Published in final edited form as:

Field methods. 2014 May 1; 26(2): 141–155. doi:10.1177/1525822X13491863.

Does Sequence Matter in Multi-Mode Surveys: Results from an Experiment

James Wagner, Jennifer Arrieta, Heidi Guyer, and Mary Beth Ofstedal

University of Michigan, Survey Research Center

Abstract

Interest in a multi-mode approach to surveys has grown substantially in recent years, in part due to increased costs of face-to-face interviewing and the emergence of the internet as a survey mode. Yet, there is little systematic evidence of the impact of a multimode approach on survey costs and errors. This paper reports the results of an experiment designed to evaluate whether a mixed-mode approach to a large screening survey would produce comparable response rates at a lower cost than a face-to-face screening effort. The experiment was carried out in the Health and Retirement Study (HRS), an ongoing panel study of Americans over age 50. In 2010, HRS conducted a household screening survey to recruit new sample members to supplement the existing sample. The experiment varied the sequence of modes with which the screening interview was delivered. One treatment offered mail first, followed by face-to-face interviewing; the other started with face-to-face and then mail. A control group was offered only face-to-face interviewing. Results suggest that the mixed mode options reduced costs without reducing response rates to the screening interview. There is some evidence, however, that the sequence of modes offered may impact the response rate for a follow-up in-depth interview

Introduction

Interest in the use of multiple modes of data collection has increased recently (de Leeuw, 2005). This interest has been spurred by several factors. First, technological development has led to the creation of new survey modes -- primarily the internet. Second, the declining coverage of landline Random-Digit Dial (RDD) surveys has turned attention to the use of Address-Based Sampling with multiple modes of data collection. For example, under such a design, addresses for which a telephone number can be obtained are assigned to the telephone mode; addresses for which there is no telephone number are assigned to the mail mode. Third, declining response rates across sampling frames and modes has led researchers to use multiple modes of data collection as a way to control costs and raise or, at least, maintain response rates (de Leeuw, 2005). Fourth, increasing survey costs have led survey organizations to consider alternative, lower cost approaches to conducting surveys.

The use of multiple modes requires consideration of the many possible impacts of collecting data in this manner. The different modes may produce different types of response error. For instance, the presence of an interviewer in either telephone or face-to-face interviewing may induce social desirability bias that is not present in self-administered survey modes such as mail or internet. Different modes may also be associated with different *nonresponse* errors. Some sampled units may be more likely to be interviewed in one mode than another. For

example, a recent meta-analysis (Manfreda et al., 2008) showed that surveys administered over the internet tend to have lower response rates than other modes. If persons who would respond to either a telephone or internet request to complete a survey are different than those who would respond to the telephone request but not an internet request, then these modes could experience differential nonresponse bias. These errors of response and nonresponse may also interact across the multiple modes. For instance, the measurement error of a mode may be worse when the mode is used as a follow-up among nonrespondents to a previous mode. In designing a multiple mode survey, it is important to consider all the potential sources of error and ways in which they may interact. This “total survey error” perspective is advocated for multiple mode surveys in particular by de Leeuw (2005).

In terms of the impact of the sequence of modes, there are disparate results in the existing literature. Among the earliest studies of the effect of mode were those conducted by Hochstim (1967) and Siemiatycki (1979). The major finding of both studies was that the response rates were very similar for several mixed mode designs including mail, telephone, and face-to-face contacts, but that costs were much less for sequences that began with the cheaper modes – mail and telephone.

Several more recent studies have reported similar findings. Rodriguez et al. (2006) found that different sequences of mail, internet and Interactive Voice Response (IVR, where a recorded voice asks questions and the respondent selects numbers on the telephone keypad to respond) surveys produced similar response rates. McMorris et al. (2009) found that the two sequences of internet and face-to-face interviewing produced comparable response rates and estimates. In a recent study Dillman et al. (2009) compared several mixed mode designs and found that mail followed by telephone performed as well on response rate as telephone followed by mail. In all these research studies, the sequence of modes has not mattered in terms of the final response rate.

On the other hand, there are several studies for which the sequence of modes does make a difference. McHorney et al. (1994) conducted a survey that used a sample of respondents to the General Social Survey (GSS, a national survey on societal trends in the U.S. that has been ongoing since 1972). As a result, they had all of the information obtained in the GSS interview for both respondents and nonrespondents to their survey. While Dillman et al. (2009) found that the sequence did not matter in relation to the response rate, McHorney and colleagues (1994) found differences in response rates across the two sequences – telephone followed by mail (68.9%) and mail followed by telephone (79.2%). Harris et al. (1997) also found differences in response rates across these two sequences, but the direction of the difference was reversed. That is, they found that a mail survey with telephone follow-up had a 50% response rate while a telephone survey with mail follow-up had a 73% response rate. This study also found differences in item missing data across the two sequences. Beebe et al. (2007) compared response to a web survey of physicians followed by mail to a mail survey followed by web. They found that the web/mail sequence had a lower response rate (62.9%) compared to the mail/web sequence (70.5%). They also found that the web/mail sequence was completed faster. Smyth et al. (2010) found that a mail survey with web follow-up produced higher response rates (71.1%) than a web survey followed by a mail survey.

(55.1%). Similar results were obtained in other studies (Converse, et al., 2008; Borkan, 2009).

However, in many of these studies, the effort applied at each phase is either not equivalent or well-controlled. For example, in the study by McMorris et al. (2009), the field period had a limited time. As a result, the treatment arm that received the web survey first and then the face-to-face treatment, may have had fewer face-to-face attempts than the arm that had face-to-face interviewing as the first mode. In the study by Smyth et al. (2010), which compared web-mail, mail-web, and the choice of web or mail, the first mode in the web-mail and mail-web arms received an additional mailing while the follow-up mode only received a single mailing, thus increasing the dosage of the first mode as well as delivering them in a different sequence. These differences in the treatment were largely a result of the fact that evaluating the sequence of modes was not an explicit goal of the study.

We were concerned that a mixed mode approach might harm response rates in a survey that had previously been conducted using only face-to-face interviewing. In the context of a mixed mode approach, the fear is that if a mode with an expected lower response rate is tried earlier in a sequence, this will increase resistance to later attempts in a different, more effective mode. This is similar to the “contamination” hypothesis from the research on doorstep interactions (Groves and Couper, 1998). In the context of large face-to-face surveys, there is a concern that if interviewers press the request to complete the survey in early interactions, they will increase resistance to the survey. In addition, as can be seen from the review of the literature, very little recent research has looked at mixed mode surveys with face-to-face interviewing as one of the modes. Following the results of the early studies conducted by Hochstim (1967) and Siemiatycki (1979), most mixed mode surveys start with the cheaper mode and reserve the more expensive mode for follow-up with a smaller number of cases.

In our case, there were concerns that beginning with a mailed survey might “contaminate” the sample and harm our ability to attain high response rates. We also hypothesized that a mixed-mode survey could be completed for a lower overall cost. The present study reviews the results of an experiment designed to determine whether a mixed-mode approach to the screening survey could attain comparable response rates to a face-to-face screening survey. The experiment would also be used to determine whether and how much cost savings the mixed-mode approach would offer. If the mixed-mode design can produce similar response rate at a lower cost, then it would be preferred. In addition, we consider whether the sequence of modes used in the mixed mode design impacts the ability to complete a follow-up, in-depth interview with persons who are identified to be eligible.

Methods

The Health and Retirement Study (HRS) is a national panel survey of persons over the age of 50 in the United States. Participants are interviewed either in-person or by telephone every two years and are also invited to participate in mail and internet surveys in-between interview waves. The primary focus of the study is on the interrelationship between health and economic status in the years leading up to and following retirement.

In 2010, the HRS was adding to its sample of persons 51 to 62 years of age (specifically, persons born between 1948 and 1959). This sample was meant to supplement a sample selected and interviewed in 2004. The sample for this supplement is a multi-stage area probability sample. The sample was selected in stages starting with Primary Sampling Units (PSUs, counties or Metropolitan Statistical Areas). Within sampled PSUs, a sample of neighborhoods was selected. These Secondary Sampling Units (SSUs) were defined by the geographic boundaries of Census blocks. Within each SSU, a sample of housing units was selected. The sample of housing units was selected from the list of all housing units in each SSU since the sampling frame does not contain information about which housing units are occupied, nor does it contain information about which housing units contain persons in the eligible age group. The selected housing units were then screened to determine if there was an eligible person present. For each sampled housing unit, a “screening interview” was attempted. This interview collected information about each person living in the household. This interview was used to determine if a person born between 1948 and 1959 lived in the household. If so, that person and his/her spouse or partner (if coupled) were asked to complete an in-depth interview for the Health and Retirement Study.

The sample of housing units was released in three replicates, or subsampled portions of a larger sample. Since the 2010 sample was a supplement meant to increase the number of minorities in the panel, the sample was selected from areas with at least 10% black population or at least 10% Hispanic population. When combined with the earlier sample, the new cohorts of persons born between 1948 and 1959 is a fully representative national sample.

For the purposes of this experiment, two random samples of SSUs were selected from across all SSUs that were part of the third release of sample, one for each of two treatment groups. As described in more detail below, one treatment group (Mail-FtF) received a mail screener survey first, followed by a Face-to-Face (FtF) interview request. The other treatment group (FtF-Mail-FtF) received a FtF interview request first, followed by a mail survey; ultimately non-respondents to the mail survey were returned to the field staff for FtF followup. The goal was to select about 1,000 sampled housing units for each treatment group. In the end, for the Mail-FtF group, 23 SSUs from 21 PSUs were selected. The selected SSUs for this group contained 992 sampled housing units. For the FtF-Mail-FtF group, 22 SSUs from 21 PSUs were selected. The selected SSUs for this group contained 1,168 sampled housing units. The remainder of the sample not selected for either experimental treatment consisted of 8,161 lines in 191 SSUs from 58 PSUs. This latter sample can be considered a control group that was given only the FtF treatment. The sample sizes varied from the expected sample sizes since the SSUs were of unequal sizes.

The first step in the process of adding new members to the HRS panel was to complete a screening interview – either by mail or face-to-face depending upon the treatment and timing of completion. The screening interview was brief – the main part of the survey included questions asking an informant to list all adult household members along with their race/ethnicity, marital status, and age. The screening survey was very similar in content to the Decennial Census. Since it closely followed the Decennial Census, it was thought that familiarity with this type of questionnaire might improve response to our survey. After the

screening interview was completed, interviewers attempted to contact identified eligible persons to set an appointment to complete the in-depth interview. The two interviews are conducted on two separate occasions. The in-depth interview included questions about health, plans for retirement, and -- for a random subset of eligible persons -- a set of physical measurements administered by the interviewer. It was a requirement of the survey that the in-depth interview be conducted face-to-face, although in some rare circumstances a telephone interview was allowed.

The mailed screening survey consisted of a letter explaining that an interviewer from the University of Michigan's Survey Research Center would be in the area conducting surveys for which members of the household may be eligible. It also included a form with questions asking an informant to list all adult members of the household and their age, marital status, race, and ethnicity. A \$2 prepaid incentive was included with the initial mailing. For the mailed survey component of each treatment arm, the nonrespondents to the initial mailed survey were sent another mailed survey after two weeks. The cases that were designated for FtF mode attempts were sent a pre-notification letter, but did not receive an incentive to conduct the screening interview as part of the FtF attempts.

The calling strategy for the FtF mode across the three arms, as is typical for a face-to-face survey, was determined by the interviewers. Interviewers are given general training about good times to call and how to organize efficient trips to sampled segments. But they make the decision about how to approach the work. In this environment, it is difficult to guarantee that equal effort will be applied to all cases. Previous studies have shown that interviewers vary in their ability to establish contact (Campanelli et al., 1997; Purdon et al., 1999; Pickery and Loosveldt, 2002; Durrant and Steele, 2009). However, productivity and response rates were monitored at the interviewer level on a daily basis. Interviewers who were performing at lower rates of efficiency or quality received advice and help from supervisors. In addition, relatively high targets for response rates were set which required high level of effort be applied to all cases that were difficult to contact or from whom it was difficult to gain cooperation.

The samples were released for production in late June-early July 2010. The samples for the FtF control group and the FtF-Mail-FtF treatment group were released to interviewers in late June. Both of these groups were sent a pre-notification letter. The letter announced that an interviewer from the University of Michigan's Survey Research Center would be in the neighborhood to determine if anyone in the household is eligible for an important survey. After five weeks in the field nonfinal sample -- that is, the portion of the sample that had not been finalized as a completed interview, non-sample unit, or final refusal -- from the FtF-Mail-FtF group was sent a mail survey. A second mailing of the mail survey was sent two weeks later during the seventh week of the field period to sample units that had not responded to the first mail survey. Finally, the nonfinal sample in the FtF-Mail-FtF group was returned to field staff five weeks after the first mailing during the tenth week of data collection.

The first mailing of the mail survey for the Mail-FtF group went out on July 13th and 14th. A second mailing was sent two weeks later. Five weeks after the initial mailing, the nonfinal

sample from the Mail-FtF group was transferred to field staff for face-to-face attempts. Figure 1 shows the sequencing of the treatments for the three arms.

All cases were worked until an overall 70% screening response rate had been achieved in December 2010. At that point, a two-phase sampling approach was implemented (Hansen and Hurwitz, 1946). A 50% subsample of the remaining active cases was selected for follow-up. At the point at which the subsampling was carried out, both arms of the experiment had completed the mailed portion of the survey data collection effort. Cases that were selected for follow-up then received a weighting factor of two to account for this sampling. The analyses reported here use those weights unless otherwise noted.

All response rates are calculated using AAPOR RR1 calculations (AAPOR, 2011). Under this calculation, the denominator of the screening response rate is the total sample size minus the count of nonsample cases (largely unoccupied housing units). The numerator is the count of completed screening interviews. There were no partial interviews to the screening interview. We also report an interview rate (in-depth interview) for identified eligible persons. These rates are reported using two sets of weights. The first set of weights account for the two-phase sampling. Since the two-phases are so strongly related to the propensity to respond, these weights are essential to accurately assess response. Estimates using these weights are useful for assessing operational effectiveness. The second set of weights account for the full probability of selection. These weights allow inference to the larger population and other sample designs.

In order to compare the costs of the different treatments, it was necessary to create a cost model for the costs related to face-to-face interviewing. Field interviewers reported time on a daily basis. On the same day, they may have visited sample from all three arms of the experiment. They did not record the time devoted to sample in each experimental treatment separately. They did, however, record every attempt made to contact each household. They also reported travel time as a separate category on their timesheets. This information was used to model the costs of each arm. The costs for the mail survey, on the other hand, were known. These costs include printing materials, assembling the mailings, and postage.

The costs for face-to-face interviewing were allocated to each arm of the experiment using the following approach. The number of trips in each arm was calculated from the data. The interviewer hours reported to be spent in travel were then allocated to each arm proportional to the number of trips in that arm. The number of calls in each arm were also calculated from the data. The interviewer hours spent in activities other than travel were then allocated to each arm proportional to the number of calls made to cases in that arm. Interviewers also charged expenses for this travel. These charges are largely for mileage in their own automobiles. These charges were allocated across the treatment arms in the same way as the hours. Averaging over all the calls simplifies the problem as not all calls take the same amount of time. An interview certainly takes longer than an attempt at which no one is home.

The treatments will be compared in terms of response rates (to the screening interview and a subsequent follow-up in-depth interview), costs (estimated dollars and select measures of

effort, such as number of face-to-face attempts per complete), and outcome measure (proportion of screened cases in the eligible age group). Design-based estimates of variance were used in all statistical comparisons.

Results

Table 1 presents several key statistics for each arm of the experiment. Overall, there was lower response to the mailed survey compared to the face-to-face survey. A combined total of 218 (144+74) households from 1,205 (553+652) interviewed in the two treatment arms with a mailed option responded to the mailed version of the survey.

The final response rates to the screening interview across the three arms are nearly identical – ranging from 84–86% (second-phase weighted, 92–94% fully weighted). These differences in response rate are not statistically significant when using a Rao-Scott design-adjusted chi-square test ($X^2(2, 116) = 1.795, p < 0.4075$). These results are encouraging as it appears that the “contamination” hypothesis does not apply. The use of a mailed screening survey in combination with face-to-face interviewing does not appear to depress response below that of the face-to-face only mode.

The eligibility rates differ across the three arms, ranging from 15–19% (second-phase weighted, 12–14% fully weighted). Using a Rao-Scott chi-square test, these differences are not statistically significant ($X^2(2, 116) = 0.4467, p < 0.7998$). Thus, it appears that the three treatments did not produce differences in the key measures.

The costs (estimated for the face-to-face portions) of the different treatment arms are shown in Table 2. The costs of the mailed survey (incentives, formatting materials, printing, assembly of materials, and mailing) were relatively small per sampled unit. The differences in response rates to the mailed survey explain the differences in cost per complete. The FtF-Mail-FtF group produced fewer completes per mailing and, therefore, has a higher cost per complete. The total screening costs include both the costs of the mailed survey and the face-to-face attempts. It appears that the Mail-FtF arm is the most cost effective approach. This arm has the largest proportion of screening interviews completed by mail. Since mail is the cheaper mode, this approach was the least costly. The FtF-Mail-FtF arm is more expensive (about 17% more) than Mail-FtF, but very close in cost to the FtF only arm. The costs of the in-depth interview were actually less for the FtF-Mail-FtF arm, but the screening costs were sufficiently higher to make this arm the most expensive treatment of the three.

Finally, the screening interview is just a means for identifying persons who will be eligible for an in-depth interview. We were concerned that the mixed-mode approach to screening might reduce our ability to obtain an in-depth interview. In other words, we wanted to evaluate the “contamination” hypothesis for this phase as well – even though the mixed modes were used in an earlier phase. This in-depth interview on the health and financial status of eligible persons is the ultimate goal of the survey. As a reminder, the in-depth interview was conducted in-person by an interviewer. That is, cases in every arm of the experiment received the same treatment for this part of the process. Table 2 includes the interview rates among the different treatment arms.

From Table 2, it appears that the FtF-Mail-FtF treatment may have led to better interview rates for the in-depth interview (73% vs. 67% for both the Mail-FtF and FtF only arms using the second-phase weight, 78% vs. 68% FtF and 70% Mail-FtF). However, using the Rao-Scott chi-square test, these differences are not statistically significant ($X^2(2, 114) = 4.4642$, $p < 0.1073$). It may be that a significant difference could be detected with a larger sample size or it may be that this difference is due to sampling error. For this comparison we had 1372 cases in the FtF only arm, 189 in the Mail-FtF group, and 180 in the FtF-Mail-FtF group. In any event, it does seem that the mixed-mode treatment does not harm response rates and may – subject to further evidence – even help response rates to the in-depth interview.

Although these response rates were very close, one important subgroup did experience significantly different response rates across the treatment groups. Interviewers record whether each sampled unit is in a locked building or gated community. It is more difficult to establish contact in these settings and, as a result, response rates are typically lower than in places without these impediments. In locked buildings and gated communities, FtF-Mail-FtF had a response rate of 67% on the in-depth interview. This was significantly higher (Rao-Scott Chi-Square test of independence, $p < 0.0001$) than for either FtF (46%) or Mail-FtF (41%).

The intent was to treat cases equally across the treatment arms. However, it was difficult to implement the same face-to-face effort across all cases. In the FtF-Mail-FtF arm, 97% of cases received at least one face-to-face attempt before the mailed survey was sent. The mean number of attempts on these cases before sending the mailed survey was 2.75 with a median of 2. The two treatment arms (Mail-FtF and FtF-Mail-FtF) received approximately equivalent effort by the end of data collection. This can be measured by the average number of attempts on nonfinal cases. This was 8.71 (SE=0.452) for the Mail-FtF arm and 9.04 (SE=0.373) for the FtF-Mail-FtF arm. These means are not significantly different. Cases in both of these arms received the same number of mail attempts – two mailings for those who failed to respond to the first mailing.

Although the in-depth interview rates were not significantly different across the treatments, there were differences in costs. The Mail-FtF treatment was the cheapest, while the FtF-Mail-FtF and FtF arms were somewhat more expensive. The in-depth interview costs were lowest for the FtF-Mail-FtF arm.

Further, an examination of the calls required to complete the in-depth interview (including calls to cases that were never completed) revealed that the FtF-Only cases required, on average, 20.6 (SE=1.18) attempts per interview. The Mail-FtF groups required 24.0 (SE=2.62) attempts per completed in-depth interview. The FtF-Mail-FtF group, on the other hand, required only 19.1 (SE=3.40) attempts per completed in-depth interview. This is in line with the expectation that the screening costs for FtF-Mail-FtF are higher, while the costs for the in-depth interviewing are lower. For a single-stage survey (e.g. no screening), Mail-FtF may be preferred as it has equivalent response rates and lower costs. The decision for a two-stage survey is more complex. The FtF-Mail-FtF mode sequence might be preferred for such a survey if a somewhat higher response rate is desired. But this will likely come at a higher cost.

The eligibility rate for the HRS was quite low given the focus on individuals born between 1948–1959. For a survey with a higher eligibility rate, the overall cost might be lower for the FtF-Mail-FtF design. In such a survey, the in-depth interview would be a larger proportion of the total costs. Since this part of the process appears to be cheaper for the FtF-Mail-FtF sequence, this may become cheaper than the other two design options at higher rates of eligibility.

Discussion

We found that it is possible to introduce a mixed-mode procedure to a screening interview process that is usually done face-to-face. The use of a mode (mail) that usually has lower response rates, did not harm the overall response rate. In other words, there was no evidence of a “contamination effect.”

In addition, the mixed-mode survey that used the mail mode as the second mode – while being somewhat more expensive -- also led to higher response rates with the follow-up in-depth survey. We note the interesting result that the FtF-Mail-FtF sequence had a higher response rate in locked buildings and gated communities than the other treatments. As a result, this sequence may be preferred, even at the higher cost, as the higher response rate might justify the increase in cost.

The experiment is subject to some limitations. Future research will be directed at these issues. The experiment was conducted on a limited domain. The sample was drawn from all neighborhoods with at least 10% black or 10% Hispanic population. The response rate to our mailed survey was quite low in both treatment arms. We speculate that this may be due to the low saliency of a screening survey that only asks for a listing of adult household members. We also only had two mailings over a 5 week period. The use of a more elaborate mail survey design, such as Total Design Method (Dillman, Smyth, Christian, 2009), could lead to increased response rates. These increased response rates would likely further tip the balance toward sequences which include a mailed component. Higher response rates to the mailed survey are likely to lead to cost savings as the mailed survey is less expensive to administer. It is also possible that higher mailed survey response rates could ameliorate the impact of the mode sequence on the in-depth interview rate. The in-depth interview was only for persons born between 1948 and 1959. Other populations may respond to second-stage requests differently. We might also consider adding content to the screening survey. The additional content may be useful for nonresponse adjustments and/or of interest substantively. It may also make the survey more salient for parts of the sample. Finally, our survey separated the screening and in-depth interviewing. In the majority of cases, each of these interviews required separate visits. If both interviews could be completed on a single visit, this might reduce the cost of face-to-face interviewing relative to the mailed survey.

Our conclusion is that given these essential survey conditions, a mixed mode approach does not harm response rates to a screening interview. Improved design for the mailed survey should lead to cost savings by increasing response rates to the mailed component, which has lower costs. We found that a face-to-face attempt, followed by a mailed follow-up may be more effective than the other treatments at producing higher response rates, although further

evidence on this point is needed. We found that it is important to look beyond the immediate phase of the survey on which an experiment is being run to see if any effects on subsequent phases may result. In this case, the modes used for a screening survey and the sequence in which they were used may have some effect on our subsequent ability to conduct an in-depth interview. For this reason, it is important to consider the survey process as a whole, rather than considering each successive stage separately.

References

- AAPOR. Standard Dispositions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7. 2011.
- Beebe TJ, Locke GR III, et al. Mixing Web and Mail Methods in a Survey of Physicians. *Health Services Research*. 2007; 42(3p1):1219–1234. [PubMed: 17489911]
- Borkan, B. *Social Science Computer Review*. 2009. The Mode Effect in Mixed-Mode Surveys: Mail and Web Surveys.
- Campanelli, P.; Sturgis, P.; Purdon, S. Can you hear me knocking?: investigation into the impact of interviewers on survey response rates. National Centre for Social Research; 1997.
- Converse PD, Wolfe EW, et al. Response Rates for Mixed-Mode Surveys Using Mail and E-mail/ Web. *American Journal of Evaluation*. 2008 1098214007313228.
- Dillman DA, Phelps G, et al. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research*. 2009; 38(1):1–18.
- Dillman, DA.; Smyth, JD.; Christian, LM. *Internet, mail, and mixed-mode surveys : the tailored design method*. Hoboken, N.J: Wiley & Sons; 2009.
- de Leeuw ED. To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*. 2005; 21(2):233–255.
- Durrant GB, Steele F. Multilevel modelling of refusal and non-contact in household surveys: evidence from six UK Government surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2009; 172(9999)
- Groves, RM.; Couper, M. *Nonresponse in Household Interview Surveys*. New York: Wiley; 1998.
- Hansen MH, Hurwitz WN. The problem of non-response in sample surveys. *Journal of the American Statistical Association*. 1946:517–529. [PubMed: 20279350]
- Harris LE, Weinberger M, et al. Assessing Inner-City Patients' Hospital Experiences: A Controlled Trial of Telephone Interviews versus Mailed Surveys. *Medical Care*. 1997; 35(1):70–76. [PubMed: 8998204]
- Hochstim JR. A Critical Comparison of Three Strategies of Collecting Data from Households. *Journal of the American Statistical Association*. 1967; 62(319):976–989.
- Manfreda KL, Bosnjak M, Berzelak J, Haas I, Vehovar V. Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*. 2008; 50(1): 79–104.
- McHorney CA, Kosinski M, et al. Comparisons of the Costs and Quality of Norms for the SF-36 Health Survey Collected by Mail versus Telephone Interview: Results from a National Survey. *Medical Care*. 1994; 32(6):551–567. [PubMed: 8189774]
- McMorris BJ, Petrie RS, et al. Use of Web and In-Person Survey Modes to Gather Data From Young Adults on Sex and Drug Use: An Evaluation of Cost, Time, and Survey Error Based on a Randomized Mixed-Mode Design. *Eval Rev*. 2009; 33(2):138–158. [PubMed: 19029360]
- Pickery J, Loosveldt G. A Multilevel Multinomial Analysis of Interviewer Effects on Various Components of Unit Nonresponse. *Quality and Quantity*. 2002; 36(4):427–437.
- Purdon S, Campanelli P, Sturgis P. Interviewers Calling Strategies on Face- to- Face Interview Surveys. *Journal of Official Statistics*. 1999; 15(2):199–216.

- Rodriguez HP, von Glahn T, et al. Evaluating patients' experiences with individual physicians: A randomized trial of mail, internet, and interactive voice response telephone administration of surveys. *Medical care*. 2006; 44(2):167. [PubMed: 16434916]
- Siemiatycki J. A comparison of mail, telephone, and home interview strategies for household health surveys. *Am J Public Health*. 1979; 69(3):238–245. [PubMed: 420369]
- Smyth JD, Dillman DA, et al. Using the Internet to Survey Small Towns and Communities: Limitations and Possibilities in the Early 21st Century. *American Behavioral Scientist*. 2010; 53(9):1423–1448.

Figure 1.
Time Sequence of Experiment

Table 1

Sample Size, Interviews, Response and Eligibility Rate by Treatment

	Mail-FtF	FtF-Mail-FtF	FtF
Sample Size	922	1,168	8,161
Completed Screening interviews	553	652	4,514
FtF	409	578	4,514
Mail	144	74	0
Second Ph Weighted Response Rate	0.84	0.86	0.85
Fully Weighted RR	0.94	0.94	0.92
Second Ph Weighted Eligibility Rate	0.19	0.15	0.18
Fully Weighted Elig Rate	0.12	0.14	0.13

Table 2

Screening Costs, In-Depth Interview Rate, and Overall Cost by Treatment

	Mail-FtF	FtF- Mail-FtF	FtF Only
Mail Cost Per Sampled Unit	\$4.28	\$4.24	NA
Mail Cost Per Complete	\$27.19	\$35.08	NA
Total Screening Cost Per Sampled Unit	\$91.72	\$102.03	\$93.11
Total Screening Cost Per Complete	\$132.97	\$152.01	\$147.14
In-Depth Interviews	126	137	922
Second-Phase Weighted In-Depth Interview Rate	.67	.73	.67
Fully Weighted In-Depth Interview Rate	.70	.78	.68
Overall Cost Per In-Depth Interview	\$1,018.77	\$1,194.89	\$1,180.73