

MEMORANDUM | March 31, 2010

TO Terell Lasane
FROM Nora Scherer, Tracy Dyke-Redmond
SUBJECT Literature Review Summary

SUMMARY OF ACTIVITIES

We have focused our initial literature review on the first evaluation question from the Evaluating Outcomes from Compliance Assistance work plan:

Is the telephone survey a valid and reliable technique for performance measurement and program evaluation?

Specifically, we concentrated on the following sub-question:

What studies have been conducted to assess the reliability and/or validity of phone surveys as a way to understand behaviors, in the context of requirements?

We conducted a thorough search of journals, databases, and academic institutions to look for relevant articles, books, or papers. Our search included databases such as Scirus, Science Direct, Dialog, EBSCO, and Ingenta. We also looked for relevant publications on federal (e.g., U.S. EPA) and state websites, on industry sites such as the American Statistical Association, and the American Evaluation Association. Finally, we searched evaluation journals, a number of university websites, and cross-checked citations from relevant publications. We found a variety of sources that generally fall into two categories:

1. **Theoretical Discussions:** This literature focused on a more theoretical discussion about different modes of survey administration. In general, these articles provide lessons and thoughts about the advantages and disadvantages of different approaches (e.g., mail, telephone, and face-to-face). Some also look at existing studies to draw bigger picture lessons (i.e., meta-analyses).
2. **Experimental Comparisons:** We found several articles that specifically tested for different measures of data quality in different modes of survey administration. These studies either use existing data, or in some cases collected new data as part of the study. They compare such factors as data quality, reliability, and response rates for different modes of survey administration.

The attached annotated bibliography has the most relevant pieces of literature that we found, separated into these two categories.

GENERAL FINDINGS

In conducting our initial, broad literature search, we found that there are several articles that compare the different modes of survey administration, but that few of those actually attempt to capture or measure those differences. We included a selection of these

“theoretical” articles, but there are several more we have not included in order to avoid repeating the same conclusions.¹ In addition, we found that several of these articles are a bit dated; most were written in the late 1970s and 1980s. This seems to be a result of the fact that telephone surveying techniques were becoming a much more popular alternative to face-to-face interviews during that time. Most of the current research (from the 1990s on) seems to focus on the emergence of internet survey techniques, and the increase in cell phone and autodialer usage, both of which are irrelevant for our purposes.

For the studies that do actually measure differences between the modes (the “experimental” studies), we found only one that focuses on compliance or compulsory behaviors: a study conducted by the U.S. EPA’s Office of Compliance. This study compared compliance and facility behaviors from a mailed survey to an on-site survey. Instead, most of the studies consider opinion surveys and surveys that attempt to measure behaviors of individuals. However, these studies do provide important lessons about the validity of different survey approaches by trying to verify the reported information from different survey modes.

While the results were somewhat varied, in general the authors found that overall there is not a substantial difference between telephone and face-to-face surveys. Any observed differences can generally be corrected by thoughtful and creative survey design. For example, if a face-to-face interview includes a visual aid to help respondents answer a question about types of equipment used in a facility, a telephone interview obviously cannot include this same visual. Therefore, survey developers will have to come up with another way to communicate this question, such as careful description of the equipment (e.g., think of easily recognizable names for the types of equipment).

Some of the differences the authors did find include:

- Face-to-face interviews have a higher response rate;
- Telephone interviews take fewer resources and less time to complete;
- Some respondents misreport their behaviors in telephone interviews (e.g., young adults were found to be more likely to underreport their smoking behaviors in a telephone interview than they were in a face-to-face interview);
- Telephone respondents are more likely to not respond to individual questions, or to give more socially desirable responses; and
- Generally, face-to-face surveys gather better quality data, but these differences are often small.

The greatest difference between modes appears to be between those with or without an interviewer (i.e., mail versus telephone or face-to-face). The presence of an interviewer seems to make a bigger difference, compared to whether or not that interviewer is on the phone or in-person.

¹ We also found some additional sources that we will most likely reference in a more thorough write-up of these issues, but we haven’t included them in the full annotated bibliography.

Overall, we find that the literature generally supports the conclusion that telephone surveys can provide accurate responses and are a valid way to collect survey data. However, we note that the literature is relatively sparse on compliance specific issues. In addition, as noted, several of the studies focus on individuals, not facilities. We suggest using the information provided here in support of the findings of the statistically valid pilot project, with the appropriate caveats.

Evaluating Outcomes from Compliance Assistance Literature Review - Annotated Bibliography

THEORETICAL ARTICLES

Van der Zouwen, Johannes and de Leeuw, Edith D. "The Relationship Between Mode of Administration and Quality of Data in Survey Research." *Bulletin of Sociological Methodology*, Vol. 29, No. 3, 1990.

The authors of this article attempt to analyze the differences between the quality of data collected from surveys conducted either over the phone, by mail, or in person. The question is whether or not the same survey given in one of three ways (i.e., by telephone, mail or face-to-face) will achieve similar results in terms of representativeness, completeness, and in the similarity of responses. In other words, because the surveys were the same, any differences captured between these surveys can be attributed to the mode of administration. They take two approaches to analyze these potential differences ("mode effects"): (1) they perform a meta-analysis on the outcomes of dozens of experiments that tested for differences between modes of administration, and (2) they conduct a field survey to test for the differences between responses from a mail, telephone, and face-to-face survey.

The meta-analyses found small but statistically significant, (and consistent) effects of the mode of administration. They found that on several indicators of data quality, the face-to-face interview scores slightly better than the telephone interview, but these differences are gradually disappearing over time. Mail surveys usually have lower response rates and less item response than face-to-face interviews and telephone interviews. Mail surveys result in more accurate, less biased answers than both forms of interviewing, especially when sensitive or embarrassing questions are being asked.

The field survey results showed that the telephone and face-to-face surveys had a significantly higher response rate than the mail survey, the face-to-face survey had the highest completion rate, and that similarity of responses depended on the type of question (e.g., whether the question was "sensitive" or not).

De Leeuw, Edith Desiree. "Data Quality in Mail, Telephone and Face to Face Surveys." Netherlands Organization for Scientific Research, 1992.

This book takes a comprehensive and comparative look, with respect to data quality, at three modes of survey administration: face-to-face interviews, telephone interviews, and mail questionnaires. The author achieves this using three primary approaches: (1) review the literature on experimental comparisons of data collection methods, (2) examine the effects of data collection mode on various aspects of data quality, and (3) examine the effects of data collection mode on research results.

The author found that a meta-analysis of experimental comparisons of these data collection methods detected small differences in data quality, suggesting a dichotomy of survey modes: modes with and modes without an interviewer. None of the modes was superior on all criteria (e.g., response validity, item nonresponse), and modes with an interviewer resulted in higher response rates and lower item nonresponse, but also produced more socially desirable answers. In general the author found that:

- Face-to-face surveys tend to obtain higher response rates than comparable telephone surveys;
- Telephone interviews are less flexible than face-to-face interviews;
- Face-to-face surveys can generally be longer than telephone as it is often more difficult for respondent to end an in-person interview; and
- Telephone surveys require fewer resources and are the fastest to complete.

The field experiment showed that the face-to-face survey resulted in the lowest response rate. The study on research results found that there were no consistent differences between the face-to-face and telephone interview with respect to item nonresponse and self-disclosure on sensitive topics. Overall, the author found no detectable difference between telephone and face-to-face interviews.

Dillman, Don A. "Mail and Telephone Surveys: The Total Design Method." New York: Wiley, 1978.

The author of this book has developed what he calls a "Total Design Method" for conducting mail and telephone surveys that he believes makes these survey administration modes competitive with, if not equal to, the face-to-face interview. The method helps researchers identify each aspect of survey administration that may affect response quantity and quality, by focusing on a theory of response behavior and developing administrative plans to direct survey implementation.

Although the author focuses on surveys gathering information about respondent's opinions or behaviors, he finds very little difference between telephone and face-to-face interviews overall. In fact, he finds that face-to-face interviews rank higher than telephone interviews in seven performance characteristics, but telephone interviews rank higher than face-to-face interviews in eight dimensions.

Essentially, he concludes that each mode has its pitfalls and benefits, and that by recognizing that, researchers can design any of the modes to provide accurate and cost-effective surveys.

Tyebjee, Tyzoon T. "Telephone Survey Methods: The State of the Art." *The Journal of Marketing*, Vol. 43, No. 3, Summer 1979, pp. 68-78.

The purpose of this paper is to identify the major methodological issues which characterize telephone survey research and present a summary of the conclusions of studies which have addressed these issues. The major issues in telephone survey research are partitioned into and discussed in the following five broad categories: research management, data validity, response rates, sampling, and questionnaire design. The author concludes that the quality of data collected by the telephone is comparable to the same information collected by personal interviews or mail questionnaires. He posits that there are four considerations to make when using telephone surveys: (1) telephone interviewing cannot grant the respondent the same anonymity as mail questionnaires, (2) only audio communication is possible over the telephone (i.e., no visual aids), (3) when observations by the interviewer are important, personal interviews are more suitable than telephone interviews, and (4) when the research is concerned with poverty-level and rural groups, a significant portion of these groups will be excluded by telephone surveys because they don't have telephones.

Jackle, Annette, Roberts, Caroline, and Lynn, Peter. "Assessing the Effect of Data Collection Mode on Measurement." Institute for Social & Economic Researchm ISER Working Paper Series, No. 2008-08, February 2008.

This study discusses the comparability of data collected in different survey modes, and some of the difficulties in evaluating whether mixing modes affects measurement and hence data comparability. Some of the difficulties include: (1) the need to avoid confounding effects, (2) the sensitivity of conclusions to methods of analyzing experimental mode comparison data, (3) the difficulty of assessing whether measurement differences matter in practice, and (4) the assessment of which mode provides better measurement. They conclude that it is extremely difficult to devise mode comparisons such that any differences in responses can clearly be attributed to the effect of mode on measurement. Moreover, even if appropriate methods are used to test for differences in responses which might be attributed to mode, this does not answer the question of whether these differences would matter in practice.

Bonnel, Patrick, and Le Nir, Michael. "The Quality of Survey Data: Telephone versus Face-to-Face Interviews." *Transportation*, Vol. 25, No. 2, May, 1998.

This paper compares the performance of telephone and face-to-face interviews with respect to several aspects of survey method: the representativeness of the sample, and the accuracy of data. The authors use results from several existing studies to conduct their comparison. Findings include:

- *Response Rate*: Appears to be lower with telephone interviews, but the difference is tending to diminish. In addition, with careful

survey design, telephone interviews can obtain response rates as high as in-person.

- *Data Accuracy:* Any differences found between the two modes are extremely small and statistically insignificant.
- *Quality of Responses:* Generally, face-to-face interviews provide better quality information; however, the difference between the two modes seems to diminish as telephone survey methods are refined.

EXPERIMENTAL ARTICLES

U.S. EPA, Office of Enforcement and Compliance Assurance. "Guide for Measuring Compliance Assistance Outcomes." Revised October 2007, EPA 300-B-07-002.

OECA publishes this guide to help states measure the goals they have articulated in their Performance Partnership Agreements (PPAs). Section VI provides an example of OECA's experience with the Dillman Method. In 2001, the Office of Compliance (OC) conducted a study to test methods for collecting outcome data from compliance assistance efforts by comparing results from a mailed survey (using the Dillman total design method) and on-site surveys. The purpose of this analysis was to test the hypothesis that the results of both data collection methods would be the same. For this study, they focused on metal finishers and marinas in Regions 1 and 5. Response rates for the metal finishing shops were less than 50 percent for both modes (43.6 percent for the mailed survey and 40.7 percent for the site visits). On the other hand, the marinas achieved a much higher response rate for the on-site visits (51.3 percent for the mailed survey and 94.6 percent for the on-site visits). They hypothesize that this difference results from a higher level of distrust for the agency in the metal finishing sector. They also tested for statistical differences between the responses gathered from the two different modes. They found that for the metal finishing facilities:

- For the most part, the two samples were identical in terms of general characteristics;
- A larger proportion of visited facilities acknowledged receiving wastewater compliance assistance compared to the mailed survey facilities;
- There were no significant differences between the two samples in terms of facilities understanding of regulatory requirements; and

- Mailed survey responses do not provide a biased estimate of performance towards key environmental regulations but there may be some concern for the air-related questions.

For the marinas, however, they found that mailed respondents indicated significantly higher compliance than the on-site observations found. Therefore, the observed differences between reported and observed compliance calls into question the validity of obtaining reliable compliance information for this sector through a mailed survey.

Rafferty, Ann P. et al. "Validity of a Household Gun Question in a Telephone Survey." *Public Health Reports*, Vol. 110, No. 3, May-June 1995, pp. 282-288.

In an effort to determine the validity of self-reported data on the presence of guns in the home obtained in a telephone survey, the authors of this article conducted a survey of households where a hunting license had been purchased or a handgun registered. They found that the proportion of respondents who reported that at least one gun was kept in their households was 87.3 percent among handgun registration households and 89.7 percent among hunting license households. If they assume that all of the households that bought a hunting license and all of the households that registered a handgun were in possession of a gun, then they conclude that 11.4 percent of the responses were invalid. In other words, 11.4 percent of the households with a hunting permit or registered handgun reported that there was no gun in their household. The authors conclude that despite some limitations, the data indicate that a question on gun presence in a household can return relatively valid responses.

Luepker, Russell V. et al. "Validity of Telephone Surveys in Assessing Cigarette Smoking in Young Adults." *American Journal of Public Health*, Vol. 79, No. 2, February 1989, pp. 202-204.

To better assess the reliability and validity of using telephone interviews to determine cigarette smoking rates in young adults, the authors compared results obtained from a telephone survey with those obtained from a face-to-face interview. They also validated the face-to-face interview with a saliva test for evidence of smoking. Interviews were conducted with young adults, ages 17 to 21 in the Minneapolis-St. Paul metropolitan area from seven area high schools. Participants were selected from a cohort of students followed annually to assess changes in smoking habits during young adulthood, and were randomly selected from four groups: non-smokers, smokers, and short- or long-term quitters. The authors found that agreement between the phone and home interviews was high for smokers and non-smokers; quitters tended to underreport their smoking (i.e., said they did not smoke during the phone interview but reported smoking at the home interview and/or the saliva test). Overall, they found that while most subjects were in agreement for their reported behaviors, the telephone methodology underestimated cigarette smoking rates by 3 to 4 per cent.

Reddish Douglas, Malinda et al. "Estimating the Proportion of Homes with Functioning Smoke Alarms: A Comparison of Telephone Survey and Household Survey Results." *American Journal of Public Health*, Vol. 89, No. 7, July 1999, pp. 1112-1114.

This study determined the proportion of homes with functioning smoke alarms in a low-income area experiencing a high rate of residential fire-related injuries. An on-site survey of households was conducted to confirm the results of a telephone survey. In the telephone survey, 71% of households reported having functioning smoke alarms. In the household survey, 66% of households reported having functioning smoke alarms; however, when the alarms were tested, the percentage dropped to 49%. The authors conclude that telephone surveys may overestimate the presence of functioning smoke alarms in some populations.

Greenfield, Thomas K. et al. "Effects of Telephone versus Face-to-Face Interview Modes on Reports of Alcohol Consumption." *Addiction*, Vol. 95, No. 2, 2000, pp. 277-284.

The authors of this study used two different survey administration modes, telephone and face-to-face interviews, to determine the effects of survey mode on alcohol consumption estimates. The primary purpose of the study was to compare the results of two different alcohol consumption surveys that were both conducted in 1990 and covered random samples from households across the entire U.S.: The 1990 National Alcohol Survey (which used face-to-face interviews) and the 1990 National Warning Labels survey (which used telephone interviews). To make the results from the two studies comparable, they defined "abstainers" as those who answered in either survey that they had not consumed alcohol in the last 12 months, and "drinkers" as those that had. They estimated the differences between the results of the two surveys by comparing the proportion of abstainers from each survey (although the surveys did not interview the same households, both samples were taken randomly from the entire population; therefore the results can be compared). They found that there was no statistical difference between the proportion of abstainers observed from the face-to-face interviews (36 percent) and the proportion observed from the telephone interviews (33 percent).

Kormendi, Eszter and Noordhoek, Johannes. "Data Quality and Telephone Interviews." Copenhagen: Danmarks Statistik, 1989.

The primary purpose of this study is to test the quality and reliability of information obtained in telephone interviews by examining the limitations and advantages of the method (compared to the face-to-face method), and how and under which conditions these limitations (or advantages) are likely to become apparent. The authors review two studies of differences between responses obtained from both methods: a major survey of the working environment and

absenteeism in the Danish Post and Telegraph Office, and interviews conducted during a multi-purpose survey in August 1984. They find that:

- Mode of interview appears to have had no influence on the quality of answers to questions that may be difficult to understand;
- The amount of recorded time for each interview type was not different between the two modes;
- No systematic difference could be found between the two groups with respect to missing or incomplete information; and
- Respondents in the face-to-face interviews had a greater tendency to provide supplementary information.

Overall, they conclude that: (1) open-ended questions require an extra degree of attention for telephone surveys, (2) telephone surveys elicit less underreporting of undesirable behaviors. They also note that telephone surveys are at a slight disadvantage because the interviewer is not able to show visual aids and cannot make eye contact with the interviewee, but these can be overcome with creative survey development.

Green, Melanie C. and Krosnick, Jon A. "Comparing Telephone and Face to Face Interviewing in Terms of Data Quality: The 1982 National Election Studies Method Comparison Project." The Ohio State University, August 1999.

In this paper, the authors report the results of a new set of analyses exploring differences in data quality across modes. They begin by offering a series of theory-grounded hypotheses about possible mode differences, and review what little evidence exists regarding their validity. Then, they report findings from an analysis of data from the 1982 National Election Study Method Comparison Project, an experiment designed to compare block-listed face-to-face interviewing with RDD telephone interviewing. Their focus is on three aspects of data quality: sample representativeness (gauged in terms of demographics), the amount of effort respondents devote to providing accurate answers (i.e., satisficing versus optimizing), and the extent to which people misportray themselves in socially desirable ways, rather than giving honest answers.

They find that the interview mode can affect both the sample representativeness and the response patterns observed in surveys. Responses from the telephone interviews showed more satisficing than the responses from the face-to-face interviews. In addition, telephone respondents said no opinion more often and showed an increased tendency toward socially-desirable responding.

Skumatz, Lisa A. "Mercury-Containing Thermostats: Estimating Inventory and Flow from Existing Residential & Commercial Buildings." Skumatz Economic Research Associates, Inc. December 28, 2009.

The authors of this study conducted two small-scale validation efforts to provide indicative data on the reliability of self-reported data for a study of the amount of mercury-containing thermostats in California: an on-site survey, and a camera study. In the first, they conducted 30 site visits centered around the Bay area to inspect the accuracy of the self-reported thermostat counts and types. The inspections found 10 percent misreported the number of thermostats. In the second effort, they mailed out 44 disposable cameras to a random sample of respondents state-wide. The sample was asked to take photographs of each thermostat in their home and return the camera, which would validate both count and type of thermostat. Overall, the combination of these two validation efforts finds that self-reports were correct in the majority of cases.

Jackle, Annette, Roberts, Caroline, and Lynn Peter. "Telephone versus Face-to-Face Interviewing: Mode Effects on Data Quality and Likely Causes - Report on Phase II of the ESS-Gallup Mixed Mode Methodology Project." Institute for Social & Economic Research, ISER Working Paper, 2006-41, August 2006.

This report presents findings from an experimental study carried out in the context of the European Social Survey, to assess the impact a change in data collection mode from the current face-to-face interviewing to telephone might have on data quality and to study the likely causes of any observed mode effects. The design included three comparison groups (two interviewed face-to-face (one with showcards, one without) and the third by telephone). They found evidence of effects caused by the presence of the interviewer, but few stimulus effects. They tested a number of hypotheses about the likely causes of mode effects on response, focusing on three forms of satisficing and social desirability bias. They found no evidence that using showcards influenced response quality, either positively or negatively. Unlike previous studies, they found no support for the hypothesis that telephone respondents were more likely to satisfice. However, consistent with expectations, they did find telephone respondents were more likely to give socially desirable responses across a range of indicators.