Journal of Official Statistics, Vol. 27, No. 2, 2011, pp. 339–357

Using Paradata and Other Auxiliary Data to Examine Mode Switch Nonresponse in a "Recruit-and-Switch" Telephone Survey

Joseph W. Sakshaug[1] *and Frauke Kreuter*[2]

Minimizing unit nonresponse and maximizing reporting accuracy about sensitive items are common goals among survey practitioners. In order to maximize reporting accuracy without compromising on response rates a common strategy is to recruit respondents over the phone and switch them to a self-administered mode (e.g., IVR, web) for answering the sensitive items. A drawback to the "recruit-and-switch" design is that a substantial portion of the sample (typically 20 percent or more) drop out during the mode switch. Recent evidence suggests that this form of nonresponse can introduce bias and offset gains in accuracy achieved by self-administration. We analyze respondents' likelihood of complying with the mode switch request in a survey of university alumni. Results indicate that paradata derived from the screening interview are related to mode switch participation. Also, we find evidence that adding reluctant sample members into the mode switch respondent pool yields improved estimates with lower nonresponse bias.

Key words: Data collection; sensitive questions; mixed mode; CATI; IVR; web.

1. Introduction

Survey organizations routinely balance tradeoffs between cost and error when choosing a mode of data collection. For example, interviewer-administered modes of data collection (e.g., face-to-face, telephone) are more costly than self-administered modes (e.g., mail, web), but tend to yield higher response rates. On the other hand, self-administered modes tend to elicit more accurate survey reports about sensitive items than interviewer-administered modes (Tourangeau and Smith 1998; Tourangeau et al. 2000; Kreuter et al. 2008). A common compromise is to combine both types of modes into a single multi-mode survey. For example, an interviewer-administered mode is used to contact, recruit, and screen respondents, and, once those tasks are completed, respondents are then switched to a self-administered mode where the sensitive items are presented. This combination is particularly common in face-to-face surveys; prominent examples are the National Study

[1] Institute for Social Research, University of Michigan, 426 Thompson Street, Room 4050, Ann Arbor, MI 48104-1248, U.S.A. Email: joesaks@umich.edu
[2] Institute for Employment Research, Nuremberg/Ludwig-Maximilians-University Munich, Germany and Joint Program in Survey Methodology, University of Maryland, 1218 Lefrak Hall, College Park, MD 20742, U.S.A. Email: fkreuter@survey.umd.edu

Acknowledgments: An earlier version of this article was presented at the 21st International Workshop on Household Survey Nonresponse in Nuremberg, Germany. We thank Paul Beatty and three anonymous reviewers for critical comments and helpful suggestions. We are especially grateful to Katharine Abraham and Mirta Galesic who oversaw the data collection. Students in the JPSM Practicum class provided assistance in the development of the study.

of Family Growth (Groves et al. 2009), the National Health and Nutrition Examination Survey (CDC 2010), and the National Survey on Drug Use and Health (SAMHSA 2009). In computer-assisted telephone interviewing (CATI), a similar mode-switch is implemented when interviewers are tasked with contacting, recruiting, screening, and subsequently switching respondents to an interactive voice response system (IVR) or providing respondents with a URL to access the questionnaire online. Both computerized methods offer a higher degree of privacy relative to CATI alone and have been shown to lessen social desirability effects (Kreuter et al. 2008).

A significant drawback to the "recruit-and-switch" technique is that a substantial portion of the sample drop out during the mode switch. For example, sample members ostensibly agree to complete the IVR portion of the questionnaire, but hang up during the transfer to the IVR system. Hang-ups during the IVR transfers are quite common (typically about 20% or more) and can be exacerbated by long transfer delays (Tourangeau et al. 2002; Couper et al. 2003). An analogous phenomenon occurs with web surveys: members of the telephone sample apparently agree to complete the survey via the Internet, but subsequently never access the online questionnaire (Kreuter et al. 2008; Fricker et al. 2005).

The effects of mode switch nonresponse can mirror those of unit nonresponse. Mode switch nonresponse reduces the effective samples size and can introduce bias if those who carry out the mode switch are systematically different than those who do not. Such bias has been experimentally shown to diminish, and, in some cases, completely offset the measurement advantages of self-administration (Sakshaug et al. 2010). That is, switching respondents from an interviewer-administered mode to a self-administered mode can yield estimates with *more* total bias than estimates obtained from respondents who were not switched to the self-administered mode.

The potential backfiring effect of switching respondents to a self-administered mode in "recruit-and-switch" surveys raises many important practical questions. First, is mode switch nonresponse avoidable? In the extreme case, such nonresponse can be completely avoided if respondents are both recruited and interviewed using a single interviewer-administered mode and not switched to a self-administered mode. For example, one could recruit and interview respondents on the telephone without switching them to IVR or the Internet, the model of current CATI surveys. However, this approach loses the potential advantages of self-administration, and even though there is evidence of offsetting nonresponse and measurement error effects (Sakshaug et al. 2010), it is unclear whether the same pattern would hold for highly sensitive topics (e.g., illicit drug use). An alternative strategy is to offer the mode switch and use design features to reduce the likelihood of nonresponse. For example, strategies shown to be effective in reducing unit nonresponse, such as offering incentives (Singer 2002) or sending reminders for web surveys (Kaplowitz et al. 2004), may be effective in persuading respondents to complete or, at the very least, start the self-administered portion of the interview.

A cost-effective strategy, in the spirit of responsive survey design (Groves and Heeringa 2006), is to preidentify likely mode switch nonrespondents and implement intervention methods to increase their likelihood of complying with the mode switch. For example, respondents with a high propensity of mode switch nonresponse may be precluded from the mode switch step and, instead, administered the questionnaire by an interviewer. Alternatively, preidentified respondents could be offered larger incentives to complete the mode switch. These strategies naturally lead in to the second question: can we identify characteristics and predictors of mode switch nonrespondents?

Identifying characteristics of mode switch nonrespondents requires auxiliary data on sampled members collected prior to the mode switch. Data already available on the sampling frame, call records, or other paradata collected during the interaction between interviewer and respondent are potential candidates for such predictors. Any of these data sources would be available for all sampled units prior to the mode switch. In order to tailor responsive design not just for increased response rates, but to reduce nonresponse bias, these data also need to be related to the survey outcomes.

The survey used here provides data from all three of these sources (frame data, contact data, and paradata derived from the interaction between interviewer and respondent). Having a rich set of data prior to the mode switch as well as an experimentally implemented switch allows us to address three questions:

1. Are contact data predictive of mode switch dropout? Specifically, are sample members who are hard-to-reach and sample members who refuse the initial survey request but subsequently agree more likely to drop out during the mode switch?
2. Can we derive paradata, such as indicators of respondent resistance, from the initial interaction with the respondent during the screening interview that are predictive of mode switch dropout?
3. Is the propensity of complying with the mode switch, based on covariates from call records, frame data, and screener paradata, related to the survey outcomes? Specifically, are survey estimates improved (i.e., is there lower nonresponse bias) if low propensity cases are recruited into the mode switch respondent pool?

With these questions in mind we re-examine data analyzed by Sakshaug et al. (2010) to explore further the effects of mode switch nonresponse. The previous analysis of these data focused on comparing the magnitude and direction of mode switch nonresponse biases with other sources of error (noncontact, refusal, and measurement). Here we focus on examining predictors of mode switch dropout and their relationship with the survey outcomes using three sources of data: call records, paradata from the screening interview, and data from the sampling frame.

2. Methods

The data we use come from a study carried out by the Joint Program in Survey Methodology (JPSM) at the University of Maryland. In 2005, students in the Practicum class designed and pretested a survey of University of Maryland alumni. The main data collection for the survey was conducted by Schulman, Ronca, and Bucavalas, Inc. during the months of August and September. Members of the sample were contacted initially by telephone, asked a brief set of screening questions about their personal and household characteristics (including access to the Internet), and then assigned to one of three methods of data collection for the main interview. The introduction described the survey as sponsored by the University of Maryland and as asking about “your college experience, interest in alumni activities, and community involvement.” The specific wording of the introduction is below:

Hello, my name is [INTERVIEWER's NAME] and I'm calling on behalf of the University of Maryland. You have been randomly selected to participate in a survey of Maryland alumni. I'm not calling to ask you for a donation. I will be asking you about your college experience, interest in alumni activities, and community involvement. Your participation is strictly voluntary and you may skip any question you don't want to answer. All of your responses will be kept confidential. The survey will take about 10 minutes.

More details of the study can be found in Kreuter et al. (2008). Relevant features of the sampling design and data collection are briefly reviewed below. Particular attention is given to the data sources we use to study the likelihood of mode switch nonresponse.

2.1. Sampling and Data Collection

A random sample (stratified by graduation year) of 20,000 graduates was drawn from a population of 55,320 alumni (class of 1989 to 2002). After sample cases were matched with Alumni Association records (to obtain telephone numbers) and various ineligible cases were dropped (e.g., those used in pretesting and those living abroad), the survey fielded 7,535 telephone numbers. Data collection took place during the months of August and September, 2005. More than a third of these telephone numbers turned out to be invalid (e.g., the number was disconnected) and the status of about another quarter could not be determined. A total of 1,501 alumni completed the screening interview and were randomly assigned to a mode of data collection (CATI, IVR, or web). There were 37 cases who reported they did not have Internet access and these were randomly assigned either to CATI or IVR data collection. The response rate (AAPOR Response Rate 1; AAPOR 2008) for the screener was 31.9% (see Appendix Table A1). Most of the nonresponse was due to difficulties in contacting alumni rather than to their unwillingness to cooperate. The refusal rate was about nine percent of the fielded phone numbers (excluding ineligibles). The overall completion rate after initial assignment to the three modes was 66.8%, leading to an overall response rate of 21.3% (AAPOR Response Rate 1 computed by multiplying the screener completion rate of 31.9% by the completion rate after initial assignment of 66.8%). Kreuter et al. (2008) found no strong overall nonresponse bias with respect to the frame data in their analysis of these data.

2.2. Call Record Data

We utilize two sets of paradata (Couper and Lyberg 2005) to study the likelihood of mode switch nonresponse. The first is call record data. Call record data is commonly used to study unit nonresponse (Bates et al. 2008; Kreuter and Kohler 2009), and is particularly useful for identifying sampled persons who are hard-to-reach or who have expressed resistance toward the survey request, all of which are potential barriers to obtaining response (Groves and Couper 1998). In the UMD alumni survey, each call attempt and dispositional outcome was recorded in the CATI system. In our analysis we use the total number of calls needed up to the screener interview as one indicator of resistance. A second resistance indicator is a binary variable indicating whether or not the respondent refused the initial survey request. These passive and active forms of resistance to the survey interview may carry over to the mode switch and yield a decreased likelihood of switching to the new mode. Such a link would indicate overlapping mechanisms, or a common cause, influencing both unit and mode switch nonresponse.

2.3. Screener Paradata

A second set of paradata available for the study of mode switch nonresponse are indicators of resistance derived from the screening interview. Indicators of respondent resistance have been thoroughly studied in surveys as correlates of cooperation. Respondent resistance during the screening interview may signal reluctance to participate in the main interview and, thus, a greater likelihood of mode switch dropout.

Borrowing from the longitudinal survey literature, there is ample evidence suggesting that indicators of resistance derived from paradata in prior waves are predictive of cooperation in subsequent waves. For example, item nonresponse is indicative of having an unpleasant or negative interview experience and is predictive of unit nonresponse in future waves (Loosveldt et al. 2002; Lee et al. 2004; Hawkes and Plewis 2006). Interview duration is another oft-cited indicator of respondent cooperation; longer interview durations are believed to be a product of how enjoyable and engaging the interview experience is to respondents, while shorter durations indicate a less enjoyable experience (Branden et al. 1995; Zabel 1998; Hill and Willis 2001). Other longitudinal evidence has shown that interviewer observations, call records, and other forms of paradata are predictive of whether respondents agree to in-survey requests, such as consent requests for physical measurements (Sakshaug et al. 2010a) and administrative data linkage (Sakshaug et al. 2010b). Adapting these findings to the study of mode switch nonresponse is possible as many of the same cooperation indicators (item nonresponse, elapsed time) are collected during the screening interview. Accordingly, we utilize these indicators to examine mode switch nonresponse.

The screening interview contained the following five questions about basic personal and household characteristics:

1. In what year were you born?
2. Are you now. . .Married, Living with a partner, Widowed, Divorced, Separated, or Never married?
3. Including yourself, how many people live in your home?
4. (IF MORE THAN ONE PERSON IN THE HOUSEHOLD): How many of these people are age 18 and under?
5. Do you have access to the internet for personal use?

Our intent is not to analyze the substantive responses given to these questions, but rather we are interested in 1) whether respondents provided a substantive response to each of these questions, 2) how fast respondents answered this set of questions, and 3) whether 1) and 2) are related to mode switch nonresponse. We constructed a dichotomous variable indicating whether a respondent answered "don't know" or was unwilling to answer at least one of the five questions (We conducted a sensitivity check by trying other categorizations of the item nonresponse variable, including not responding to multiple items, but they did not have a stronger effect on mode switch nonresponse). We constructed an elapsed time variable based on timestamp data collected at the beginning and end of the screening interview. We created two dummy variables based on quartiles of the elapsed time distribution – whether respondents placed in the lowest quartile of the distribution ("fastest," IVR range: 0–87 seconds; web range: 0–111 seconds) and whether respondents placed in the highest quartile of the distribution ("slowest"; IVR range: 135 + seconds; web range: 179 + seconds). The reference groups consist of the middle quartiles. The longer screener response times in the web group reflect additional remarks made to respondents about logging into the web instrument. Another unique feature of the web group is that they were asked to provide their email address to receive the URL to access the questionnaire online. Respondents who did not feel comfortable providing an email address (or did not have one) were read the URL over the phone or this information was faxed to them. We suspected that not having an email address (or unwillingness to provide one) might impact their likelihood of logging into the web instrument, and constructed an indicator of whether the URL was read or faxed versus sent via email.

2.4. *Frame Data*

Another source of auxiliary information available to study the effects of mode switch nonresponse is frame data. Data from the sampling frame is by far the most common source of auxiliary information used to study unit nonresponse and for the construction of nonresponse bias adjustments. However, the range of available variables is often limited for the purposes of studying nonresponse (Lynn 2003). Our study is no exception. We draw upon a limited set of frame variables collected from university records to study the effects of mode switch nonresponse: whether the sample person lived in Maryland at the time of graduation, age at the time of interview, gender, race (White vs non-White), and years since graduation with respect to the survey year (2005).

2.5. *Mode Switch Procedure*

Upon completing the screening questions, respondents were randomly assigned to one of three modes of data collection (CATI, IVR, and web) for the main interview. Respondents assigned to the web group were offered a $20 incentive to complete the web survey questionnaire. The specific wording of the mode switch introduction for the IVR and web groups is below:

IVR introduction:

> *Thank you for answering those background questions. Now I'd like to switch you over to an automated response system for some questions about your college experience and community involvement. We have designed the system to be easy to use and hope it will expedite your response. The rest of the interview will take about 10 minutes.*

Web introduction:

> *Thank you for answering those background questions. Based on your responses, you have qualified to participate in our web Survey. We would like to send you a link to a short web questionnaire about your college experience and community involvement that you can complete in your own time. If you complete the survey online, you will receive $20 in the mail.*

The initial goal was to achieve 360 completed cases for each mode. Based on response rate estimates by mode from the operations pretest, the initial allocation of sample cases was set at 19, 56, and 25 percent for CATI, web, and IVR, respectively. The allocation by mode was adjusted to 24, 36, and 40 percent at the end of August. During the final days of data collection, all the remaining cases were designated to the IVR mode of data collection. A total of 1,501 respondents completed the screening interview and were assigned to one of the three modes. The number of screener respondents allocated to CATI, web, and IVR was 338, 639, and 524, respectively (see Appendix Table A1).

2.6. Questionnaire and Validation Data

The main interview consisted of 37 questions. Most of these were contributed by the Alumni Association and are not relevant to the purposes of this article. In our analysis, we examine the effects of mode switch nonresponse using six questions for which validation data were available from university records for both respondents and nonrespondents:

1. What was your cumulative overall undergraduate grade point average or GPA at the time you received your undergraduate degree?
2. Did you ever receive a grade of "D" or "F" for a class?
3. During the time you were an undergraduate at the University of Maryland, did you ever drop a class and receive a grade of "W"?
4. . . .did you graduate with cum laude, magna cum laude, or summa cum laude?
5. Since you graduated, have you ever donated financially to the University of Maryland?
6. Are you a dues-paying member of the University of Maryland Alumni Association?

These questions came early in the questionnaire, but our numbering of the items above does not correspond to the item numbers in the questionnaire. We constructed two dichotomous variables based on the GPA item – whether the student graduated with a cumulative GPA less than 2.5 and whether they graduated with a GPA higher than 3.5. We thought that respondents with GPAs lower than 2.5 would be less motivated to carry out the mode switch (a GPA of 2.0 or less triggers an academic warning at the University of Maryland) than those with GPAs higher than 3.5 (a GPA that high or higher in a given term qualifies the student for the Dean's List). Items 2 through 6 are also dichotomous variables.

3. Results

3.1. Mode Switch Nonresponse

Of the 1,501 respondents who completed the screening interview and were assigned to a mode of data collection, 394 (or 26.2%) never started the main interview. The dropout rates for the IVR and web groups were 21.8% (or 114 out of 524) and 42.4% (or 271 out of 639), respectively. The total number of dropouts includes 9 screener respondents who were assigned to CATI but dropped out prior to starting the main interview. We exclude these cases and restrict our analysis to cases assigned to either of the two methods of computerized self-administration. In addition, there were a total of 29 screener respondents (IVR: 8; web: 21) who explicitly refused to participate in the mode switch in their assigned mode, at which point the interview was terminated. Although these cases contribute to the overall nonresponse associated with switching modes, we exclude them from our analysis and focus on nonrespondents who agreed to participate in the mode switch but dropped out prior to starting the main interview in the self-administered mode. Further exclusions include 18 cases (IVR: 13; web: 5) who dropped out before the full range of timestamp data could be collected. The final analytic sample consists of 1,116 cases allocated across IVR ($n = 503$) and web ($n = 613$) modes.

3.2. Relationship Between Call Record Data and Mode Switch Nonresponse

Table 1 shows the results of several logistic regression models predicting the propensity of contacting a sample member (Column 1), obtaining their cooperation (Column 2), and completing the mode switch (i.e., starting the main interview in the new mode; Columns 3–5) on call record, sampling frame, and screener paradata variables. The mode switch response models are shown separately for the IVR and web groups; the combined IVR/web group is presented in the last column. We present the unit nonresponse models (contact and cooperation) in an attempt to identify call record or frame variables that are linked to both unit and mode switch nonresponse. If such overlap is found, then intervention strategies or weighting adjustments aimed at dampening the effects of unit nonresponse may also be effective in ameliorating the effects of mode switch nonresponse.

Several questions can be answered from Table 1. The first question we address is whether hard-to-reach sample members and those who initially refuse the survey request, both known characteristics of unit nonresponders, are less likely to switch modes. We suspected that these active and passive forms of resistance would translate into a decreased likelihood of carrying out the mode switch, given that they completed the screening interview. However, we did not find evidence to support this conclusion. The number of callback attempts made to a sample case and whether an initial refusal was given were both highly correlated with either contact or screener cooperation, or both, as expected, but did not appear to decrease the likelihood of mode switch participation when controlling for frame and screener variables. Every additional contact attempt increased the logarithmic chance of being contacted by 0.39 ($p < 0.05$); the estimated logarithmic chance of participation on the other hand is reduced with every additional contact attempt by 0.74 ($p < 0.05$). The logarithmic chance for a screener respondent to continue with the main interview on the other hand never changed by more than 0.20 with each additional contact attempt in any of the modes and was not significant at the 5% level. The bivariate associations (not shown) were similarly weak and did not yield a reliable relationship.

3.3. Relationship Between Frame Data and Mode Switch Nonresponse

The second question we address (using Table 1) is whether auxiliary variables from the sampling frame are associated with respondents' likelihood of carrying out the mode switch. Covariates from the sampling frame are commonly used to study unit nonresponse and to create nonresponse weighting adjustments, but it is unknown whether such covariates are related to the likelihood of complying with the mode switch. Two covariates from the sampling frame were found to be significantly related to the mode switch outcome: gender and the number of years since degree was received. Females tended to be

Table 1. Contact, screener, and mode switch response propensity models by call records, frame data, and screener paradata. Coefficients are log-odds

Explanatory variables	Predicting contact ($N = 7{,}535$)	Predicting screener cooperation, conditional on contact ($N = 3{,}497$)	Predicting mode switch response ($Y = 1$)		
			IVR ($N = 503$)	Web ($N = 613$)	Combined ($N = 1{,}116$)
Intercept	−1.56***	0.23	1.30	0.63	0.81*
Call records					
# calls (log)	0.39***	−0.74***	−0.18	0.18	0.09
Ever refused	–	−1.81***	0.63	0.57	0.44
Frame data					
Lived in MD	0.08	−0.34***	0.11	0.03	0.11
Age	0.03***	0.03***	0.01	−0.03	−0.01
Female	0.01	−0.14	0.57*	0.14	0.30*
Non-White	−0.17**	0.003	−0.20	−0.18	−0.18
Years since degree	−0.01	0.002	−0.003	0.06*	0.03
Screener paradata					
At least one item missing			−0.18	−0.98*	−0.88**
Screener time < 25th pct			−0.36	0.17	−0.09
Screener time > 75th pct			−0.08	0.08	−0.02
Read/Fax'd web link			–	−2.04***	–
Likelihood Ratio	343.57***	634.20***	12.07	66.51***	22.76*
Pseudo R^2	0.05	0.17	0.02	0.10	0.02
Pseudo R^2 (max rescaled)	0.06	0.22	0.04	0.14	0.03

*** <0.0001; ** <0.01; * <0.05.

more likely to comply with the IVR mode switch (and the combined IVR/web mode switch) than did males ($b = 0.57$, $p < 0.05$). Also, less recent university graduates tended to have a greater likelihood of starting the main interview in the web mode than their predecessors ($b = 0.06$, $p < 0.05$). Neither of these variables is related to respondents' propensity of being contacted or cooperating with the survey request. Conversely, none of the frame variables associated with unit nonresponse (i.e., living in Maryland, age, and race) were associated with mode switch nonresponse. These findings suggest that factors related to unit and mode switch nonresponse do not necessarily overlap and can be distinct from each other.

3.4. Relationship Between Screener Paradata and Mode Switch Nonresponse

Next we examine whether information gleaned from the screener interview might signal whether a respondent participates in the mode switch. A key question is whether indicators of respondent resistance captured during the screening interview decrease the likelihood of carrying out the mode switch. We operationalized respondent resistance using two paradata outcomes – whether a respondent did not provide a substantive response (i.e., refused or "don't know") to at least one of the screener questions and the elapsed time of the screening interview. We hypothesized that respondents with item missing data and fast screener response times (i.e., first quartile of elapsed time distribution) would be less likely to carry out the mode switch than those with slower response times. The results in Table 1 support the first part of this hypothesis: respondents who did not provide a substantive response to at least one screener item were less likely to switch to the web mode ($b = -0.98$, $p < 0.05$); the same result was found for the combined IVR/web group ($b = -0.88$, $p < 0.01$). Respondents who finished the screening interview the fastest were less likely to participate in the IVR mode switch than slower respondents who finished within the middle quartiles of the elapsed time distribution ($b = -0.36$), but this result did not reach statistical significance ($p < 0.20$). There was no evidence that respondents with the slowest screener response times (i.e., last quartile of elapsed time distribution) participated in the mode switch at a lower rate than faster respondents ($b = -0.02$). Another finding, limited to the web group, was that respondents who did not have, or refused to provide, an email address and thereby received the web survey URL over the phone (or via facsimile) tended to comply with the mode switch at a much lower rate than those who provided an email address ($b = -2.04$, $p < 0.001$).

3.5. Relationship Between Auxiliary Variables and Survey Outcomes

So far, we have shown that utilizing auxiliary variables from different data sources can be useful for characterizing and identifying respondents who are unlikely to switch to a self-administered data collection mode. In order to tailor intervention strategies not just for increasing mode switch response rates, but also to reduce nonresponse bias, these data also need to be related to the survey outcomes. The next set of analyses examine the relationship between the auxiliary variables and the survey outcomes (based on university records). Figure 1 presents the absolute bivariate correlations of the auxiliary variables obtained from all data sources (call records, frame, and screener paradata) with the

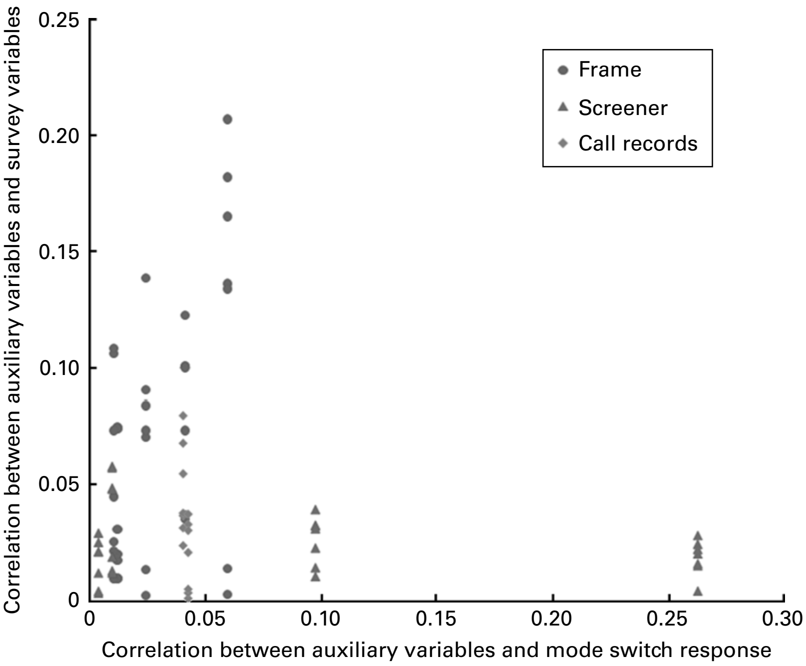

Fig. 1. Relationship between the correlation of the auxiliary variables and survey variables of interest and correlation of auxiliary variables and mode switch response indicator. All correlations are shown in absolute values

mode switch response indicator (x-axis) and the key survey outcomes described in Section 2.6 (y-axis).

A few observations can be made from Figure 1. First, most of the auxiliary variables are only weakly correlated with the mode switch response indicator. The only auxiliary variable that exceeds an absolute correlation of 0.20 is an indicator of whether an email address was provided by respondents assigned to the web mode. The next highest correlation is with an indicator of item missing data in the screening interview. Although both frame and screener paradata variables were found to be associated with the mode switch outcome (see Table 1), the strongest predictors of mode switch response, in general, appear to be the screener paradata variables, not the frame variables. The second observation is that the auxiliary variables are not highly correlated with the key survey variables of interest. Only one absolute correlation exceeds 0.20 – the correlation between gender and graduating with a cumulative GPA > 3.5. In general, the frame variables appear to have higher correlations with the survey outcomes than the screener paradata.

Although the auxiliary variables show only modest potential to be useful for reducing nonresponse bias, survey practitioners are ultimately interested in whether estimates are improved (i.e., lower nonresponse bias) by recruiting reluctant sampled units into the respondent pool; this is a critical question behind nonresponse reduction efforts. We estimated the reluctance of respondents to participate in the mode switch using propensity scores based on the combined IVR/web response propensity model in Table 1. Following suggestions from Rosenbaum and Rubin (1983), five roughly equal-sized propensity score groups were formed, ordered from low to high propensity of participating in the mode switch. Respondents were assigned the value 1 in the newly formed variable if their propensity to continue the interview after the mode switch lay between 0.37 and 0.65; the value 2 if their predictive propensity fell within the interval of 0.65 and 0.69; the value 3 for propensities between 0.69 and 0.72, the value 4 if the estimated propensity to switch 0.72 and 0.75, and the value 5 for the remaining respondents indicating the highest response propensity range of 0.75 to 0.86.

Figures 2–8 show record means for each survey outcome cumulated over propensity strata for respondents who complied with the mode switch. Moving from left to right on each graph indicates how the estimated mean changes through adding the lower propensity cases into the mode switch respondent pool. The dotted line in each graph represents the record mean for mode switch respondents. The solid line denotes the overall, or target, mean based on records for the full mode switch-eligible sample. Differences between the two lines indicate nonresponse bias for the respondent mean. The Wilcoxon rank-sum test indicated that the differences between the two lines were statistically significant ($p < 0.05$) in Figures 3–6.

A couple of observations can be made from these figures. First, in most instances (Figures 2–6), adding the low propensity cases into the respondent pool moves the estimates roughly towards the "true" record value. For example, the nonresponse bias in the mean proportion of students graduating with honors (Figure 6) decreases monotonically over the propensity strata as the lower propensity cases are added into the respondent pool. This is an indication that survey estimates can be improved by persuading low propensity cases to participate in the mode switch.

Second, the rate of change in estimates over propensity strata is not always constant. For example, the nonresponse bias in the proportion of respondents who received a grade of "D" or "F" (Figure 2) decreases rapidly as the highest propensity groups are added to the respondent pool, but flattens out as the most reluctant respondents are added, suggesting that the most reluctant cases do not differ substantially from the higher propensity cases (conditional on the propensity model).

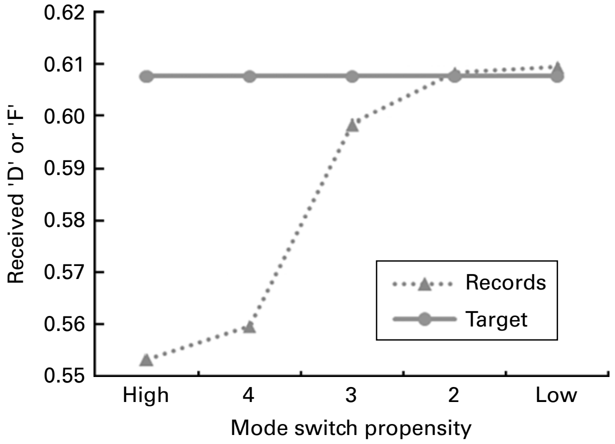

Fig. 2. Cumulative mean of mode switch propensity strata, received "D" or "F" ("D" and "F" reflect the lowest grades in the U.S. college grading system). Difference between lines is not statistically significant ($p = 0.33$)

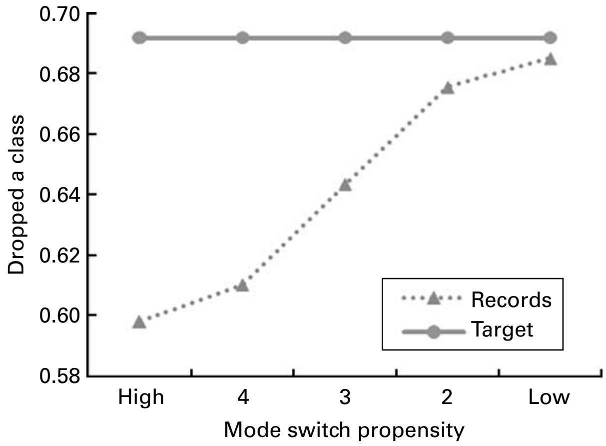

Fig. 3. Cumulative mean of mode switch propensity strata, dropped a class. Difference between lines is statistically significant ($p = 0.01$)

Lastly, in some instances adding lower propensity cases to the respondent pool does not improve estimates and may lead to increases in nonresponse bias. For example, Figures 7–8 exhibit roughly "U"-shaped patterns where the most accurate estimates are obtained by excluding the lowest propensity groups from the respondent pool. Such patterns may indicate misspecification of the propensity model, such as a missing covariate or interaction term.

4. Discussion

This analysis has five main findings. First, there was no indication that hard-to-reach sample members or sample members who refused the initial survey request – common characteristics of unit nonresponders – were more or less likely to comply with the mode switch than sample members who were initially cooperative or easier-to-reach. Second,

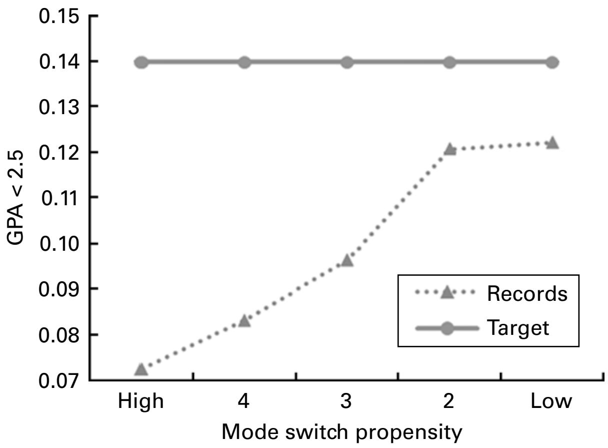

Fig. 4. Cumulative mean of mode switch propensity strata, GPA < 2.5. Difference between lines is statistically significant ($p = 0.01$)

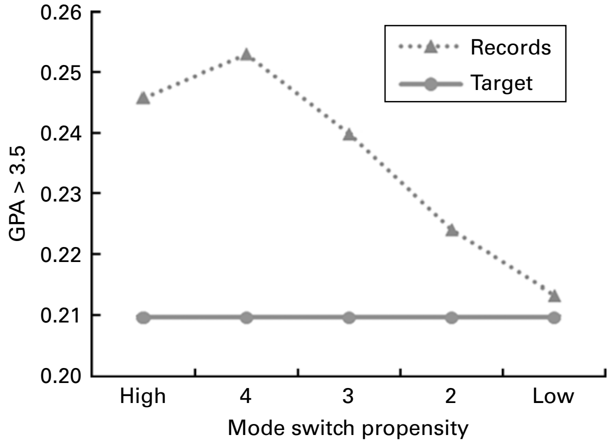

Fig. 5. Cumulative mean of mode switch propensity strata, GPA $>$ 3.5. Difference between lines is statistically significant ($p = 0.01$)

sampling frame covariates associated with respondents' likelihood of complying with the mode switch did not overlap with covariates associated with unit nonresponse (contact and cooperation), and vice-versa. Third, some paradata derived from the screening interview were useful in identifying indicators of respondent resistance that led to a decreased likelihood of participating in the mode switch. Fourth, the screener paradata variables yielded higher correlations with the mode switch outcome than the frame variables, but the frame variables yielded higher correlations with the survey outcomes than the screener paradata. Fifth, in most instances survey estimates were improved (i.e., lower nonresponse bias) by adding reluctant or low propensity cases to the mode switch respondent pool.

The options for survey practitioners considering a mode-switch survey design but worried about losing a substantial portion of their sample during the switch to self-administration (and possibly introducing bias into their results), are the following: 1) Forego the mode switch and use interviewer-administration for all interviews; 2) Carry

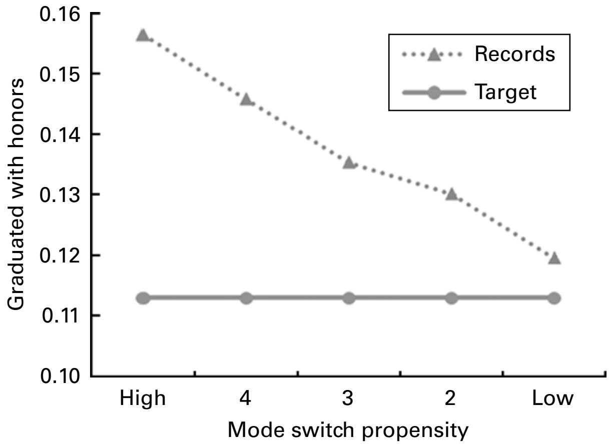

Fig. 6. Cumulative mean of mode switch propensity strata, graduated with honors. Difference between lines is statistically significant ($p = 0.01$)

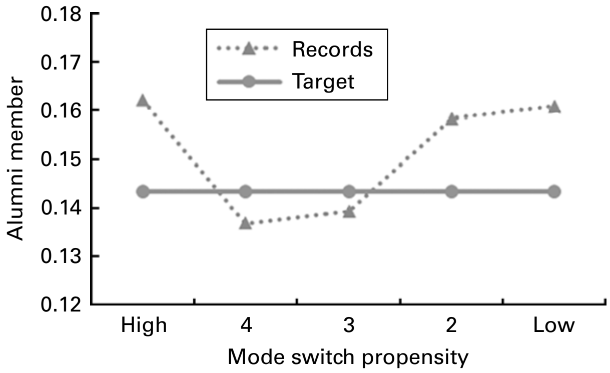

Fig. 7. Cumulative mean of mode switch propensity strata, alumni member. Difference between lines is not statistically significant ($p = 0.68$)

out the mode switch and ignore the effects of mode switch nonresponse; 3) Use auxiliary information to characterize mode switch nonrespondents post-data collection and caution data users on any imbalances found; or 4) Preidentify likely mode switch nonrespondents prior to the switch and implement tailored intervention methods to increase their likelihood of switching modes.

For the study described here, Sakshaug et al. (2010) found that implementing the mode switch increased nonresponse error and, thus, offset the advantages of reduced measurement error in self-administered modes for some variables but not for others. Furthermore, implementing the mode switch did not lead to an overall reduction in total bias. Hence, they recommend from a total survey error perspective to bypass the mode switch and use interviewer-administration for all interviews (Option 1). However, if the main focus is on measurement error bias a mode switch may be appropriate because it is not necessarily increasing the overall bias and in some cases does reduce measurement error.

Option (2) is the current practice where the main focus is on decreasing measurement error, even if that decrease comes at the expense of greater nonresponse. Indeed,

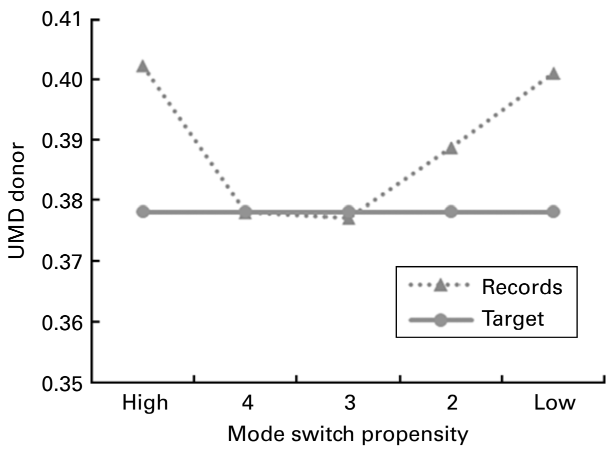

Fig. 8. Cumulative mean of mode switch propensity strata, UMD donor. Difference between lines is not statistically significant ($p = 0.33$)

measurement error can be the dominant source of error, particularly in surveys involving highly sensitive topics (e.g., illicit drug use). Yet, as already noted, the increase in nonresponse can diminish (or offset) gains in accuracy due to self-administration. Weighting adjustments meant to correct for unit nonresponse bias may correct for additional biases attributed to mode switch nonresponse. However, we did not find evidence that covariates related to unit nonresponse overlapped with those related to mode switch nonresponse (see Table 1). Thus, nonresponse weighting adjustments may not be effective for reducing mode switch biases unless a wide range of adjustment variables are used. Utilizing a wide range of adjustment variables from different sources, including paradata variables, has recently been considered for adjustment of unit nonresponse (Kreuter et al. 2010) and may be useful for mitigating the effects of mode switch nonresponse. This is an area for future work.

Option (3) depends on the richness of the auxiliary variables available to study and characterize mode switch nonrespondents. As is the case with studies of unit nonresponse, finding quality covariates on both respondents and nonrespondents is a persistent challenge. We utilized covariates from three different data sources (call records, sampling frame, and screener paradata), yet only a few were related to the mode switch outcome. Alternatively, we could have expanded our analysis by utilizing some of the substantive responses given to the screener questions (excluding the item missing data) to characterize mode switch nonrespondents, but preliminary analyses on these covariates (not reported) did not yield any associations with the mode switch outcome. As a result, we found that screener paradata variables achieved the highest correlations with the mode switch outcome, relative to the other data sources used (call records and sampling frame).

Preidentifying and intervening on likely mode switch nonrespondents (Option 4), again, requires a rich set of covariates that are predictive of mode switch behavior and collected prior to the switch. Predicting mode switch behavior requires either 1) a sequential learning algorithm, embedded in the computer-assisted interviewing software, capable of examining patterns and estimating posterior probabilities of mode switch behavior in real-time conditional on fixed characteristics and a history of previous outcomes, such as used in a responsive design framework (Wagner and Raghunathan 2007; Groves and Heeringa 2006), or 2) easily observable characteristics that signal reluctance to participate in the mode switch and that are generalizable across different surveys. The current study focused on the latter requirement. Indicators of respondent resistance identified during the screening interview, such as item missing data and inability to provide an email address (web group), were predictive of respondents' reluctance to carry out the mode switch, even though they ostensibly agreed to do so. These indicators are easily observable by interviewers or the computer-assisted interviewing software and may be used to trigger responsive design strategies aimed at increasing the likelihood that a respondent will comply with the mode switch. (e.g., offering larger incentives or sending periodic reminders when switching to a web mode). Such efforts are likely to be fruitful as we found adding reluctant sample members to the mode switch respondent pool tended to reduce nonresponse bias and improve the majority of the estimates (see Figures 2–8). Of course, implementing such intervention strategies should be considered with regard to the survey budget. Surveys that include a budget for nonresponse reduction and follow-up should consider allocating a portion of their funds to efforts aimed at reducing the effects of mode switch nonresponse.

A certain limitation of the current study is its reliance on alumni from a single university. However, given the size of the University of Maryland (in the last 20 years on average about 25,000 undergraduates) and the wide range of alumni (reaching from graduating classes from 1989 to 2002) we feel confident that the sampling frame reflected a rather diverse population. Nevertheless, attempts to replicate our findings should be mounted on a broader target population. Replication is needed to confirm, for example, whether the same indicators of respondent resistance collected during the screening interview are similarly effective in predicting mode switch behavior in different surveys. In addition, future work may consider the study of interviewer effects on the mode switch outcome. Interviewers play an important role in motivating respondents to comply with the mode switch. Identifying characteristics of the most successful interviewers may be useful for purposes of enhancing interviewer training procedures aimed at motivating respondents to carry out the switch. Finally, as is the case for studies of unit nonresponse, future work is needed to identify additional characteristics and predictors of mode switch nonresponse and optimal methods of reducing and/or adjusting for its effects; in particular in times of rapid changes in technological environment new predictors are likely to appear.

Appendix

Table A1. Final disposition codes

	Total	%	%	%
Seemingly usable phone numbers fielded	7,591	100.0	–	–
Not eligible and deceased	2,889	38.1	–	–
Eligible cases and unknown eligibility	4,702	61.9	100.0	–
Unknown eligibility	1,914	–	40.7	–
Eligible, no-interview				
Language barrier	33	–	0.7	–
Physically/mentally unable	7	–	0.1	–
Noncontact	797	–	17.0	–
Refusal	441	–	9.4	–
Partial screener completion	9	–	0.2	–
Screener completed and assigned to mode	1,501	–	31.9	
Initially assigned to CATI	338	–	–	100.0
Started main questionnaire in CATI	328	–	–	97.0
Initially assigned to web	639	–	–	100.0
Provided e-mail address	617	–	–	96.6
Started main questionnaire web	368	–	–	57.6
Initially assigned to IVR	524	–	–	100.0
Started main questionnaire IVR	389	–	–	74.2

5. References

American Association for Public Opinion Research (AAPOR). (2008). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. Technical Report.

Bates, N., Dahlhamer, J., and Singer, E. (2008). Privacy Concerns, Too Busy, or Just Not Interested: Using Doorstep Concerns to Predict Survey Nonresponse. Journal of Official Statistics, 24, 591–612.

Branden, L., Gritz, R.M., and Pergamit, M. (1995). The Effect of Interview Length on Attrition in the National Longitudinal Study of Youth, National Longitudinal Surveys Discussion Paper, No. 28. Washington, DC: U.S. Bureau of Labor Statistics, U.S. Department of Labor.

Centers for Disease Control and Prevention (CDC). (2010). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Questionnaire. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. Available at http://www.cdc.gov/nchs/nhanes/nhanes2009-2010/questexam09_10.htm

Couper, M.P. and Lyberg, L. (2005). The Use of Paradata in Survey Research. Proceedings of the 55th Session of the International Statistical Institute (CD-ROM).

Couper, M.P., Singer, E., and Tourangeau, R. (2003). Understanding the Effects of Audio-CASI on Self-Reports of Sensitive Behavior. Public Opinion Quarterly, 67, 385–395.

Fricker, S., Galesic, M., Tourangeau, R., and Yan, T. (2005). An Experimental Comparison of Web and Telephone Surveys. Public Opinion Quarterly, 69, 370–392.

Groves, R.M. and Couper, M.P. (1998). Nonresponse in Household Interview Surveys. New York: John Wiley.

Groves, R.M. and Heeringa, S.G. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. Journal of the Royal Statistical Society: Series A (Statistics in Society), 169, 439–457.

Groves, R.M., Mosher, W.D., Lepkowski, J.M., and Kirgis, N.G. (2009). Planning and Development of the Continuous National Survey of Family Growth. National Center for Health Statistics. Vital and Health Statistics, 1(48).

Hawkes, D. and Plewis, I. (2006). Modelling Nonresponse in the National Child Development Study. Journal of the Royal Statistical Society, Series A, 169, 479–492.

Hill, D.H. and Willis, R.J. (2001). Reducing Panel Attrition: A Search for Effective Policy Instruments. Journal of Human Resources, 36, 416–438.

Kaplowitz, M.D., Hadlock, T.D., and Levine, R. (2004). A Comparison of Web and Mail Survey Response Rates. Public Opinion Quarterly, 68, 94–101.

Kreuter, F. and Kohler, U. (2009). Analyzing Contact Sequences in Call Record Data. Potential and Limitations of Sequence Indicators for Nonresponse Adjustments in the European Social Survey. Journal of Official Statistics, 25, 203–226.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R.M., and Raghunathan, T.E. (2010). Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Non-Response: Examples from Multiple Surveys. Journal of the Royal Statistical Society, Series A, 173, 1–19.

Kreuter, F., Presser, S., and Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. Public Opinion Quarterly, 72, 847–865.

Lee, E., Hu, M.Y., and Toh, R.S. (2004). Respondent Non-Cooperation in Surveys and Diaries: An Analysis of Item Nonresponse and Panel Attrition. International Journal of Market Research, 46, 311–326.

Loosveldt, G., Pickery, J., and Billiet, J. (2002). Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey. Journal of Official Statistics, 18, 545–557.

Lynn, P. (2003). PEDAKSI: Methodology for Collecting Data about Survey Non-Respondents. Quality and Quantity, 37, 239–261.

Rosenbaum, P.R. and Rubin, D.R. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 70, 41–55.

Sakshaug, J.W., Couper, M.P., and Ofstedal, M.B. (2010a). Characteristics of Physicial Measurement Consent in a Population-Based Survey of Older Adults. Medical Care, 48, 64–71.

Sakshaug, J.W., Couper, M.P., and Ofstedal, M.B. (2010b). Patterns of Consent: Linking Longitudinal Health Survey and Social Security Administration Records. Paper presented at the 8th International Conference on Health Policy Statistics, Washington, D.C., February.

Sakshaug, J.W., Yan, T., and Tourangeau, R. (2010). Nonresponse Error, Measurement Error, and Mode of Data Collection: Tradeoffs in a Multi-Mode Survey of Sensitive and Non-Sensitive Items. Public Opinion Quarterly, 74, 907–933.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2009). 2008 National Survey on Drug Use and Health: Data Collection Final Report. Technical Report Prepared by RTI International. Available at http://www.oas.samhsa.gov/nsduh/2k8MRB/2k8DCFR.pdf

Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. In Survey Nonresponse, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds). New York: John Wiley.

Tourangeau, R., Rips, L.J., and Rasinski, K. (2000). The Psychology of Survey Response. New York: Cambridge University Press.

Tourangeau, R. and Smith, T.W. (1998). Collecting Sensitive Information with Different Modes of Data Collection. Computer Assisted Survey Information Collection, M.P. Couper, R.P. Baker, J. Bethlehem, C.Z. Clark, J. Martin, W.L. Nicholls, and J. O'Reilly (eds). New York: John Wiley.

Tourangeau, R., Steiger, D.M., and Wilson, D. (2002). Self-Administered Questions by Telephone – Evaluating Interactive Voice Response. Public Opinion Quarterly, 66, 265–278.

Wagner, J. and Raghunathan, T.E. (2007). Bayesian Approaches to Sequential Selection of Survey Design Protocols. Proceedings of the American Statistical Association, Survey Research Methods Section, 3333–3340.

Zabel, J.E. (1998). An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labour Market Behavior. Journal of Human Resources, 33, 479–506.

Received May 2010
Revised January 2011