

MEMO

To: The National Election Study Board of Overseers

From: The National Election Study Ad Hoc Committee on Survey Mode

Members: Norman Bradburn, University of Chicago
Charles Franklin, University of Wisconsin
Graham Kalton, Westat
Jon Krosnick, Ohio State University (chair)
Merrill Shanks, University of California – Berkeley
Sidney Verba, Harvard University

Re: The State of Scientific Knowledge on the Advantages and Disadvantages of Telephone vs. Face-to-face Interviewing

Date: May 28, 1999

At the request of the Board of Overseers of the National Election Study, our committee was formed to consider the possible advantages and disadvantages of shifting principal data collection for the NES from face-to-face to telephone interviewing. To facilitate deliberations, two papers were written for the committee. Melanie Green and Jon Krosnick wrote a review of literature on mode differences in social desirability response bias and survey satisficing and reported new evidence on these issues. And Charles Ellis and Jon Krosnick write a literature review and conducted a quantitative meta-analysis of the existing literature on the differences between respondent samples generated by traditional block-listing methods and those generated by RDD methods.

Once these papers were written and distributed to committee members, Jon Krosnick conducted individual interviews with the other committee members to ascertain their views on the various issues of interest. And a conference call was held during which Jon summarized individuals' views, and a discussion of the emerging issues was held. This memo summarizes the committee's collective views on the issues that it considered.

The committee's deliberations were directed by two possible shifts that NES might make in its data collection methods in efforts to reduce costs. The most dramatic step would be to shift all interviewing to the telephone and to sample American households via RDD or list assisted telephone number based methods. A less dramatic step would be to continue sampling households by the traditional block listing method but to conduct some portion of the interviews by telephone. The questions addressed below about mode differences therefore consider (1) the impact of a shift in sampling to RDD methods, and (2) the impact of conducting an interview by telephone, rather than face-to-face.

The memo is organized in nine principal sections, each one addressing a particular question that the committee attempted to answer, involving: sample quality, response rates, interview length, respondent satisfaction, social desirability response bias, satisficing, responses to open questions, show cards, and respondent misunderstandings. We end with a brief discussion of some additional miscellaneous issues that emerged in our discussions and a recommendation.

Sample Quality

The first question the committee considered was whether the quality of samples generated by block listing and RDD differ from one another. Although a substantial number of published studies might appear at first glance to have compared the quality of samples obtained by block listing and RDD methods, only a small set of them do indeed offer the potential for such insight. In order for a study to provide such information, it must have implemented block listing and RDD sampling side by side, using approaches as comparable as possible. But many studies that compared telephone and face-to-face methodology in fact sampled all households by block listing methods and then assigned respondents to either face-to-face or telephone interviewing. Only the studies that in fact compared block listing to RDD sampling are relevant to address this question.

In order to assess sample quality, one must select a set of variables as criteria. We did so by focusing on demographic characteristics, for two reasons. First, such characteristics of sample respondents can be compared to census data to assess accuracy. And second, we presumed that reports of demographics would be relatively consistently accurate across mode of interview. Therefore, any observed differences between an RDD sample interviewed by telephone and a block listing sample interviewed face-to-face would be attributable to the sampling method and not the method of interviewing.

Our review of the existing studies in this regard indicates that RDD samples are generally less representative of the population than block listing samples. In particular, as compared to block listing samples, RDD samples appear to under-represent low income, low education, and racial minority respondents. And the block listing samples are generally closer to census figures for the population than are the RDD samples.

In general, there are two reasons why block listing and RDD samples might differ: non-coverage and non-response. Non-coverage refers to segments of the population not within the sampling frame. There is some non-coverage even for block listing samples, which is inherent in the current NES approach. But moving to RDD sampling would introduce a new type of non-coverage: residents of households without telephones. And one reason for the gap in accuracy between block listing and RDD samples is the fact that low income, low education, and racial minority individuals are especially likely not to have a working telephone in their homes.

Non-response refers to refusals of some sampled individuals to be interviewed, and this can vary depending upon sampling mode. In particular, members of households sampled by block listing are approached on the doorstep, whereas respondents sampled

by RDD are initially contacted over the telephone. Different sorts of people may refuse to be interviewed in these two settings. And in fact, a second reason why block listing samples are more representative than RDD samples is the fact that low income and racial minority respondents are more likely to refuse to participate in a survey when contacted initially by telephone than when contacted face-to-face initially.

For the many investigators interesting in studying differences between low income, low education, or racial minority respondents and higher income, more education, non-minority respondents in terms of political attitudes and behavior, this sampling bias may be quite a significant cost inherent in a shift to RDD sampling. The under-represented groups of respondents are relatively small in the population to begin with, and shrinking their representation in a survey sample would then make it more difficult to reliably assess their other characteristics and compare them to other groups.

One possible solution to this problem might be statistical weighting of data points to yield results more closely matching the population. The committee recognizes that this is a widely-used strategy for this purpose and recognizes that it is really the only option available to an investigator in order to address this problem *ex post facto*. But the committee is less sure of the extent to which weighting would indeed improve a sample's quality with regard to variables not employed in the weighting algorithm. We therefore believe that the possibility of weighting should not be used as a basis for concluding that apparent sample quality differences between block listing and RDD samples can be ignored.

In conclusion, then, the committee views the apparently reliably greater biases in RDD samples than in block listing samples as a notable cost of shifting to RDD sampling for the NES.

Response Rates

Response rates for RDD telephone surveys have been falling in recent decades, whereas response rates for block listing face-to-face surveys, including the NES, have not been dropping in any significant way. Furthermore, even with the greatest effort expended to maximize response rates, RDD telephone survey response rates are likely to be approximately 15% lower than that for face-to-face surveys.

Recently reported evidence from a number of studies suggests that this 15% difference in response rate may not notably compromise sample representativeness. Furthermore, working hard to convince reluctant respondents to be interviewed may yield data of unusually poor quality, because these individuals may not be highly motivated to answer questions carefully and precisely. Nonetheless, a response rate represents one benchmark used by many scholars to assess the quality of a survey. Therefore, lower response rates represent an additional likely cost of telephone surveys.

In recent years, survey data collection firms have been experimenting with various procedures to increase response rates of RDD telephone surveys. For example,

using reverse directories, it may be possible to obtain names and addresses for as many as 60% of selected households and mail them letters in advance, informing them about the study and telling them that they will be receiving a phone call. In addition, financial incentives might be used to improve response rates. Published studies document that such approaches can indeed improve response rates. Nonetheless, the committee suspects that even extensive efforts along such lines are unlikely to yield response rates as high as those for face-to-face surveys.

Interview Length

NES interviews often average substantially longer than an hour, and the total time needed to conduct the pre-election and post-election interviews is typically 130 minutes on average. Yet conventional wisdom among experienced telephone interviewing organizations is that telephone interviews lasting longer than 30 minutes are very difficult to accomplish. In fact, the Gallup Organization has a general principle of keeping telephone interviews shorter than 12 minutes unless very unusual circumstances occur. This is based upon the presumption that the intrusion in people's daily lives represented by a longer telephone interview substantially threatens sample and data quality. In particular, respondent fatigue and the potential for break-offs seem likely to rise as telephone interviews lengthen. Furthermore, response rates may be lower for long RDD telephone interviews, because interviewers' morale may flag when they know interviews will be long and respondents may become frustrated with the length.

The committee believes that little if any reliable evidence exists with which to assess the validity of these concerns about interview length. A number of telephone studies have been implemented with quite long interviews, including some experiences by NES itself. But data from these experiences have not been reported in ways that would permit an assessment of compromises in data quality due to interview length specific to the telephone mode. Furthermore, members of the committee who have attempted to conduct long telephone interviews in their own research have found them to work badly in general.

The committee shares a concern that long interviews may be more difficult to accomplish on the telephone than face-to-face, but we also feel that data are needed to test the validity of this concern.

Respondent Satisfaction and Willingness to be Reinterviewed

Also of concern for the NES is respondent willingness to participate in reinterviews after participating in an initial interview. Over the years, the NES has conducted a number of important panel studies, and the validity of data obtained from such studies hinges on a sufficiently large proportion of respondents being willing to be reinterviewed.

Existing studies that have compared respondent enjoyment of telephone versus face-to-face interviews indicate that respondents enjoy the latter in substantially larger

numbers than the former. This therefore represents an additional potential cost of moving from face-to-face to telephone interviewing.

Social Desirability

Another issue considered by the committee sets aside issues of sampling and shifts focus to the impact of the mode of data collection on respondent behavior during the interview. In particular, we explored whether people are more or less willing to reveal socially undesirable facts about themselves in telephone interviews than in face-to-face interviews.

Many observers presume that respondents will be more honest in telephone interviews than in face-to-face interviews. This reasoning is based upon the presumption that respondents misrepresent themselves in survey interviews in order to cultivate favorable impressions of themselves in the eyes of their interviewers and in order to avoid any subtle disapproval interviewers might communicate if an embarrassing fact were to be revealed by a respondent. Presumably, an interviewer in the same room with the respondent can communicate disapproval non-verbally and is therefore more potentially threatening than a distant interviewer at the other end of a telephone wire.

Surprisingly, however, the accumulated evidence suggests just the opposite. A number of studies have compared people's willingness to report socially embarrassing facts about themselves in face-to-face and telephone interviews. These facts include not having voted in the last election, using illegal drugs, and being racially prejudiced. Some studies have found no reliable differences between modes in willingness to report such characteristics. But more such studies have found more honesty in face-to-face than in telephone interviews, and a meta-analysis of 25 studies showed this to be the reliable general tendency in the existing literature.

One possible explanation for this finding involves the identity and trustworthiness of the interviewer. In telephone interviews, the identity of the interviewer cannot be known for sure and cannot even be independently validated by the respondent (unless letters are sent to respondents in advance alerting them to the upcoming telephone call). This may lead to a slight hesitation to completely trust the interviewer, which may lead to a hesitation to report potentially incriminating facts. In contrast, a face-to-face interviewer can produce written materials to confirm his or her identity. And a face-to-face interviewer can most likely develop a greater sense of rapport with the respondent by smiling and performing other behaviors that cannot be observed if performed by a telephone interviewer.

This therefore represents a potential cost of a shift to telephone interviewing from face-to-face interviewing.

Satisficing

Another issue of interest to the committee was the question of whether respondents invest less effort to provide valid self-reports on non-sensitive topics during telephone interviews than during face-to-face interviews. The concern here stems from the notion that telephone respondents may be less motivated by their interviewers than face-to-face respondents, whose interviewers can illustrate their own professionalism and commitment to high data quality by their non-verbal behavior. Also, if interviewers are indeed able to generate better rapport face-to-face than over the telephone, this rapport may also serve to motivate respondents to be thoughtful and careful in generating answers, rather than simply generating thoughtless answers in order to complete the interview as quickly as possible. And interviewers and respondents alike seem more likely to be comfortable with longer silences and a slower pace in face-to-face conversation than in a conversation over the telephone, which is likely to speed up the pace of telephone interviewing and therefore make careful deliberation more difficult for respondents to accomplish. Consequently, telephone respondents may be more likely to satisfice rather than optimize in generating self-reports.

A number of studies exist with which to compare telephone and face-to-face interviewing in terms of satisficing. Specific manifestations of satisficing studied include acquiescence response bias, the selection of "don't know" response options, and mental coin flipping. And across these studies, reliable patterns indicate more satisficing in telephone interviews than in face-to-face interviews. This therefore represents an additional cost of moving from face-to-face to telephone interviewing.

Responses to Open Questions

One set of widely analyzed and important questions in the NES have asked respondents to describe the good and bad points of candidates for public office and of political parties. It is now well-established that respondents answer such open-ended questions more briefly and in less detail when interviewed over the telephone than when interviewed face-to-face. This therefore represents an additional potential cost of telephone interviewing.

Show Cards

A final principal issue considered by the committee is the use of showcards. Many of the most often analyzed NES core items have involved presentation of response choices on showcards, including the thermometers and 7-point scale policy preference questions. The use of these showcards has been motivated by the belief that they substantially improve respondent ability to answer these questions accurately.

A move to telephone interviewing would make use of showcards more difficult, in at least two ways. First, although it is possible to get showcards to respondents prior to telephone interviews (either by mailing them or by dropping them off at the doorstep), such a procedure does not assure that all respondents will have the showcards available at

the time of the interview. For example, a recent NES experience indicated that only 70% of the respondents who received showcards on their doorsteps had them available at the time of later telephone interviews.

A second problem with this is that respondents must turn the pages of a showcard booklet themselves, rather than having an interviewer present to assure that this process is done properly. Some mistakes made by respondents in handling the showcards may be detectable by telephone interviewers, but some may not. And this may be especially challenging for respondents with more limited literacy.

One way to deal with this problem would be to rewrite the items currently using showcards to be administerable without them. Indeed, some committee members see potential advantages to such shifts, believing that current items formats (e.g., 7-point policy scales, thermometers) are not optimally designed to measure attitudes most reliably and validly. However, it is important to note that rewriting some NES items to avoid the need for showcards would involve such dramatic changes so as to render the new question incomparable to the old one.

But even in cases where the rewriting is relatively simple, such a shift in presentation approach from visual to oral is very likely to shift the nature of systematic measurement error due to response choice order. A great deal of evidence now documents that the order in which response choices for closed-ended questions are presented to respondents influences their answers. Furthermore, these effects appear to be the result of satisficing and vary in nature depending upon mode of administration. Nearly all documented primacy effects in categorical questions occurred under visual presentation of response choices, and nearly all documented recency effects in such questions occurred under oral presentation. Therefore, dropping the use of showcards for some items is likely to reverse the direction of systematic bias due to this form of satisficing.

Detecting Respondent Misunderstanding or Difficulty

In face-to-face interviews, interviewers can see nonverbal, visual cues from respondents signaling confusion or misunderstanding on their part, the need to take a short break in the interviewing, or other such phenomena. These cues can include furrowed brows, looks of puzzlement, looks of exhaustion, and so on. In the face of such cues, face-to-face interviewers can be responsive and tailor the interaction to maximize interview completion and data quality. In a telephone interview situation, however, interviewers cannot see such cues, so they cannot adjust their behavior to suite respondents' needs as well. In the case of long and complex interviews, such as the NES, this may represent a significant disadvantage of telephone interviewing as compared to face-to-face interviewing, although no data currently exist with which to assess the validity of the committee's suspicion on this point.

Other Issues

A few additional issues emerged in the committee's deliberations, and we list them below.

Time series comparability. A number of committee members have extensive experience with studying long-term trends over time using survey data. And their experience indicates that significant shifts in data collection methods have generally introduced substantial uncertainty into this sort of scholarship that is very difficult if not impossible to overcome. The committee therefore believes that methodological shifts of the magnitude considered here (either from block listing sampling to RDD, or from face-to-face to telephone interviewing, or both) should be seen as starting a different, new study of time trends, rather than continuing the existing study.

The committee noted that administrators of the General Social Survey are not currently considering a move to telephone interviewing. This survey series is much like the NES, in that it is designed to track social change in beliefs, attitudes, and behavior over time, and it does so via interviews typically lasting 90 minutes or longer on average. The scholars shepherding the GSS believe that the mission of the study to track trends would be seriously compromised by a mode shift, so it has not been considered.

The 1996 NES mode experiment. One set of information that the committee considered when assessing the impact of interviewing mode on response quality was from the experiment conducted in 1996 by the NES, randomly assigning respondents to be interviewed either face-to-face or by telephone. After careful consideration, the committee concluded that this study was not optimally informative for the issues under current consideration, because of the respondent selection procedure used.

In short, respondents who participated in this experiment had been interviewed at least once previously, and some had been interviewed a number of times previously. Furthermore, they were not a representative sample of the nation but rather represented a subset of such a sample who could be successfully recontacted and agreed to be interviewed more than once. All this means that the respondents were experienced with NES-style instruments and were likely to have been unusually motivated to provide high-quality data. If a fresh, fully representative cross-section had been interviewed, the committee suspects that mode effects would have been more apparent. For this reason, we did not place significant weight on this study in reaching our final conclusions.

If the NES Board is considering making a permanent move to a blending of telephone and face-to-face interviewing following the selection of households via block listing, the committee feels that there is sufficient convincing evidence available to justify concern about a number of potential drawbacks, as outlined above. Specifically, data quality seems likely to be lower in the telephone interviews than in the face-to-face interviews.

However, this conclusion is based upon studies done by different survey organizations for different purposes on a range of different topics. It is conceivable that the drawbacks of telephone interviewing may be less for the NES than for these other surveys. The committee therefore recommends to the Board that it consider seeking funds for a new mode experiment, to more effectively assess the impact of such a mode shift for the NES in particular.

The purpose of this study would be to assess whether telephone interviewing *per se* induces more satisficing, more social desirability response bias, less data accuracy, and different patterns of unit non-response than does face-to-face interview *per se*. The design of this study would resemble that of the 1996 NES experiment in some ways. Households would be selected through block listing with household enumeration and respondent selection being done on the doorstep. Then, selected respondents would be randomly assigned to interview mode. However, two important differences could be incorporated into the methodology. First, the sample of respondents could be composed of people never before interviewed for the NES, to permit confident generalization of findings to such fresh samples. And second, the questionnaire could build in experiments explicitly designed to assess the magnitude of satisficing and social desirability bias and to permit record checks to validate factual reports. As a result, strong conclusions can be reached about any impact of interview mode *per se* on data quality.

One way to accomplish this experiment would be to seek funds to supplement the 2000 NES. Specifically, the National Science Foundation Measurement, Methods, and Statistics program and or private foundations could be approached seeking funds to permit the basic NES to be done purely via face-to-face interviewing. In addition, funds would be used to permit fielding additional interviews to be done by telephone. Each selected respondent would be randomly assigned to be interviewed either face-to-face (for the 2000 NES main study) or by telephone (for this methodological experiment).

Voters vs. no-voters. Analyses of the 1982 MCP data suggest that the responses of voters are less susceptible to mode effects than the responses of non-voters. Of course, the NES is equally concerned with understanding the beliefs, attitudes, and behavior of both groups of respondents. But it is useful to note that analyses focussed on voters appear less likely to be biased by mode effects.

Sample size. One additional consideration raised by the committee has to do with sample size. For many NES analysts, sample sizes have shrunk in recent years to numbers of cases that severely limit one's ability to yield reliable results from complex multivariate analyses. It would therefore be preferable to increase sample sizes beyond those of recent NES surveys. If the funds available to run the NES were to be held constant in the future, a shift to telephone interviewing would permit a greater number of people to be interviewed. This would be a significant advantage of telephone interviewing.

Geographical clustering. An advantage of the RDD telephone approach is that it avoids the geographical clustering usually implemented in block-listing studies to cut costs and maximize efficiency.

Interviewer supervision. Another advantage of telephone interviewing is the greater ability it allows for close interviewer supervision. That is, supervisors can listen in on conversations between interviewers and respondents and provide immediate feedback to interviewers in efforts to improve their administration behavior and maximize standardization of administration across interviewers.

Publication of the committee's informational memos. A number of committee members expressed the view that the Krosnick and Green memo and the Ellis and Krosnick memo integrate and review the existing literature in ways that yield conclusions of which many survey researchers are unaware. It would therefore be useful to attempt to publish those memos in top quality outlets at the earliest possible time, to help scholars who have not thought a great deal about the issues addressed to get in better touch with the existing evidence. The committee suggested that some particular changes be made to the memos, and those changes should obviously be made prior to submission of them for publication.

Conclusion

Although the committee sees a basis for expecting a number of differences between the quality of data obtained by telephone and face-to-face interviewing, the magnitudes of these differences can ultimately be weighted against cost in making a judgment about which mode is most appropriate. We have not sought information on costs, so we have not attempted to judge this tradeoff for the Board. Furthermore, to do so would require more precise estimates of the magnitudes of data quality decrements that would result from a switch to telephone interviewing in the NES than are currently available. This is an additional justification for the notion of doing a fresh mode experiment in the future.

The committee appreciates the opportunity to provide these opinions to the Board and hopes we have been helpful to the Board for its deliberations. We would be happy to elaborate upon or clarify the opinions expressed herein if that would be useful, and we wish the Board the best of luck in grappling with these important issues.