone surveys (Gaziano 2005), and increasingly used in mail surveys (Battaglia et al. 2008; Hicks and Cantor 2012; Olson, Stange, and Smyth 2014). Although a true probability selection method like the Kish method would be ideal in theory, we used this quasi-probability method because true probability methods are complex and burdensome and thus are rarely, if ever, used in mail surveys. We used the next-birthday method over other quasi-probability selection methods like the last birthday or oldest/ youngest adult methods because it is a commonly used method and because research indicates little difference across these methods in response rates, sample composition, and selection accuracy (Olson, Stange, and Smyth 2014). Each household was randomly assigned to one of three within-household selection treatments: (1) standard placement of the instructions in the cover letter only (*n* = 2,000); (2) instructions in the cover letter and a sentence with instructions on the front of the questionnaire itself (*n* = 2,000); and (3) instructions in the cover letter and an explicit question asking the survey respondent if they were the person in the household with the next birthday (*n* = 2,000). In all three treatments, the instructions in the cover letter were given in exactly the same place and used the same words: "To assure that we have heard from people of all types, we ask that the adult (age 18 or older) in your household who will have the *next birthday* complete the enclosed survey." An example cover letter is shown in **Figure 1,** and the three questionnaire cover treatments are shown in **Figure 2.**

UNIVERSITY OF
Nebraska
Lincoln

BUREAU OF SOCIOLOGICAL RESEARCH

Department of Sociology
301 Benton Hall
P.O. Box 886102
Lincoln, NE 68588-6102
(402) 472-3672
FAX (402) 472-4568
1-800-480-4549
Email: bosr@unl.edu

<DATE>

<STATE> Resident
<ADDRESS>
<CITY>, <STATE> <ZIP>

Dear <STATE> Resident,

We are writing today to ask for your help with an important nationwide survey about the opinions and well-being of people living in the United States. This survey is being conducted by researchers at the University of Nebraska-Lincoln with the goal of gaining a better understanding of the successes and challenges that people in the US experience today. The only way to understand these circumstances is to hear the experiences and opinions of people like you.

Your household was one of a small portion of US households randomly selected to complete this survey. To assure that we have heard from people of all types, we ask that the adult (age 18 or older) in your household who will have the **next birthday** complete the enclosed survey.

Participation in the survey is voluntary, but will help us understand this important topic. It should only take about 15-20 minutes to complete this survey. Answers will be kept confidential. You might notice that there is a unique identification number on the questionnaire. The unique ID simply allows us to keep track of which households have already completed the questionnaire so that we can manage our mailings appropriately. Answers will never be associated with your address.

If you have any questions or would like more information about this study, please contact me at 1-800-480-4549 or by email at bosr@unl.edu. This study has been approved by the UNL Institutional Review Board (IRB# 20150315200 EP; 402-472-6965).

Thank you in advance for participating in this study. To say thank you, we have included a small token of appreciation. We look forward to receiving your responses and hope you enjoy answering the questionnaire.

Sincerely,

Jolene Smyth

Jolene Smyth
Director, Bureau of Sociological Research
University of Nebraska-Lincoln

University of Nebraska at Lincoln   University of Nebraska at Omaha   University of Nebraska at Kearney   University of Nebraska Medical Center

Figure 1. Example Cover Letter from First Mailing. Across the three within-household selection experimental treatments, all sample members received the same letter, with the exception that mention of the incentive was eliminated from the last paragraph for those also assigned to treatments that did not receive an incentive.
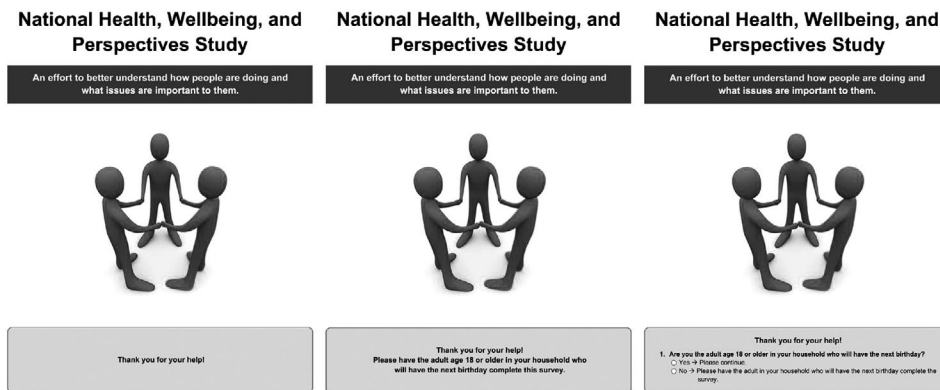
Figure 2. Experimental Questionnaire Cover Treatments for Within- Household Selection Instructions.

The three cover-letter treatments were fully crossed with three randomly assigned incentive treatments: (1) no incentive (*n* = 2,004); (2) a $1 incentive included in the first mailing sent to the household (*n* = 1,998); and (3) a $1 incentive included in the second full mailing sent to the household (*n* = 1,998, the first replacement questionnaire sent to nonrespondents).[1] Four respondents tore the ID number off their questionnaire, and we thus do not know the incentive group to which they were assigned. As a result, they are dropped from all analyses in this paper, and the analyses here are conducted on *n* = 998 responding households. For each set of analyses below, we assessed whether the incentive experiment impacted the results of the within-household selection experiment. Our results (not shown) indicate that the incentive experiment had no impact on the within-household selection experiment. In addition to the within-household instruction and incentive experimental treatments, two versions of the questionnaire were administered to test questionnaire design issues; these were fully crossed with the instructions and cover letter treatments, yielding a full factorial 3×3×2 design. The primary questions used to evaluate within-household selection accuracy were not varied across the two questionnaire forms, although some of the wording for the covariates varied across the two forms.[2]

1. The cover letter shown in Figure 1 is from the treatment with an upfront incentive and thus contains a brief reference to the incentive in the final paragraph. No mention of an incentive was made for those not assigned to receive an incentive.
2. Experimental variation on the wording of the covariates occurred in the trust, suspicion, and privacy questions, but did not affect the univariate distribution for these items (Smyth and Olson 2016). Additionally, the question about children varied between using a full filter and quasi-filter across questionnaire forms, leading to some differences in the number of children (Watanabe, Olson, and Smyth 2015).

The response rate analyses are unweighted; all other analyses are probability of selection weighted. Missing data rates on individual variables used in the analyses ranged from 1 to 18 percent, yielding a listwise deletion rate of over 30 percent in the logistic regression analyses. Missing data in the demographic and other predictor variables was multiply imputed 10 times using sequential regression methods in Stata 13.1 (ice procedure). The fraction of missing information (a relative measure of between imputation variance to total variance; Little and Rubin [2002]) on any of the estimates was less than 8 percent and was generally much smaller.

We begin by evaluating response rates (AAPOR RR1, AAPOR 2016) across the experimental conditions, using chi-square tests to evaluate differences. Then, we evaluate composition of the respondent pool using five demographic characteristics—sex, race/ethnicity, age, education, and presence of children. We compare these five characteristics to the 2014 five-year estimates from the American Community Survey (ACS) as benchmark data (U.S. Census Bureau 2015). We look at each characteristic individually, and summarize the average absolute deviation for the characteristics from the benchmark data. We also examine whether respondents and nonrespondents to each experimental treatment differ based on information available from the frame, including Census division, county size, and Census tract-level race and income information. This frame information was provided by SSI with the frame delivery.

We also examine eight measures collected in the survey that we anticipate being related to confusion, concealment, and commitment (see Appendix A for question wording). As correlates of concealment, we include measures of identity theft, concern over threats to personal privacy, trust in others, and suspicion of others. As correlates of commitment, we include a measure of volunteerism, self-reported likelihood of answering surveys like this one, being the household member who opens the mail, and being the household member who opens the door to strangers. We anticipate that the number of people in the household, age, education, and the presence of children will be related to confusion over the household task (Olson and Smyth 2014; Olson, Stange, and Smyth 2014), and that sex and race/ethnicity will be associated with concealment-related concerns (Tourangeau et al. 1997). Differences across experimental conditions for each factor are examined using survey design-adjusted *F*-tests; deviations from the population data are tested using survey design-adjusted *t*-tests.

Each questionnaire contained a household roster (**Figure 3**) in which the names, relationship to the respondent, birthdate, and sex were provided. This roster information allows us to evaluate accuracy of selection, that is, whether the person completing the questionnaire actually is the person in the household with the next birthday. Not all households completed the

Figure 3. Household Roster.

household roster; 12.9 percent skipped the roster entirely. We are able to evaluate accuracy of selection only among the 869 households that completed the roster. We look at accuracy overall and by the experimental treatments. Then, we predict accuracy of selection in a logistic regression model using our experimental treatments and the proxy variables for confusion, concealment, and commitment. We examine predictors of accuracy overall and for households with two or more adults. Next, we examine whether a data quality outcome—item nonresponse— varies across the experimental treatments. We look at item nonresponse rates across 215 questions asked in the NHWPS that were not included in skip patterns and were not part of the household roster.

Finally, we examine whether survey estimates vary for correctly and incorrectly selected respondents. We examine the theoretically guided proxy variables for confusion, concealment, and commitment, as well as other items for which we anticipate the true values will vary across members in the household— that is, items related to doing household tasks—and thus vary for correctly and incorrectly selected respondents.

Findings

Response Rates

Response rates for each of the experimental treatments are shown in **Table 1**. Response rates across the three selection instruction treatments were not significantly different overall (chi-squared = 3.10, p = .212). The cover-letter-only treatment (AAPOR RR1 = 17.8 percent) had a marginally significantly higher response rate than the verification question treatment (AAPOR RR1 = 15.7%, chi-squared = 3.01, p = .082), with the instruction on the cover treatment being between the two other treatments (AAPOR RR1 = 16.5 percent).

Composition

In reporting the composition of the sample in each of the experimental treatments, we begin with the demographic characteristics that we can compare with ACS benchmark values and with aggregate information available on the frame. We then examine survey estimates that more closely reflect the mechanisms of concealment and commitment that can only be compared across the experimental treatments.

Demographic characteristics: The five demographic characteristics for which data could be obtained from the American Community Survey (**Table 2**)—sex, race, education, age, and presence of children—do not significantly differ across the within-household selection experimental treatments. For example, 65.0 percent of the respondents in the no-instruction treatment were female, compared to 62.6 percent of respondents in the instruction-on-cover treatment, and 58.6 percent of the verification question treatment (F = 1.19, p = 0.31). Additionally, supplemental material presented in the online

Table 1. Response Rates by Within-Household Selection and Incentive Condition

	Number of respondents	Overall RR1
Overall	998	16.6%
No instruction	355	17.8%
Instruction on cover	329	16.5%
Verification question	314	15.7%

All treatments received 2,000 sample cases. Four questionnaires were returned with the ID number removed, and thus cannot be assigned to an experimental treatment. All response rates are calculated treating these four cases as nonrespondents and thus should be viewed as conservative estimates. This is why the overall AAPOR RR1 here differs from that reported in the Data and Methods section of the paper.

Table 2. Demographic Composition, ACS Benchmarks, and Deviations from ACS 2010–2014 Benchmarks

	Percentages for experimental conditions (standard errors in parentheses)												Difference from ACS benchmark data			
	No Instruction (n = 355)		Instruction (n = 329)		Verification question (n = 314)		Overall (n = 998)		F	p-value	ACS 2014		No instruction	Instruction	Verification	Overall
	%	(SE)	%	(SE)	%	(SE)	%	(SE)								
Sex																
Male	35.0	(2.8)	37.4	(2.9)	41.4	(3.1)	37.8	(1.7)	1.19	0.31	48.6		13.6****	11.2****	7.2*	10.8****
Female	65.0	(2.8)	62.6	(2.9)	58.6	(3.1)	62.2	(1.7)			51.4					
Race									1.07	0.38						
White	80.5	(2.3)	80.0	(2.4)	79.2	(2.5)	79.9	(1.4)			62.8		17.7****	17.2****	16.4****	17.1****
Black	7.3	(1.5)	6.0	(1.4)	7.6	(1.6)	7.0	(0.9)			12.2		4.9***	6.2****	4.6***	5.2****
Other/Multiracial	5.1	(1.3)	9.7	(1.8)	7.1	(1.6)	7.2	(0.9)			8.1		3.0*	1.6	1.0	0.9
Hispanic	7.0	(1.5)	4.2	(1.2)	6.1	(1.4)	5.8	(0.8)			16.9		9.9****	12.7****	10.8****	11.1****
Education									1.48	0.18						
Less than HS	3.2	(1.0)	4.0	(1.2)	4.3	(1.4)	3.8	(0.7)			13.8		10.6****	9.8****	9.5****	10.0****
HS graduate	16.0	(2.1)	15.3	(2.2)	16.7	(2.2)	16.0	(1.3)			28.2		12.2****	12.9****	11.5****	12.2****
Some college/AA	34.9	(2.8)	26.0	(2.6)	34.7	(3.0)	31.9	(1.6)			31.3		3.6	5.3*	3.4	0.6
BA+	45.9	(2.9)	54.8	(3.0)	44.3	(3.1)	48.3	(1.7)			26.7		19.2****	28.1****	17.6****	21.6****
Age									0.60	0.78						
18–34	13.3	(2.2)	15.2	(2.5)	12.6	(2.3)	13.7	(1.4)			30.6		17.3****	15.4****	18.0****	16.9****
35–44	11.7	(2.1)	12.0	(2.1)	16.5	(2.4)	13.3	(1.3)			16.9		5.2**	4.9**	0.4	3.6**
45–54	17.1	(2.6)	15.6	(2.4)	13.3	(2.3)	15.4	(1.5)			18.4		1.3	2.8	5.1*	3.0*
55–64	22.4	(2.5)	20.8	(2.6)	23.8	(2.9)	22.3	(1.5)			16.1		6.3**	4.7*	7.7**	6.2****
65+	35.4	(3.0)	36.4	(3.1)	33.8	(3.0)	35.2	(1.8)			18.0		17.4****	18.4****	15.8****	17.2****
HH with children									0.12	0.89						
No	69.8	(2.6)	70.9	(2.7)	71.7	(2.8)	70.7	(1.6)			67.4		2.4	3.5	4.3	3.3*
Yes	30.2	(2.6)	29.1	(2.7)	28.3	(2.8)	29.3	(1.6)			32.6					
Average absolute deviation													9.6	10.3	8.9	9.3

* $p < .05$; ** $p < .01$; *** $p < .001$; **** $p < .0001$

appendix shows that although respondents and nonrespondents differ on the address's Census tract characteristics, these differences did not significantly vary across the experimental conditions.

Moreover, the experimental selection treatments significantly differ from the ACS benchmarks in similar ways. As mail surveys have found (Battaglia et al. 2008; Hicks and Cantor 2012; Olson, Smyth, and Wood 2012; Olson and Smyth 2014; Olson, Stange, and Smyth 2014), the NHWPS overrepresents non-Hispanic Whites (79.9 percent overall compared to 62.8 percent in the ACS), older (35.2 percent age 65+ versus 18 percent in the ACS), more educated (48.3 percent BA+ versus 26.7 percent in the ACS), and female (62.2 percent overall versus 51.4 percent in the ACS) respondents. All treatments are generally representative of households with children. When the average absolute difference between the survey estimates and the ACS benchmark values is calculated, the treatment containing the verification question has the smallest average deviation from the benchmark (8.9 percentage points), followed by the cover-letter-only treatment (9.6 percentage points) and the cover instructions (10.3 percentage points). The verification question thus appears to have the smallest average error relative to the benchmarks.

We also examined eight other survey estimates that more closely represent the mechanisms of commitment and concealment (**Table 3**). No difference emerged across the three experimental treatments for any of these estimates ($p > 0.22$ for all estimates). For example, the average level of concern about identity theft in the no-instruction treatment is 3.10 (where 3 = sometimes and 4 = often; se = 0.06), 3.10 (se = 0.06) in the cover instruction treatment, and 3.07 (se = 0.06) in the verification question ($F = 0.07$; $p = 0.93$). Additionally, the percentage of respondents who report being the household's mail opener is between 75 and 78 percent in all of the treatments ($F = 0.18$, $p = 0.84$). Thus, no evidence exists that any of the treatments yielded significant differences in survey estimates related to commitment or concealment.

Accuracy

We now turn to examining how well the households followed the within-household selection procedure. First, we examine whether households that completed the roster are different from those who did not. There are no significant differences in completing the roster across the within-household selection treatments ($F = 0.18$, $p = 0.84$). In multivariate models predicting roster completion, no differences exist in the rate of completing the roster for male versus female respondents, respondents of different age or race groups, or households with children (results not shown). There are modest differences in completing the roster between respondents with less than

Table 3. Proxy Measures across Experimental Treatments

	No instruction	Instruction	Verification question	Overall	F	p-value
I worry about identity theft						
Mean	3.10	3.10	3.07	3.09	0.07	0.93
Std err	0.06	0.06	0.07	0.04		
Concerned about threats to personal privacy						
Mean	3.16	3.22	3.25	3.20	0.85	0.43
Std err	0.05	0.05	0.05	0.03		
People can be trusted						
Most people can be trusted	64.7%	63.3%	63.0%	63.7%	0.11	0.90
You cannot be too careful in dealing with people	35.3%	36.7%	37.0%	36.3%		
Suspicion of others						
Suspicious of other people	23.5%	29.0%	24.5%	25.6%	1.49	0.23
Open to other people	76.5%	71.0%	75.5%	74.4%		
Likelihood to answer surveys like this one						
Mean	3.35	3.33	3.26	3.32	0.75	0.47
Std err	0.05	0.05	0.05	0.03		
# times volunteered						
Mean	2.87	2.96	2.90	2.91	0.18	0.83
Std err	0.10	0.10	0.10	0.06		
Opening the mail						
No	22.3%	24.2%	24.2%	23.5%	0.18	0.84
Yes	77.7%	75.8%	75.8%	76.5%		
Opening door for strangers						
No	39.8%	36.3%	37.8%	38.0%	0.37	0.69
Yes	60.2%	63.7%	62.2%	62.0%		

All estimates are selection weighted and multiply imputed.

high school education (4 percent missing roster information) versus other education groups (about 15 percent missing roster information); however, less than 4 percent of respondents had less than high school education, reducing the reliability of this estimate. There was also no association between completing the roster and concerns about identity theft, concerns about privacy, feelings of trust or suspicion of other people, or mail or door opening behavior. Respondents who rate themselves as being more likely to complete other surveys like this one are more likely to complete the roster (e.g., 22.1 percent missing for Not at all likely vs. 11.4 percent missing for Very likely, $p < .0001$), but people who volunteer more frequently (e.g., 10.6 percent missing for never vs. 20.8 percent missing for 5+ times volunteers, $p = 0.02$) and persons in larger households (e.g., 3.1 percent missing for one-adult households vs. 20 percent missing for 3+-adult households, $p < .0001$) are less likely to complete the roster.

We now look at the accuracy of selection among households that completed the roster. In the treatment that contained the verification question on the cover of the questionnaire, seven of the 252 respondents indicated that they were not the person in the household with the next birthday. Only six of the seven respondents have complete roster information. Because we do not have this kind of screening for the other experimental treatments, these six cases are retained in the following analyses. None of the findings change when these respondents are excluded from the analyses.

Overall, 67.8 percent of the households correctly selected the household respondent (**Table 4**). When single-person households are excluded from the analysis, 60.4 percent of households accurately selected a respondent. These accuracy rates are similar to those found in previous research in mail surveys (Battaglia et al. 2008; Olson, Stange, and Smyth 2014; Stange, Smyth, and Olson 2016). Looking across the selection instructions, there are no significant differences among all households in accuracy rates ($F = 1.49$, $p = 0.84$) or in households with two or more adults ($F = 1.53$, $p = 0.22$). However, in planned pairwise comparisons across the treatments, selection accuracy

Table 4. Percent Accurate Selections by Instruction Condition

	Overall 2+	*adult HH*
Overall	67.8	60.4
Selection instructions		
No instruction	64.7[a]	56.8[a]
Instruction on cover	66.9	59.2
Verification question	72.0[a]	65.7[a]
F-statistic	1.49	1.53
p-value	0.84	0.22

a. These treatments differ from each other with $p = 0.09$. All analyses account for selection weighting and multiple imputation.

is significantly higher (p = 0.09) with the verification question compared to no instruction at all, both overall and for 2+-adult households.

We now examine multivariate models predicting accuracy of selection with the experimental treatments and proxy measures of concealment, confusion, and commitment (**Table 5**). We examine all households and households with two or more adults because one-adult households can only be selected correctly. The treatment with the verification question is marginally significantly different (p = 0.08) from the treatment with the instructions only in the cover letter—the odds of an accurate selection are 40 percent larger when there is a verification question compared to the cover letter alone. When additional predictor variables are considered, the effect of the verification question increases in statistical significance—the odds of an accurate selection are 59 percent greater when there is a verification question on the cover of the questionnaire compared to instructions in the cover letter alone. There is no difference in accuracy when the instructions are on the questionnaire cover relative to having them in the cover letter alone.

Regarding the proxies for confusion, larger households have less accurate selections, a finding consistent with prior work (Battaglia et al. 2008; Olson and Smyth 2014; Stange, Smyth, and Olson 2016). This holds overall (OR = 0.49, p < .0001) and when the sample is restricted only to households with two or more adults (OR = 0.75, p < .05). Surprisingly, respondents who have some college or an associate's degree are *less* likely to be selected accurately than those who are only high school graduates. Age and selection accuracy are not associated.

For concealment proxies, none of the proxy variables for concealment are significant predictors of accuracy of selection (p > 0.20). That is, concerns about identity theft, privacy concerns, being trustful or suspicious of others, sex, and race all are not statistically associated with accuracy of selection. This holds overall and for households with two or more adults.

Finally, only one proxy variable for commitment significantly predicts accuracy of selection. Respondents who self-report volunteering more frequently are more likely to be selected accurately in all households (OR = 1.18, p < .01) and in households with two or more adults (OR = 1.20, p < .01). The other measures of commitment, including self-reported response propensity and the gatekeeping behaviors, are not associated with the accuracy of selection.

Item Nonresponse

With respect to the single measure of data quality, item nonresponse, if the verification question brings in less motivated respondents, higher item nonresponse rates should exist among respondents in the verification question

Table 5. Odds Ratios and Standard Errors Predicting Accuracy of Selection, All Households and 2+-Adult Households

	All households				2+-adult households			
	Model 1 (n = 869)		Model 2 (n = 869)		Model 1 (n = 557)		Model 2 (n = 557)	
	OR	(SE)	OR	(SE)	OR	(SE)	OR	(SE)
Selection instruction treatments								
No instructions	–		–		–		–	
Instruction on cover	1.08	(0.20)	1.12	(0.23)	1.06	(0.22)	1.09	(0.24)
Verification question on cover	1.40+	(0.27)	1.59*	(0.35)	1.50+	(0.33)	1.59*	(0.37)
Incentive treatments								
No incentive	–		–		–		–	
Pre-paid incentive	1.11	(0.22)	1.13	(0.25)	1.03	(0.23)	1.06	(0.25)
Incentive with reminder	0.99	(0.20)	1.00	(0.23)	0.96	(0.21)	0.98	(0.24)
Confusion proxies								
Number of adults in HH			0.49****	(0.06)			0.75*	(0.11)
Children in HH								
No kids in HH			–				–	
Kids in household			0.85	(0.21)			1.00	(0.26)
Education								
Less than HS			0.98	(0.56)			0.79	(0.44)
HS graduate			–				–	
Some college or AA			0.49*	(0.15)			0.48*	(0.15)
BA+			0.59+	(0.17)			0.68	(0.21)
Age								
18–34			–				–	
35–44			0.62	(0.22)			0.62	(0.22)
45–54			0.95	(0.33)			0.85	(0.30)
55–64			0.90	(0.31)			0.95	(0.33)
65+			0.69	(0.23)			0.69	(0.23)

(continued)

Table 5. Odds Ratios and Standard Errors Predicting Accuracy of Selection, All Households and 2+-Adult Households *(continued)*

	All households				2+-adult households			
	Model 1 (n = 869)		Model 2 (n = 869)		Model 1 (n = 557)		Model 2 (n = 557)	
	OR	(SE)	OR	(SE)	OR	(SE)	OR	(SE)
Concealment proxies								
Worry about identity theft			0.94	(0.08)			0.97	(0.09)
(1 = never, 5 = always)								
Concern about privacy			0.91	(0.10)			0.88	(0.11)
(1 = Not at all,								
4 = very concerned)								
Trust								
Most people can be trusted			–				–	
Cannot be too careful in			1.25	(0.29)			1.22	(0.30)
dealing with people								
Suspicion								
Suspicious of other people			–				–	
Open to other people			0.80	(0.20)			0.76	(0.20)
Sex								
Male			–				–	
Female			1.19	(0.23)			1.13	(0.23)
Race/Ethnicity								
White, non-Hispanic			–				–	
Black, non-Hispanic			1.17	(0.44)			0.92	(0.38)
Other, non-Hispanic			1.21	(0.45)			1.16	(0.43)
Hispanic			1.44	(0.58)			1.34	(0.54)
Commitment proxies								
Likelihood to answer surveys			0.98	(0.11)			0.92	(0.11)
(1=Not at all, 4= very likely)								
Frequency volunteer			1.18**	(0.07)			1.20**	(0.08)
(1 = Never, 5 = 5+ times)								
Mail opener in HH			1.31	(0.28)			1.16	(0.24)
Open door for strangers			1.25	(0.25)			1.00	(0.20)
Model fit								
F-test	0.93		2.62****		0.98		1.07	

All analyses account for selection weighting and multiple imputation.
+ *p* < .10 ; * *p* < .05 ; ** *p* < .01 ; **** *p* < .0001

condition. The overall item-nonresponse rate is 9.2 percent of the questions, and this does not differ across the experimental treatments ($F = 0.22$, $p = 0.81$). In the no-instruction treatment, the item missing data rate is 9.3 percent, compared to 8.9 percent in the cover instructions treatment, and 9.4 percent in the verification question condition. Pairwise comparisons between the treatments also show that the treatments are not statistically different from each other.

Survey Estimates

Table 6 presents estimates for the survey variables used as proxy indicators of confusion, concealment, and commitment for accurately selected and inaccurately selected respondents. Consistent with previous research (Stange, Smyth, and Olson 2016), estimates for being worried about identity theft, concerns about personal privacy, trusting other people, being suspicious of other people, and the self-rated likelihood to answer surveys like this one do not differ between accurately and inaccurately selected respondents ($p > 0.12$ for all of these estimates). Correctly selected respondents report volunteering more often (2.66 times for inaccurately selected vs. 2.92 times for accurately selected respondents, $p = 0.05$). However, for other variables related to prosocial behaviors (e.g., donating money, participating in service organizations, voting), accurately and inaccurately selected respondents do not differ significantly in these survey estimates. There are also no significant differences in self-described party ID or political ideology or general political attitudes about how things are going in the country for accurately versus inaccurately selected respondents (results not shown).

In contrast, estimates related to doing household tasks significantly differ between correctly and incorrectly selected respondents. Accurately selected respondents are more likely to report opening the mail (71.3 percent for inaccurately selected vs. 80.1 percent for accurately selected respondents, $p = 0.01$) and opening the door for strangers (57.5 percent for inaccurately selected vs. 66.7 percent for accurately selected respondents, $p = 0.02$). That is, the odds for reporting opening the mail are 62 percent greater, and for opening the door for strangers are 48 percent greater for accurately selected respondents than for inaccurately selected respondents. This pattern holds for other questions about engaging in household tasks, including grocery shopping, doing household repairs, paying bills, and housekeeping, with the accurately selected respondents more likely to report doing these behaviors (results not shown). Thus, questions that ask about concepts with anticipated heterogeneity across household members significantly differ for accurately selected respondents compared to inaccurately selected respondents, but other survey estimates do not.

Table 6. Survey Estimates for Theoretically Guided Proxy Variables, Estimated on Correctly Selected and Not Correctly Selected Respondents

	Not correctly selected	Correctly selected	t	p-value
I worry about identity theft				
Mean	3.12	3.06	–0.71	0.48
Std err	0.07	0.04		
Concerned about threats to personal privacy				
Mean	.23	3.17	–0.92	0.36
Std err	0.05	0.04		
People can be trusted				
Most people can be trusted	68.3%	62.3%	1.57	0.12
You cannot be too careful in dealing with people	31.7%	37.7%		
Suspicion of others				
Suspicious of other people	22.7%	27.2%	–1.27	0.21
Open to other people	77.3%	72.8%		
Likelihood to answer surveys like this one				
Mean	3.36	3.35	–0.18	0.86
Std err	0.06	0.04		
# times volunteered				
Mean	2.66	2.92	1.93	0.05
Std err	0.12	0.07		
Opening the mail				
No	28.7%	19.9%	2.52	0.01
Yes	71.3%	80.1%		
Opening door for strangers				
No	42.5%	33.3%	2.33	0.02
Yes	57.5%	66.7%		

n = 869; all analyses account for selection weighting and multiple imputation.

Conclusions and Implications

In this paper, we examined three different methods of presenting within-household selection instructions to sampled households: a fully passive display in the cover letter alone, a moderately passive display in the cover letter and a statement on the front of the questionnaire, and a statement in the cover letter plus an active verification question confirming that the sampled person is the correct respondent. Presenting a verification question reduced response rates modestly, but improved the composition of the sample relative to benchmark data, increased the odds of an accurate selection by 59 percent, and had no effect on item nonresponse compared to the more passive treatments. This effect was not modified by the use of incentives, a common motivational tool, despite suggestions in previous studies that incentives may help improve accuracy of selections (e.g., Battaglia et al. 2008).

This study suggests that survey researchers who conduct mail surveys should make the selection task an active part of the survey-taking process. Rather than simply asking respondents to follow instructions for within-household selection in a cover letter or even reinforcing this set of instructions on the cover, the active act of confirming the respondent increased the probability of the household selecting the correct respondent. This came with a modest tradeoff in nonresponse rates, but not with increased nonresponse bias on frame variables. The verification question also yielded an increased following of the within-household selection instructions, and thus an increase in the adherence to (quasi-)probability sampling. Even with the decreased response rate, the sample itself was more representative, reinforcing that higher response rates do not necessarily yield lower nonresponse biases on survey estimates (Groves and Peytcheva 2008) and suggesting an improvement in total survey error for these estimates despite a lower response rate. Of course, despite this modest increase in representativeness, all of the treatments overrepresented female, white, older, and highly educated adults. Moreover, the decrease in response rates likely increased costs per complete, a factor that should be considered.

None of our estimates for correlates of confusion, concealment, and commitment significantly varied over the experimental treatments. Although the verification question yielded respondent pools with fewer highly educated (BA+) and adults aged 65 and above (as indicators of confusion), these proportions are not statistically different from the other treatments. Additionally, we found no evidence that the verification question recruited individuals who were systematically different in their concerns about identity theft or privacy or in commitment to being a survey respondent. With a 16.7 percent response rate, it is possible that people with stronger concerns in these areas failed to participate in this survey at all. Similarly, only

the confusion indicators of number of adults in the household and education level and the commitment indicator of volunteerism were statistically significant predictors of accuracy of selection among these theoretically motivated predictors; none of the concealment proxies predicted the accuracy of selection. That is, most of the variables included as predictors of selection accuracy do not differ between correctly versus incorrectly selected respondents. It is possible that these theoretically derived covariates do not fully reflect the constructs of confusion, concealment, and commitment, and that other measures may be better indicators.

Although most survey estimates examined did not differ between accurately and incorrectly selected respondents, the more notable differences in estimates directly related to activities expected to vary across household members. These differences for estimates of engaging in various types of household tasks across accurately and inaccurately selected individuals are clear indications that getting the selection correct can matter for survey estimates when those estimates are less homogeneous across members of the same household, thus increasing the risk of coverage error. In particular, we would expect that future research will find that concepts measured at the household level (e.g., income, household health insurance) and that members of the same household have similar attitudes toward or experiences with will not vary for accurately versus inaccurately selected respondents, whereas items with strong intrahousehold variation such as participation in household tasks will show much larger differences.

This study examined only one method for asking respondents to actively confirm that they were the correctly selected individual. Other methods could include asking the household to write the number of people in the household and the name of the person who is the selected respondent on the cover (Hicks and Cantor 2012), rather than simply selecting yes or no, or expanding the task to use a household roster for selection, as in the Kish method for interviewer- administered surveys (Kish 1949). These other methods add burden to the respondent task, but could increase attention to the task. The advantage of the method studied here is that the verification question is only one question, and can be added with minimal cost to the survey organization and minimal burden to the respondent. This study is also limited in that it examined only one within-household selection method—the next birthday method. Whether this would hold for other selection methods (e.g., last birthday, oldest/youngest adult; Gaziano [2005]) has not been examined.

Over 12 percent of households failed to complete the roster, and less than 20 percent of households participated. Thus, our understanding of the accuracy of selection is limited to the subset of households that completed the roster, a group that is different on household size, volunteerism, and

self-rated response propensity; household size and volunteerism are also significant predictors of accuracy of selection. This mail survey had a somewhat sensitive topic (mental health), which likely dampened response rates. We do not know whether the effects would differ for a less sensitive topic or a survey with a higher response rate. One additional limitation of this analysis is that these measures of confusion, concealment, and commitment are not available for all members of the household, just the participant. Thus, we cannot evaluate directly whether these mechanisms operate within households—that is, whether within a particular household, the incorrectly selected mail respondent is less confused, has lower concealment concerns, and/or is more committed.

Despite these limitations, this study was the first study of which we are aware to experimentally vary the placement of selection instructions for the next birthday selection method in a national mail survey. The finding that a single verification question on the cover of the questionnaire increased representativeness of the sample and accuracy of selection without harming data quality is important for survey practitioners, even with modest effect sizes. To date, practitioners have had few empirically evaluated tools that increase the rate at which within-household selections are made accurately, thus reducing coverage error. This study has shown that a single question asking if the person completing the survey is the household member who matches the selection criteria can do this (but incentives cannot). It also suggests that future research in this area should examine other ways of including active participation in the selection task for the household, including in other modes. This active participation mimics, in some way, the interaction between the interviewer and respondent in telephone surveys in identifying the person in the household with the next birthday ("Can you tell me the name of the person in the household with the next birthday?"). Even though the verification question improved accuracy of selection, about a third of households with two or more adults still "got it wrong" in this mail survey. Thus, future research into improving within-household selection in mail surveys continues to be needed. Additionally, active verification that the household member is the one with the next birthday could improve the accuracy of selection in a telephone mode, a mode in which respondents are inaccurately selected between 5 and 20 percent of the time (Olson and Smyth 2014). Future work also should investigate how the findings from mail translate to other survey modes.

Appendix A. Question Wording

Sex. Your sex: Male, Female.

Hispanic Ethnicity. Are you Spanish, Hispanic, or Latino? Yes, No

Race: What is your race? (Check all that apply) White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Other, please specify:

Education: What is the highest degree or level of school you have completed? Less than high school, High school or equivalent (GED), Some college but no degree, Associate's degree (AA, AS), Bachelor's degree (BA, BS), Master's Degree (MA, MS, MEng, Med, MSW, MBA), Professional degree beyond a bachelor's degree (MD, DDS, DVM, LLB, JD), Doctorate degree (PhD, EdD)

Age: Your date of birth: MM/DD/YYYY

Presence of children: Not including yourself or your spouse/partner, how many dependents (children or adults) are currently living with you? That is, how many others receive at least one half of their financial support from you? If no dependents in a category, enter "0." Under age 1, Aged 1–5, Aged 6–11, Aged 12–17, Aged 18 or older

Identity theft: I worry about identity theft. Never, Rarely, Sometimes, Often, Always

Threats to personal privacy: How concerned are you about threats to personal privacy in America today? Very concerned, Somewhat concerned, A little concerned, Not at all concerned

Trust: Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Most people can be trusted, You cannot be too careful in dealing with people

Suspicion: In general, would you say that you tend to be suspicious of other people or open to other people? Suspicious of other people, Open to other people

Self-rated response propensity: Please indicate how likely each person you listed in question #48 [household roster] is to answer surveys like this one. You. Very likely, Somewhat likely, Slightly likely, Not at all likely

Volunteerism: In the past 12 months, how many times did you do each of the following? You volunteered at school, church, or another organization. Never, Once, Twice, 3–4 times, 5 or more times

Mail opener: Thinking about the people you listed in question #48 [household roster], who is most likely to do each of the following tasks? Opening the mail. You, Person 2, Person 3, Person 4, Person 5, Person 6

Door opener: Thinking about the people you listed in question #48 [household roster], who is most likely to do each of the following tasks? Opening the door for strangers. You, Person 2, Person 3, Person 4, Person 5, Person 6

Supplementary data are attached to the online record in this database.

References

AAPOR. 2016. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, 9th ed. Oakbrook Terrace, IL: American Association for Public Opinion Research.

Battaglia, Michael P., Michael W. Link, Martin R. Frankel, Larry Osborn, and Ali H. Mokdad. 2008. "An Evaluation of Respondent Selection Methods for Household Mail Surveys." *Public Opinion Quarterly* 72:459–69.

Gaziano, Cecilie. 2005. "Comparative Analysis of Within-Household Respondent Selection Techniques." *Public Opinion Quarterly* 69:124–57.

Groves, Robert M., and Emilia Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72:167–89.

Hicks, Wendy, and David Cantor. 2012. "Evaluating Methods to Select a Respondent for a General Population Mail Survey." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Orlando, FL, USA.

Kish, Leslie. 1949. "A Procedure for Objective Respondent Selection within the Household." *Journal of the American Statistical Association* 44:380–87.

Little, Roderick J. A., and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*. Hoboken, NJ: Wiley.

Martin, Elizabeth. 1999. "Who Knows Who Lives Here? Within-Household Disagreements as a Source of Survey Coverage Error." *Public Opinion Quarterly* 63:220–36.

Olson, Kristen, and Jolene D. Smyth. 2014. "Accuracy of Within-Household Selection in Web and Mail Surveys of the General Population." *Field Methods* 26:56–69.

Olson, Kristen, Jolene D. Smyth, and Heather Wood. 2012. "Does Giving People Their Preferred Survey Mode Actually Increase Survey Participation Rates?" *Public Opinion Quarterly* 76:611–35.

Olson, Kristen, Mathew Stange, and Jolene Smyth. 2014. "Assessing Within-Household Selection Methods in Household Mail Surveys." *Public Opinion Quarterly* 78:656–78.

Smyth, Jolene, and Kristen Olson. 2016. Mismatches. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Austin, TX, USA.

Stange, Mathew, Jolene D. Smyth, and Kristen Olson. 2016. "Using a Calendar and Explanatory Instructions to Aid Within-Household Selection in Mail Surveys." *Field Methods* 28:64–78.

Tourangeau, Roger, Gary Shapiro, Anne Kearney, and Lawrence Ernst. 1997. "Who Lives Here? Survey Undercoverage and Household Roster Questions." *Journal of Official Statistics* 13:1–18.

U.S. Census Bureau. 2015. *2010–2014 American Community Survey 5-Year Esti-mates*. Washington, DC: U.S. Department of Commerce, U.S. Census Bureau.

Watanabe, Megumi, Kristen Olson, and Jolene D. Smyth. 2015. "A Comparison of Full and Quasi-Filter Formats for Behavioral Questions." Paper presented at the Annual Meeting of the Midwest Association for Public Opinion Research, Chicago, IL, USA.