Annu. Rev. Sociol. 1991. 17:225–49

THE DESIGN AND ADMINISTRATION OF MAIL SURVEYS

Don A. Dillman

Washington State University, Pullman, Washington 99164-4014

KEY WORDS: survey research, measurement error, nonresponse error, response rate, noncoverage error, mixed mode surveys

Abstract

For reasons of cost and ease of implementation, mail surveys are more frequently used for social research than are either telephone or face-to-face interviews. In this chapter, the last two decades of research aimed at improving mail survey methods are examined. Discussion of this research is organized around progress made in overcoming four important sources of error: sampling, noncoverage, measurement, and nonresponse. Progress has been especially great in improving response rates as a means of reducing nonresponse error. Significant progress has also been made in finding means of overcoming measurement error. Because mail surveys generally present few, if any, special sampling error problems, little research in this area has been conducted. The lack of research on noncoverage issues is a major deficiency in research to date, and noncoverage error presents the most significant impediment to the increased use of mail surveys. The 1990s are likely to see increased research on mail surveys, as efforts are made to incorporate mail into cost-effective mixed mode designs.

INTRODUCTION

More sample surveys are done by mail than by any other survey method. An examination of all federal surveys approved by the US Office of Management

0360-0572/91/0815-0225$02.00

and Budget and active in August 1981 revealed that nearly 8 out of 10 utilized self-administered questionnaires. Of these 2137 surveys 69% were conducted solely by self-administered procedures, versus 2.2% solely by telephone, and 9.4% solely by face-to-face interviews. An additional 11% involved self-administered questionnaires in combination with one of the other methods. The report concluded that at least 90% of those categorized as self-administered surveys were implemented by using mail procedures (US Office of Management and Budget 1984).

Two major reasons for the extensive use of mail survey methods are, first, there is the much lower cost for completing them; second, procedures for mail surveys are often deemed simple enough that individuals and organizations conduct their own rather than relying upon survey research organizations. If viewed from the standpoint of large-scale general public surveys done through national survey organizations, there is little doubt that telephone surveys are now the nation's most used method for large-scale general public surveys. However, when surveys of all populations, including those implemented by professors, graduate students, government, schools, and businesses, as well as the national survey organizations are included, the prevalence of mail survey procedures is evident.

The literature on mail survey methods is voluminous. A bibliography in the mid-1970s on research to improve mail survey procedures included nearly 250 entries (Dillman 1978). A more recent bibliography, including only items published since 1970, compiled in preparation for the writing of this chapter, included more than 400 entries (Dillman & Sangster 1990).

Our purpose here is selectively and critically to review major developments in the design and administration of mail surveys during the last two decades, as they are conveyed by this literature. Important issues for future research are also identified.

Much but by no means all of the literature has been authored by sociologists. Sociology is only one of many disciplines that depend on data collected by mail surveys for the scientific advancement of the discipline. Statistics, psychology, marketing research, economics, and the various health sciences are disciplines in which research efforts to improve mail survey methods are regularly conducted and reported. In this review we are less concerned with the origin of particular research than its relevance to improving the quality and quantity of response to mail surveys, and in particular to resolving the major error issues. To these issues we now turn.

PERSPECTIVE TAKEN IN THIS REVIEW

The sample survey is distinguished from other research methods frequently used by sociologists by its ability to estimate quantitatively the distribution of a characteristic in a population, and to accomplish this by obtaining informa-

tion (answers to questions) from only a small proportion of that population, usually a few hundred or thousand. Thus, one can learn from a sample survey such things as what percent of a city's residents own the home in which they live, what proportion of state government employees are satisfied with their jobs, and the extent to which residents of large cities have greater fear of crime than do residents of small cities.

To generalize the results for the sample to the population from which it is drawn, one must normally contend with at least four potential sources of error, any one of which may make the survey results unacceptable. These sources of error include: (*a*) sampling error, (*b*) noncoverage error, (*c*) nonresponse error, and (*d*) measurement error (Groves 1989). Efforts to do quality mail, or for that matter telephone or face-to-face, surveys require that we attempt to eliminate, or at least reduce, all four types of error. Each kind of error describes a reason for the discrepancy between the population attribute measured and the estimate derived from the sample survey.

Sampling error results from heterogeneity on the survey measures among members of the population. It is attributable to the fact that certain members of the population are deliberately excluded by selection of the subset of members for which responses were obtained. In practice, sampling error is mostly a reflection of the number of respondents surveyed and provides the basis for statements that one frequently hears such as, "This randomly drawn sample of 400 respondents provides results that are accurate within 5 percentage points." In general, quadrupling sample size decreases sampling error by one half. Sampling error is the aspect of survey quality examined through inferential statistics applied to sample survey results, from which conclusions about significant differences in the population are obtained.

The principles for sampling were, for the most part, developed early in this century. They have presented particular challenges for face-to-face interviews where multistage cluster designs are typically used to reduce costs, and for telephone procedures where methods for efficiently selecting random digits to access households had to be developed. Mail surveys typically rely on lists for which simple random or simple systematic sampling procedures are used so that very little research on ways of reducing sampling error in mail surveys has been done. For this reason, sampling error receives little attention in this paper.

Noncoverage error arises because some members of the population are not covered by the sampling frame and therefore have no chance of being selected into the sample. Reducing this source of error is an important research topic for mail surveys. Noncoverage error is one of the major reasons that mail surveys have not been as useful as desired in surveying the general public. Because of the small amount of mail survey research in this area we shall consider it only briefly, although in an important way, in this paper.

Nonresponse error stems from the fact that some of the members of the sample population do not respond to the survey questions. The vast majority of research on improving mail survey methods has focused on response rates, the generally accepted indicator of nonresponse error. This near singular focus on response rates as a means of improving results to mail surveys occurs because low response has long been considered the major problem of mail surveys. Just as important, however, is the perception fostered by much of the literature that such research could be successful in identifying means of improving response rates, in contrast to the limited opportunities for overcoming noncoverage error. Thus, nonresponse error is the major focus of this paper.

Measurement error refers to the discrepancy between underlying, unobserved variables (whether opinions, behaviors, or attributes) and the observed survey responses. Whereas the three preceding types of error (sampling, noncoverage, and nonresponse) stem from nonobservation, measurement error results from the process of observation. In practical terms, measurement error results from respondent characteristics, e.g. their inability to provide accurate information or a motivation for whatever reason to provide inaccurate information. Measurement error may also result from characteristics of the question (e.g. a question phrased so that it cannot be answered correctly) or of the questionnaire, (e.g. the order in which questions are presented).

In recent years, there has been a great deal of interest in mixed mode surveys, i.e. data collected by two or more methods for a single survey purpose (Dillman & Tarnai 1988). This interest, propelled by rising costs and a recognition that different members of a single household may be accessible to one method but not another, forces an explicit recognition of the survey mode itself as a source of measurement error. The fact that mail surveys depend upon visual communication and the ability of people to read and write, that telephone surveys depend on aural communication, and that face-to-face interviews may involve all of the senses raises the specter of different methods producing different results. Thus, it should not come as a surprise that measurement error is a rapidly expanding topic of research for mail as well as other types of surveys.

A good sample survey, by whatever method, is one in which all members of a population have a known opportunity to be sampled for inclusion in the survey (noncoverage error is avoided); the people to be surveyed are sampled by random methods in sufficiently large numbers to provide a desired level of precision (sampling survey error is limited); questions are selected and phrased in ways that result in people providing accurate information (measurement error is avoided); and everyone who is included in the sample responds (nonresponse error is avoided). Seldom if ever does a sample survey

accomplish all of these goals. However, this multifaceted concern becomes a standard by which progress in the development of mail survey methods can be measured.

REDUCING NONRESPONSE ERROR

A low response rate does not necessarily entail nonresponse error, i.e. a discrepancy between the frequency of a population characteristic and that estimated by the survey that occurs because some people did not respond. Those who respond to a survey may not differ in any measurable way from those who do respond. However, the usual reason that surveys are commissioned is that the distribution of certain population characteristics is unknown and a survey becomes the tool to find out the distribution of those characteristics. Thus, it is usually impossible to compare respondents with nonrespondents on precisely those variables of most survey interest. For populations in which a few characteristics of the completed sample are known (e.g. grades and age for a student sample), limited respondent-nonrespondent comparisons can sometimes be made. However, even when one can be confident that no differences exist on these variables, one still does not know whether differences exist on those variables of interest that led to the decision to conduct the survey. It is for this reason that response rate has de facto become the accepted proxy for nonresponse error. The general assumption is that the higher the response rate the lower the potential of nonresponse error and therefore the better the survey.

Low response rate, more than any other issue, has given mail surveys a poor image among social researchers. Prior to the 1970s, mail surveys were generally seen as producing response rates unacceptable for serious sociological research (Dillman 1978). Thus, it is not surprising that the vast majority of mail survey research has, and continues to be, focused on this issue.

The research literature is dominated by hundreds of tests of specific procedures and techniques for improving response rates, typically manipulated one or two at a time. Examples of techniques tested include: financial incentives, material incentives (e.g. ballpoint pens), follow-up reminders, timing of follow-ups, personalization of correspondence, anonymity of response, questionnaire layout, questionnaire length, color of questionnaire, type of outgoing postage, type of return postage, content of cover letter, source of survey sponsorship, and higher rate of postage (for example, see Dillman 1978).

Most such articles are in themselves quite limited in what can be learned about how to improve response rates. Studies differ dramatically with respect to the population that is surveyed, topic, length of questionnaire, and especially procedures other than the test procedures used to achieve a higher

response rate. Thus, when one learns that prior notice is found to increase response rate in a survey in which no follow-up is used, it is impossible to deduce what would happen if several follow-ups (now, a generally accepted procedure) were to be employed. Also, if one finds that the color of stationery increases responses from 25% to 30%, still quite a low final response rate, one is left wondering whether such a difference would also be observed if additional follow-ups, personalized correspondence, or other techniques were used. Would these results be replicated with at least a 50% response rate in the control group? To put it another way, would a technique that was found to make a difference in conditions that led to a low response rate work less effectively under conditions that led to a higher response rate?

The last two decades have witnessed three distinctively different efforts to move beyond the limitations of these individual studies. Each effort attempts to assess results from a large number of previous efforts and to ascertain how response rates can be improved. We label these efforts here as (*a*) subjective assessments, (*b*) quantitative assessments, and (*c*) design of comprehensive systems.

Subjective Assessments

In 1961, Christopher Scott reported a lengthy and detailed comprehensive review of existing research, intended to reach conclusions about which techniques could be reasonably expected to improve response rates and which could not. His major conclusion was that follow-up mailings were most effective, a finding reaffirmed by every comprehensive literature review since that time. Other factors judged by Scott to make a positive difference included prior notice, financial incentives, special postage, sponsorship, stamps on the return envelope, personalization, and interest in the survey topic (see Table 1). More than a decade later, Linsky (1975), in another exhaustive review of the literature, reaffirmed the importance of all but one of these considerations ("interest in the topic" was not included in the review). Similar conclusions were reached by Kanuck & Berensen (1975), in a review that focused in large part on research conducted in the marketing research tradition, and Duncan (1979) in a review that emphasized management research. Techniques or study characteristics that did not make a difference, according one or more of these reviews, included nature of the questionnaire, deadline date, promise of anonymity, nature of the cover letter, and questionnaire length.

Another review, limited to the use of financial incentives, sought to specify the nature of incentives that worked best (Armstrong 1975). Only prepaid (as opposed to postpaid) monetary incentives appeared consistently to make a difference, and in general, nonresponse was reduced by about one third through their use. Perhaps the most striking finding of these studies is that on virtually every topic (with the possible exceptions of follow-ups and financial

incentives), contradictory findings existed that could not be easily dismissed as the result of differences in study designs.

Although these studies each made an important contribution to assessing systematically the relative importance of various factors influencing response rate, their conclusions were necessarily limited by the author's subjective assessment of effects. In addition, each represented only a compilation of findings from previous literature; there was no effort to integrate them theoretically to determine which elements should be combined in what way to obtain the highest possible response rate. The limitations of this approach are explicitly stated in a quite recent review by Harvey (1987) done in this tradition. He reaffirms the importance of precontact, stamped return envelopes, follow-ups, and economic incentives but concludes that very different factors may be at work in different situations and that these are probably not stable across populations.

Quantitative Assessments

In 1978, Heberlein & Baumgartner published an innovative, quantitative assessment of procedures used in previous mail survey research. Two hundred forty (240) manipulations applied to 98 mail survey experiments were treated as if they were respondents to a survey. Seventy-one characteristics of each of these "respondents," i.e. surveys, were coded; they included four general category characteristics: (*a*) research study attributes (e.g. type of sponsoring research organization); (*b*) sampling and sample characteristics (e.g. type of population such as students or government employees); (*c*) questionnaire characteristics (e.g. salience of topic); and (*d*) research procedures (e.g. a special third contact). These researchers then used a regression analysis procedure to predict final response rates, using the 71 characteristics as independent variables.

Heberlein & Baumgartner found that number of contacts and salience of the topic explained 51% of the variation in the final response rate. A 10-variable model, including these and other variables, predicted 66% of the variation in the final response rate. The predictive variables included: market research, sponsorship, general population, and length (all negative effects); and government sponsorship, employee population, school or army population, special third contact, and incentive on the first contact (all positive effects).

Goyder (1982) replicated this study on an extended sample of 330 surveys with similar results, except that the negative effect of market research organization sponsorship disappeared. The negative effect of general population and the positive effect of employee population were also significantly less. Another replication by Eichner & Habbermehl (1981), utilizing studies from Austria and West Germany, suggested potential cross-cultural differences.

A major contribution of these quantitative assessments is to focus attention

squarely on structural constraints of doing surveys (e.g. sponsorship and population to be surveyed). Prior to this time research had mostly focused on identifying manipulable procedures the might be shaped to improve response. The identification of salience of topic as one of the most important determinants of response, along with the effects of structural constraints, made it clear that the potential for reducing nonresponse error varies considerably across types of surveys. One legacy of this model is a quantitative means for researchers to learn how response to a survey might be improved by making particular changes (e.g. adding an incentive or cutting pages from the questionnaire), an idea eventually expanded further by Bruvold & Comer (1988). Heberlein & Baumgartner also added a new slant to response rate research by attempting to predict response to the initial mailing and then using that result to predict overall response rate; this idea usefully separated sources of early and late effects in the survey process.

These quantitative approaches also exhibit certain unavoidable limitations: missing data (because authors of the original individual studies did not report details on nonexperimental factors), and subjective decisions on how variables such as salience were to be coded. The quantitative assessment technique is also limited in its ability to provide information about techniques not tried or used in only a few past studies. Further, the analysis is empirically driven, as is most of the past literature, so that the process of combining techniques becomes mechanical, devoid of guidance from conceptions of why people behave as respondents or nonrespondents. Finally, as generally noted by the authors, regression analysis is not a substitute for a very large factorial design in which many variables would simultaneously be manipulated.

Fox et al (1988) developed a somewhat different quantitative assessment of past studies in an effort to overcome some of the limitations of the regression analysis method. Their meta-analysis involved accumulating statistical evidence against a null hypothesis across individual studies to determine an overall level of significance at which the hypothesis could be rejected. This study was, in many ways, less ambitious than that of Heberlein & Baumgartner as only nine variables were tested. The six shown to be significant included university sponsorship, prenotification by letter, postcard follow-up, first-class outgoing postage, and questionnaire color (Table 1). In contrast, notification of a cutoff date and a postscript to the letter were not significant. Some of their conclusions (e.g. questionnaire color) may be questionable because of the small number of studies available for review. Armstrong & Luske (1987) provide a similar meta-analysis with respect only to the use of postage and presumably show that applying postage to a return letter (versus business reply envelopes) makes a positive difference.

Design of Comprehensive Systems

Efforts to design comprehensive systems with the explicit goal of maximizing response rates represent a quite different approach to the problem of reducing nonresponse error. Though informed by past research, such an approach does not seek to summarize it either statistically or qualitatively. Rather, an attempt is made to construct a system of procedures and techniques that will maximize response rate, in ways consistent with obtaining quality responses.

An example of this approach is the Total Design Method (TDM) (Dillman 1972, 1978, 1983, Dillman et al 1974). The TDM utilizes social exchange theory to guide the careful integration of specific procedures and techniques, many of which were found important by both the subjective and quantitative assessments just reviewed. The theoretical framework used in this approach posits that questionnaire recipients are most likely to respond if they expect that the perceived benefits of doing so will outweigh the perceived costs of responding. Thus, every visible aspect of the questionnaire development and survey implementation process is subjected to three design considerations: the reduction of perceived costs (e.g. making the questionnaire appear easier and less time-consuming to complete), increasing perceived rewards (e.g. making the questionnaire itself interesting to fill out by adding interest-getting questions), and increasing trust (e.g. by use of official stationery and sponsorship) that the promised rewards will be realized (Dillman 1978, 1983).

Among the consequent TDM design recommendations are the following: ordering questions to assure that interesting ones related to the topic described in the cover letter come first; use of graphical design and numerous question-writing principles to ease the task of reading and answering questions (e.g. consistency in the use of large [capital or dark] letters and contrasting small letters); printing the questionnaire in a booklet format with a topically neutral but interesting cover; use of photo reduction of regular sized type to make pages seem smaller and easier to complete; use of four carefully spaced mailings, including a postcard follow-up one week after the original mailing, a replacement questionnaire and cover letter informing the recipient the questionnaire has not yet been received four weeks after the original mailing, and a second replacement questionnaire and cover letter seven weeks after the first mailing, sent to nonrespondents by certified mail; individually printed, addressed, and signed letters; addresses printed onto envelopes rather than on address labels; use of smaller than usual business stationery to reduce costs and make the entire request appear smaller and easier to comply with; cover letter content that includes descriptions of the study's social usefulness and why the respondent is important (repeated in different ways in each of the subsequent mailings); explanation of identification numbers and how con-

fidentiality is protected; folding of outgoing materials in a way that contrasts with advertising mail. Other details of the TDM, in particular how they are integrated to create a holistic effect, are described elsewhere (Dillman 1978, 1983).

The major strength of the Total Design Method as a comprehensive system is that meticulously following the prescribed procedures consistently produces high response rates for virtually all survey populations. Response rates typically reach 50–70% for general public surveys, and 60–80% for more homogenous groups where low education is not a characteristic of the population (Dillman 1978, 1983, Mullen et al 1987). Response rates greater than 80% have frequently been reported (cf Crosby et al 1989, Dillman 1978). To my knowledge, no study that has utilized a 12-page or smaller booklet and followed the TDM in complete detail, from questionnaire through the full set of implementation procedures, has obtained less than a 50% response rate. The TDM has also produced comparable response rates in Europe (Nederhof 1983a, b, 1985, DeLeeuw & Hox 1988), Japan (Jussaume & Yamada 1990), and Australia (Graetz 1985).

Another strength of the TDM is that it is theoretically driven, removing emphasis from individual techniques and stressing how elements fit together. The emphasis is placed not on a particular technique (e.g. personalization or follow-ups), but on how these procedures can be linked to influence questionnaire recipients positively. This theoretical guidance focuses the design upon projecting a consistent positive image of the respondent. Although the TDM incorporates many features found important in past research, it treats some of them differently. For example, Heberlein & Baumgartner's work (1978) revealed that lack of salience of the questionnaire topic was a constraint to high response. The TDM approach treats salience as a manipulatable variable and incorporates a number of efforts to increase salience of topics (e.g. questionnaire cover, question order, and addition of interest-getting question).

The TDM also exhibits a number of significant shortcomings. First, there is a lack of empirical evidence that all, or even most, of the dozens of prescribed details make a difference. Some details are particularly open to question: the size of stationery, size of the booklet, and method of folding questionnaires. Work by Jansen (1985) has questioned the usefulness of photo reduction and shown that a larger size can produce a higher response rate. Limited evidence exists that the degree of overall adherence to the TDM significantly influences responses; the evidence is provided by a study in which research investigators located in 13 different states each administered the identical questionnaire (which closely followed TDM procedures) to similarly drawn samples of the general public in their respective states. An "Adherence to TDM" index was calculated which scored each state on how many of the 22 implementation

details of the TDM were utilized. State scores on this index ranged from 6.5 to 21.5. Response rates ranged from a low of 38 percent (of the questionnaires sent in the original mailing) for the state with the lowest score, to 65.1 percent in the state with the highest score. A Spearman rank-difference correlation of .81 confirmed that adherence to details of the TDM implementation process correlated closely with the final response rates (Dillman et al 1984).

Evidence also exists that certain specific features of the TDM make a significant difference as part of this comprehensive mail survey system. These features include follow-ups (Dillman et al 1984), the certified third follow-up (House et al 1977, Dillman & Moore 1983, Nederhof 1985a, b, and DeLeeuw & Hox 1988), follow-ups that inform people their questionnaire has not yet been received versus those that provide a deadline date (Nevin & Ford 1976), and personalization (Dillman & Frey 1974).

The effect of graphic cover design has also been tested. Grembowski (1985) found that response rates were higher for a cover that had more salience to the questionnaire recipients. A graphical representation emphasizing fluoridation as a means of protecting children's dental health performed better than a cover emphasizing reduction of dental costs in a survey of parents, all of whom who had dental insurance plans as part of their employee benefits, and for whom cost was probably not a major issue of direct concern. Nederhof (1988) conducted a study in which identical covers were developed, except that one questionnaire cover was made to stand out more than the other by filling open spaces with black. This "high contrast" cover was intended to be more memorable and thus easier for the questionnaire recipient to recall and retrieve it. His hypothesis proved correct in a survey of Dutch biotechnologists, with 75.6% of the recipients returning the black-dominated questionnaire versus 60% who returned the low contrast (or white) version. More importantly, the entire difference was attributable to returns received from the second wave, i.e., the postcard reminder when no replacement questionnaire was enclosed. Nearly identical returns were received from the first and third mailings (a fourth one was not used), suggesting that the effect was due entirely to the "greater retrievability" at the time of the postcard follow-up that encouraged the recipient to find and return the questionnaire.

Other research has demonstrated that nearly comparable response rates can be obtained by substituting a telephone follow-up for the final certified mailing, when a telephone number is available (Dillman & Moore 1983, Dillman et al 1984). This finding is particularly important because of societal trends toward greater work force participation for all adults in a household and toward the existence of more single person households. When a certified letter is delivered and no one is at home to sign for it, then a note is left asking the respondent to come to the post office to pick it up. This procedure allows an

opportunity to identify ineligible respondents who had simply ignored other mailings, thus improving the researcher's ability to calculate response rates accurately.

In addition, four experiments (Dillman 1978, Dillman & Moore 1983) have consistently shown that a modest increase in response, from 2–4%, can be obtained by using stamped return envelopes instead of business reply envelopes. Except for these changes—in the substitution of a final telephone reminder for the certified mailing, the use of stamped return envelopes, and use of modern word processing equipment to print questionnaires and provide better graphical design on them—the TDM remains mostly unchanged. We still know relatively little about the importance of many of the originally prescribed details. It does appear that the advent of desk-top publishing has led many users of TDM procedures, including the author, to switch to slightly larger questionnaire sizes (folded legal size paper), more use of color, and less attention to folding and inserting the questionnaires in the prescribed manner.

A particularly glaring shortcoming of the TDM is that it ignores the use of financial incentives, whose importance in improving response rates according to past research appears second only to the use of follow-ups. Relatively few studies have used a full TDM procedure and experimented with various types of financial incentives. An exception is a recently completed study by Roberts & Hite (1990) which showed that respondents to a TDM household survey who received a $1.00 check were significantly more likely to respond, 72% to 65%. A study by James & Bolstein (1990) has also demonstrated positive effects for both $1.00 and $2.00 incentives on TDM surveys. Finally, a recent study by Johnson & McLaughlin (1990), using a 32-page questionnaire, obtained an 82.6% response rate with a $5.00 cash prepayment versus nearly identical questionnaire but lower response rates for a $10.00 check after completion (72.3%) and for no financial incentive (71.3%). However, the $10.00 post payment resulted in a significantly lower item nonresponse, 5.3% versus 9% for other respondents. The use of material incentives in TDM surveys has been attempted but generally they have not had any discernable effect on response rates (Nederhof 1983b, Mullen et al 1987).

A third limitation of the TDM may be the reliance on social exchange theory. This theory was developed to explain the nature of face-to-face interaction and the social structures that evolved from those interactions (Blau 1964 and Homans 1961). Mail survey respondents and researchers do not, as a general rule, come into contact with one another. Although the communications efforts (as many as four of which originate from the researcher) can reasonably be viewed as attempts at interaction, this application of social exchange theory goes somewhat beyond its traditional development. Other theoretical perspectives might provide the basis for development of comprehensive mail survey systems (e.g. dissonance theory) (Hackler & Bour-

gette 1973), self-perception theory (Hansen 1980), attitude-behavior and altruism theoretical frameworks (Heberlein & Baumgartner 1978), and a broad social-values perspective that places emphasis on respondent's concerns of privacy (Goyder & Leiper 1985). However, none of these theoretical orientations has, as yet, been utilized individually or in concert with others to develop a comprehensive survey system aimed at minimizing nonresponse error.

A fourth limitation of the TDM is that the same set of procedures is recommended for all populations, thus giving it a one-size-fits-all nature. This puts the social exchange approach that underlies the TDM somewhat at odds with the TDM procedures derived from it. It is reasonable to expect that different people value somewhat different returns for their actions to somewhat different degrees, and that these differences would be reflected in the perceived balance among rewards and costs associated with completing a questionnaire. For example, the reduced size of print and questionnaire advocated in the TDM may produce a positive response among young people, whereas enlarged print and size of questionnaire might do the same for older people. Thus, customization of detailed procedures to particular populations might be expected to maximize response for different survey populations.

This possibility is demonstrated by Teske et al (1983), who has used a modified version of the TDM for numerous crime, victimization, and abuse surveys in Texas. Their procedures involved a TDM-like booklet questionnaire and five contacts, including a prenotice, first mailing of a questionnaire, a postcard reminder, second mailing of a questionnaire, and a final reminder postcard. Personalized letters and stamped return envelopes were also used. Eleven statewide general public surveys that utilized these procedures obtained response rates of 61–77%.

Closely related to this one-method-fits-all limitation of the TDM is the lack of recognition of how different survey situations may require quite different survey procedures. For example, one of the biggest uses of mail surveys in the United States is requests to complete television viewing diaries in order to estimate audiences in each of the nation's television markets. Typically these surveys ask large numbers of respondents to fill out their diaries for a specific week; an emphasis is placed on not doing it by recall after that week is over. A standard TDM approach with contacts spread out over seven weeks simply would not work or be useful. A procedure commonly used by television research companies to complete such surveys involves two or more precontacts, two more contacts by mail and/or telephone during the diary week to remind people of the need to complete the questionnaire as they watch television, and another notice to return the questionnaire at the end of the diary week.

Another situation which a standard TDM procedure does not fit is that

when questionnaires can only be hand-delivered to respondents (e.g. airports, shopping centers, and museums where the questionnaire is handed to visitors who are asked to return it at a later time, or even the drop-off of questionnaires to an area of a probability sample of households). Machlis & Dillman (1990) have developed a modified TDM procedure for use in such situations. They combine social exchange and foot-in-the-door principles to develop a sequence of requests that involves first asking respondents to pull to the side of the road, second, asking if they would be willing to answer three short questions and complete a mail-back questionnaire after their visit to the park, and third, asking for a mailing address so that a "thank-you" can be sent to them. The resultant procedure produced response rates of 84–88% from four national parks versus 31–42% in three parks where the original procedure was simply handing out questionnaires to random samples of visitors as they paid the park entrance fee.

Conclusions and Needed Research

In a regression analysis of the response rate determinants of 500 different surveys, Goyder found that on average the response rate for mail surveys was only 7.5% lower than for face-to-face interview surveys included in the sample (1985). Heberlein & Baumgartner, in the previously discussed quantitative assessment of mail survey responses, found that the surveys they examined had obtained response rates of over 60%, and one fourth had response rates higher than 80%. DeLeeuw (1990) in a meta analysis of method comparison studies found that mail survey obtained response rates only eight percentage points lower than face-to-face interviews. Juxtaposition of these findings with those from TDM studies leads me to conclude that there is no longer any reason to accept low response rates to mail surveys. Procedures are available to assure response rates of at least 50% and higher than that for most populations. Lower response rates can no longer be excused on the grounds of the inadequacy of the available mail survey procedures.

Further, the consistency among major findings from past assessment efforts, as shown in Table 1, suggests substantial agreement on which techniques influence response. Improvements in our knowledge of how to reduce response error in surveys will not be advanced much further by the type of article that has dominated and continues to dominate mail research, i.e. experiments on one or two techniques applied to base procedures that obtain low response rates. It can be assumed that there is a theoretical maximum response asymptote, which we cannot hope to exceed except through the use of culturally unacceptable inducements. Although for some populations this response asymptote may be 100%, for others it may be only 80–85%. Application of current knowledge can be expected to take response to within 10–20 percentage points of this theoretical maximum. The kind of technique-

Table 1 Techniques found important in three approaches to identifying influences on mail survey response rates

Dependent variable	Subjective assessment					Quantitative assessment				Comprehensive system
	1961 Scott	1975 Kanuk Berenson	1975 Linsky	1979 Duncan	1987 Harvey	1979 Heberlein Baumgart.	1981 Eichner Habermehl	1982 Goyder	1988 Fox et al	1978 Dillman
Follow-up	***	***	***	***	***	***	***	***	**	***
Prior notice	*	**	***	***	**	**	—	—	***	—
Financial incentive	**	***	***	***	***	**	.	**	***	—
Special postage	**	**	***	?	**	**	.	**	**	***
Sponsor	**	*	***	**	?	**	**(−)	**	—	—
Stamped return envelope	**	**	**	***	***	.	—	—	***	***
Personalization	*	*	*	***	**	.	—	—	—	***
Interest (salience)	**	—	.	—	**	***	***	***	—	***
Questionnaire length	.	.	.	.	*	**	**	**	—	*
Cover letter	**	?	.	**	**	—	—	—	—	***
Anonymity	.	.	.	.	.	—	—	—	—	—
Deadline date	—	.	—	.	—	—	—	—	.	—
Size, color, +/or reproduction	.	.	.	.	**	—	—	—	*	***
Population surveyed	—	—	—	—	*	**	**	**	.	.

Relative importance ascribed by the researcher's conclusions as judged by the author: *** very important; ** somewhat important; * slightly important; · not important; ? no conclusion reached; — no effect; (−) direction opposite to predicted direction

oriented response rate experiments needed in the future are ones that use a comprehensive system as a base procedure and then are designed so that small incremental effects, e.g. two or three percentage points, can be identified as the asymptote is approached.

A second research need is to shift from a focus on techniques like those described in Table 1 to an emphasis on how parts of the survey process complement one another. There is a need for an entirely new set of concepts to capture these survey dynamics. For example, observation of people who make repeated requests of another person to do something, with which that person has not yet complied, suggests that the requester tends to pose the question in different ways each time the request is made, presumably in search of an appeal that works. This observation suggests that "variety of contact" might be tested as a way of improving response rates. A survey procedure could be designed which uses different although compatible sizes and/or color of stationery, quite different letter content, sent by different types of mail in each contact. The questionnaire might even be redesigned to make it shorter for a final mailing. Another concept worthy of testing is the "escalation and de-escalation of intensity" associated with a sequence of contacts. One might use a first contact that is very "soft", for example, mentioning that a survey will be coming and that the sender would appreciate the respondent taking a look at it, and then increasing the intensity of the request culminating in a special delivery or express mailing that includes a substantial financial incentive in it. Alternatively, one might initiate a survey process with a handwritten, very personalized mailing that includes a financial incentive and provides a deadline date, sent by express mail, but then followed up with softer, more voluntary types of messages. Thus, one might test the hypothesis of whether a declining or escalating intensity or an even balance among all contacts is most effective. The importance of the "variety" and "escalation" hypotheses here is less that these particular concepts themselves may bear fruit than to illustrate how we might move beyond a focus on techniques to testing processes of an entire contact procedure for reducing nonresponse error.

Finally, as we approach the asymptote of the highest reasonably obtainable response, there is a great need to test for reactive effects that result in less accurate data due to pressing too intensely, for higher response rates. The dilemma has been demonstrated in Sweden by Christoffersen (1987), who found an educational bias in mail questionnaires, even when very high response rates were obtained; based upon other work, however, he concludes that intensive persuasion greatly compounds the error of responses. A contrasting perspective is provided by DeLeeuw & Hox, who found in a TDM experiment that the use of personalization without a certified letter resulted in more socially desirable answers than when the certified final follow-up was

used (1988). This issue underscores the very close linkage between nonresponse error and measurement error, the issue to which we now turn.

REDUCING MEASUREMENT ERROR

Traditionally, concerns over measurement error in mail surveys have focused on item nonresponse, failure to obtain adequate answers to open-ended questions, the effect of low educational levels, the lack of success with complex skip patterns, and the ability of the researcher to control the sequence in which questions were asked. These problem areas tended to be balanced against two apparent advantages, i.e. that the absence of an interviewer lessens the likelihood both of socially desirable responses and of interviewer subversion (Dillman 1978).

During the 1980s, however, renewed attention has been focused on measurement error in mail surveys for two reasons. First, new research had persuasively demonstrated measurement error concerns in telephone and face-to-face interviews that are somewhat worse than many researchers had expected (Schuman & Presser 1981). Second, increasing numbers of mixed mode surveys apparently are being done to reduce costs and to deal with increasing error problems in face-to-face and telephone surveys, e.g. falling response rates (Dillman & Tarnai 1988, Goodfellow et al 1988, Brambilla & McKinlay 1987). Thus, the important questions about measurement error are not only how accurate are data obtained by mail surveys, but whether the answers are the same as those obtained for telephone and face-to-face interviews. The latter issue is especially important because many questions have no absolute standard of accuracy, e.g. attitude questions that utilize vague quantifier answer choices such as "strongly agree" or "somewhat agree." Mail survey research in these issues has been expanding but many questions remain unanswered.

Do Mode Differences Exist?

It is reasonable to expect that certain differences may occur when questions are answered in mail versus interview surveys. Mail survey respondents depend upon a visual presentation of information for which reading skills are necessary. In contrast, telephone respondents depend entirely upon aural skills for hearing the questions and upon memory and verbal skills for responding. All of these skills can be utilized in face-to-face interviews, but in general very little reading is likely to be required. The physical presence of the interviewer in the face-to-face situation, and the necessity of interacting with an interviewer for telephone interviews, also provide a basis for expecting that respondents might provide different responses.

Thus, it is encouraging that relatively few differences have been reported

for personal attributes of respondents or their households that could not be attributed to nonmeasurement error differences such as differential response rates or noncoverage issues. The data also appear mostly equivalent for easy-to-recall behavioral characteristics.

In one of the first major self-administered (mail retrieval was not used) versus telephone comparisons, Bishop et al (1988) found that the order effects previously identified by Schuman & Presser (1981) were less likely to occur in mail surveys. Ayidiya & McClendon (1990) also tested several effects reported in the Schuman & Presser (1981) study. Their work generally supports that of Bishop et al (1988) in suggesting that response effects in mail surveys are somewhat less than in telephone surveys. They found that for the most part recency effects (choosing the last response category) were reduced, but that primacy effects (choosing the first offered category) persisted. They found no evidence that question form effects produced by nonopinion filters, such as offering a middle alternative on agree or disagree questions, would be eliminated in a mail survey. A limitation of the Ayidiya & McClendon study is that they report results only for a mail survey and, therefore, can only compare their mail survey results with the original interview responses from the Schuman & Presser study.

Another type of difference frequently observed in mode research comparisons is a tendency for telephone and face-to-face respondents to select more extreme answers than do mail respondents when vaguely quantified scaler categories are used. Mail respondents, on the other hand, tend to distribute themselves across the full scale. Such an effect was first reported by Hochstim (1967), who found that the percentage of respondents saying their health was excellent (compared to good, fair, or poor) was 30% by mail versus 44% by telephone and 47% in the face-to-face interview. Mail respondents were also more likely to report having discussed medical problems with their husbands, an issue which, Hochstim reasoned, respondents might have been embarrassed to admit to the female interviewer.

In a survey of drinking behavior, Mangione et al (1982) reported no difference between mail and face-to-face or telephone surveys on 12 demographic items, but found self-administered respondents more likely to use middle categories as opposed to extreme ones on questions that involved elements of socially desirability. Similarly, Dillman & Tarnai (in press) found that telephone respondents were more likely than mail respondents to say that they "never" (as opposed to frequently, sometimes, or occasionally) drove a car after they had been drinking alcoholic beverages.

A dental health study by Zapka et al (1988) showed that telephone respondents reported brushing more frequently than did mail respondents. In a survey of former hospital patients, Walker & Restuccia (1984) found that telephone respondents reported greater satisfaction with their recent stay. A needs assessment of university psychological services revealed that 49 of 50

items were rated higher by telephone respondents (Talley et al 1983). In a comparison of mail, telephone, and face-to-face procedures, on a wide range of scales, DeLeeuw (1990) found small differences in psychometric reliability, with the mail method being most reliable and the telephone being least reliable.

In a three-way comparison of methods, Dillman & Mason reported no significant differences on demographic items (1984). However, in a series of nine items that required respondents to assess specific community issues as "not a problem," "a small problem," "a medium problem," or "a serious problem" in their community, telephone respondents and face-to-face respondents were more likely to choose the extreme response choice (not a problem) for all nine items. A similar tendency was observed on a second set of nine items-in-a-series measuring neighborhood quality on a seven-point numerical scale, and on five individual items using various scaler responses (Dillman & Mason 1984). In a later study, the nine community problem items and neighborhood quality items were adapted to student living situations and readministered to samples of students by telephone, as well as by in-class self-administered questionnaires, with similar results. Whereas a field survey had resulted in a difference of 6.7 to 19.3 percentage points in use of the "not a problem" category between the mail and telephone, the student survey produced even greater differences of 12.6 to 38.1 percentage points (Tarnai & Dillman in press).

In sum, there is substantial evidence that some differences exist between responses to certain questions asked in mail and telephone interview surveys. However, most of these studies have been empirically focused and have made only very limited attempts to provide theoretical explanations for the differences.

Potential Explanations for Mode Differences

At least three possibilities have been posed as explanations for the telephone extremeness effect noted above. Perhaps the most frequently offered explanation is that respondents to self-administered questionnaires do not have to interact with an interviewer and are therefore less inclined to offer socially desirable answers. This explanation seems likely to account, at least in part, for the differences reported by Hochstim (1967), Mangione et al (1982), Tarnai & Dillman (in press), Zapka et al (1988) and Walker & Restuccia (1984). Each of these surveys dealt with items for which a clear direction of social desirability could be identified. However, the nature of these studies was such that other effects could not be ruled out.

A second possible explanation is based upon the fact that telephone respondents must hear and respond to one question at a time, and mail respondents are able to look at full pages of questions. Thus, mail respondents are able to

consider answers to questions that both precede and follow particular items when deciding upon an answer. This "context effect" has been hypothesized to account for the extremeness effect in the series of community and neighborhood needs assessment items. In an experiment that varied the ability of respondents to see the context of each question as it was being asked of them over the telephone, Dillman & Tarnai have shown that context accounted for some, but by no means all, of the extremeness effect (in press).

A third explanation concerns differences in pace and control of the answering process in telephone versus self-administered surveys (Hippler & Schwarz 1987), (Schwartz et al (in press). Telephone interviewers read each question at a pace set by them, and there is considerable pressure on respondents to listen intently and respond when asked. This pressure may lead to a top-of-the-head tendency, whereby the respondents base their judgment on the first piece of information that comes to mind. This perspective may explain why self-administered questionnaires were found by Bishop et al (1988) and Ayidiya & McClendon (1990) to be somewhat less subject to order effects. It may also explain the extremeness effect under the assumption that the end categories are more likely to be remembered than are the intermediate categories.

Much remains to be learned about responses to self-administered and interview surveys, but the above research suggests that progress is being made. A theory of response-order effects appears to be emerging from research done in interview surveys. It is based on a theoretical perspective reported by Krosnick & Alwin (1987). They argued that when long lists of items were presented to respondents in visual form, the respondents were more likely to choose responses that appeared early in the list; in oral form later items were more likely to be chosen. Israel and Taylor (in press) demonstrate this tendency for choosing the first item in a mail experiment, but no telephone comparison was made. Hippler, Schwarz, Noelle-Neumann, and their colleauges have extended this work in a series of papers, showing, for example, that the theory applies as well to dichotomous questions (Hippler et al 1989). They also show that the need to compute a judgment on the spot is a crucial prerequisite (Schwarz et al 1989). However, this theory cannot account for the extremeness effect described by Tarnai & Dillman (in press), which resulted in telephone respondents frequently picking the first offered category more often than did mail respondents. These differences may stem from the fact that Krosnick & Alwin (1987) examined the choosing of answers from a list of nominal alternatives whereas Tarnai & Dillman (in press) used ordinal response choices. This response task required the respondent to determine where his or her answer fit on an implied scale.

Finally, in face-to-face surveys the ability of respondents accurately to recall visits to doctors can be improved by the use of recall techniques such as

reconstructing details of the visit and the use of benchmark dates important in people's lives (Means et al 1989 and Jobe & Mingay 1989). Recently, these techniques were applied to mail survey respondents by asking about the extent to which they wore seatbelts on their three most recent trips in a car. The use of a four-question sequence to reconstruct details of each trip resulted in increases of 13–16% of the mail and 11–14% of the telephone respondents who reported wearing their seatbelts all of the time (Dillman & Tarnai, forthcoming). In the past, one of the perceived advantages of mail surveys was that people had time to carefully think through questions before answering them. These data raise the possibility that this image of mail questionnaires is more a myth than reality, and that cognitively designed series of questions may be just as important for use in mail surveys as in interview surveys.

Future Research Needs

The last few years have witnessed a very promising start in raising critical questions about measurement error in mail surveys. Potential explanations of some of these effects have been offered but much remains to be done in order to develop a theory of response effects. It is critical that this work proceed and that it be done in concert with work on telephone and face-to-face interview surveys if the benefits of this research are to be realized for the development of mixed mode survey applications.

REDUCING NONCOVERAGE ERROR

When complete, up-to-date lists of populations are available, as is frequently the case because of the increased prevalence of computers for compiling and storing lists, the noncoverage error is not usually a problem. Studies of students, professional association members, employers, health insurance recipients, and many other groups are frequent examples. Yet, the problem of noncoverage error is perhaps the single greatest barrier to the expanded use of mail surveys, primarily because of the problem it presents for doing household surveys of the general public. There are no up-to-date listings that provide complete coverage of households in the United States to the degree that can be accomplished by area probability samplings for face-to-face interviews and random digit dialing for telephone surveys. Telephone directories present the most readily available source of names and addresses, but are usually out-of-date. In some regions of the country half or more of the households are unlisted, and this trend continues upward (Survey Sampling 1990). Other lists, such as city directories and drivers license lists, are frequently unavailable and suffer from other problems of noncoverage.

The solution to the noncoverage problem involves learning how to develop

efficient sampling and other procedures for dropping off and retrieving questionnaires from households and perhaps developing response inducement procedures that can rely on household listings without names. As interest in conducting mixed mode surveys develops, it seems likely that such procedures may be developed because of the tremendous cost advantage to delivering questionnaires to people's households versus having interviewers return many times, in order to obtain acceptable response rates. This issue represents a very high priority for future research.

THE FUTURE OF MAIL SURVEYS

We live in an information age. Information is increasingly important to decisionmaking in all units of US society and is generally being substituted for other resources in the production of goods and services (Dillman 1989). The ability of sample surveys to find out information about populations by surveying only limited samples suggests that the use of all types of surveys, and especially mail, will increase. For example, low cost desktop publishing techniques create the capability for conducting quality mail surveys in even small businesses, agencies, and institutions. Another factor suggesting the increased use of mail surveys is the increased cost and nonresponse problems associated with telephone and face-to-face interviews, which have moved us toward mixed mode surveys.

During the last three decades, tremendous progress has been made in improving response rates to mail surveys. Measurement error issues have been identified, and much research is now being directed toward the resolution of these issues. In contrast, the lack of research on noncoverage issues represents a major barrier to advancement in the uses of mail surveys, especially for the difficult-to-survey general public.

In the past, research on improving mail surveys has mostly been conducted outside the major research centers of the United States, where attention has been focused in large part on face-to-face and telephone methods. The advent of interest in mixed mode surveys, in an effort to find the most cost-effective mix of survey procedures for conducting important societal surveys, suggests that the attention of these research centers may increasingly be focused on mail surveys in this context. Should such interest develop, the benefits for the advancement of mail survey research will be considerable.

ACKNOWLEDGMENTS

The author wishes to acknowledge, with thanks, the research assistance of Ms. Roberta L. Sangster for conducting the extensive literature search that was critical to the completion of this chapter. Review comments provided by

Ms. Sangster, Edith DeLeeuw, Joop Hox, and Priscilla Salant are also gratefully acknowledged.

Appreciation is expressed to the Center for Survey Methods Research (ZUMA) in Mannheim, Germany. The review framework and initial draft of some of the issues addressed in this paper were developed there while the author served as Guest Professor in 1985 and 1987. These ideas benefitted from many discussions with Hans-Jurgen Hippler of the Center's staff.

Literature Cited

Armstrong, J. S. 1975. Monetary incentives in mail surveys. *Publ. Opin. Q.* 39:111–16

Armstrong, J. S., Luske, E. J. 1987. Return postage in mail surveys: a meta-analysis. *Public Opin. Q.* 51:(1)233–48

Ayidiya, S. A., McClendon, M. J. 1990. Response effects in mail surveys. *Public Opin. Q.* 54:229–47

Bishop, G. G., Hippler, H., Schwarz, F. 1988. A comparison of response effects in self-administered and telephone surveys. In *Telephone Survey Methodology,* ed. R. M. Groves et al, New York: Wiley & Sons

Blau, P. M. 1964. *Exchange and Power in Social Life*. New York: Wiley & Sons

Brambilla, D. J., McKinlay, S. M. 1987. A comparison of responses to mailed questionnaires and telephone interviews in a mixed-mode health survey. *Am. J. Epidemiol.* 126:962–71

Bruvold, N. T., Comer, J. M. 1988. A model for estimating the response rates to a mailed survey. *J. Bus. Res.* 16:101–16

Christoffersen, M. N. 1987. The educational bias of mail questionnaires. *J. Off. Statist.* 3:459–64

Crosby, F. E., Ventura, M. R., Feldman, M. J. 1989. Examination of a survey methodology: Dillman Total Design Method. *Nursing Res.* 38:56–58

DeLeeuw, E. D. 1990. *Data quality in mail, telephone, and face-to-face surveys: a model comparison in the Netherlands*. Free University. Unpublished ms.

DeLeeuw, E. D., Hox, J. J. 1988. The effects of response-stimulating factors on response rates and data quality in mail surveys: A test of Dillman's Total Design Method. *J. Off. Statist.* 4:241–49

Dillman, D. A., 1972. Increasing mail questionnaire response for large samples of the general public. *Public Opin. Q.* 35(3):254–57

Dillman, D. A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley & Sons

Dillman, D. A. 1983. Mail and self-administered surveys. In *Handbook of Survey Research,* ed. Peter H. Rossi, James D. Wright, Andy B. Anderson. 10:359–77

Dillman, D. A. 1989. Our changing sample survey techniques. *Choices* 4(4):40–41

Dillman, D. A., Carpenter, E., Christenson, J., Brooks, R. 1974. Increasing mail questionnaire response: A four state comparison. *Am. Sociol. Rev.* 39:744–56

Dillman, D. A., Dillman, J. J., Makela, C. J. 1984. The importance of adhering to detail of the Total Design Method (TDM) for mail surveys. In *New Directions for Program Evaluation,* ed. D. C. Lockhart, pp. 49–64. San Francisco: Jossey-Bass.

Dillman, D. A., Frey, J. H. 1974. Contribution of personalization to mail questionnaire response as an element of a previously tested method. *J. Appl. Psychol.* 59:297–301

Dillman, D. L., Mason, R. G. 1984. *The influence of survey methods on question response*. Pap. pres. Annu. meet. Am. Assoc. Public Opin. Res., Delavan, Wis.

Dillman, D. L., Moore, D. 1983. *Improving response to mail surveys: Results from five experiments*. Pres. Annu. Meet. Am. Assoc. Publ. Opin. Res., Buck Hill Falls, Penn.

Dillman, D. A., Sangster, R. L. 1990. *Mail Surveys: A comprehensive bibliography, 1974–1989. Technical Report*. The Soc. & Econ. Sci. Res. Ctr., Wash. State Univ.

Dillman, D. A., Tarnai, J. 1988. Administrative issues in mixed mode surveys. In *Telephone Survey Methodology*, ed. Robert Groves et al, pp. 509–28. New York: Wiley & Sons

Dillman, D. A., Tarnai, J. 1991. Mode effects of a cognitively designed recall question: A comparison of answers to telephone and mail surveys. In *Measurement Errors in Surveys*, ed. Paul, Biemer, et al. New York: Wiley & Sons. In press

Duncan, W. J. 1979. Mail questionnaires in survey research: a review of response inducement techniques. *J. Manage,* 5:39–55

Eichner, K., Habermehl, W. 1981. Predicting

response rates to mailed questionnaires. *Am. Sociol. Rev.* 46:361–63
Fox, R. J., Crask, M. R., Kim, J. 1988. Mail survey response rate: a meta-analysis of selected techniques for inducing response. *Public Opin. Q.* 52(4):467–91
Goodfellow, M., Kiernan, N. E., Ahern, F., Smyer, M. A. 1988. Response bias using 2-stage data-collection—a study of elderly participants in a program. *Eval. Rev.* 12: 638–54
Goyder, J. 1982. Further evidence on factors affecting response rates to mailed questionnaires. *Am. Sociol. Rev.* 47:550–53
Goyder, J. C. 1985. Face-to-face interviews and mail questionnaires: the net difference in response rate. *Public Opin. Q.* 49:234–52
Goyder, J., Leiper, J. M. 1985. The decline in survey response -a social values interpretation. *Sociology—J. Br. Sociol. Assoc.* 19: 55–71
Graetz, B. 1985. The potential of mail surveys. *Austr. New Z. J. Sociol.* 21:445–55
Grembowski, D. 1985. Survey questionnaire salience. (letter). *Am. J. Publ. Health* 75: 1350
Groves, R. M. 1989. Research on survey data quality. *Public Opin. Q.* 51:S156:S172
Hackler, J. C., Bourgette, P. 1973. Dollars, dissonance and survey returns. *Public Opin. Q.* 37(Summer):276–81
Hansen, R. A. 1980. A self-perception interpretation of the effect of monetary and nonmonetary incentives on mail survey respondent behavior. *J. Marketing Res.* 17: 77–83
Harvey, L. 1987. Factors affecting response rates to mailed questionnaires: A comprehensive literature review. *J. Market Res. Society* 29(3):342–53
Heberlein, T. A., Baumgartner, R. 1978. Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature. *Am. Sociol. Rev.* 43: 447–62
Hippler, H. J., Schwarz, N. 1987. Chapter 6: Response effects in surveys. In *Social Information Processing and Survey Methodology*, ed. H. J. Hippler, N. Schwarz, S. Sudman. New York: Springer
Hippler, H. J., Schwarz, N., Noelle-Neumann, E. 1989. *Response order effects and dichotomous questions: the impact of administration mode*. Pap. pres. Annu. meet. Am. Assoc. Public Opin. Res., St. Petersburg, Fla.
Hochstim, J. R. 1967. A critical comparison of three strategies of collecting data from households. *J. Am. Statist. Assoc.* 62:976–89
Homans, G. C. 1961. *Social Behavior: Its Elementary Forms*. New York: Harcourt Brace & World
House, J. S., Gerber, W., Mc Michael, A. J. 1977. Increasing mail questionnaire response: a controlled replication and extension. *Publ. Opin. Q.* 41:95–99
Israel, G. D., Taylor, C. L. 1991. Can response order bias evaluations? *Eval. Prog. Plan.* In press
James, J. M. 1990. Effect of monetary incentives and follow-up on the response rate and response quality in mail surveys. *Public Opin. Q.* 54:346–61
Jansen, J. H. 1985. Effects of questionnaire layout and size and issue: involvement on response rates in mail surveys. *Perceptual & Motor Skills* 61:139–42
Jobe, J. B., Mingay, D. J. 1989. Cognitive research improves questionnaires. *Am. J. Publ. Health* 79(8):1053–55
Johnson, T., McLaughlin, S. 1990. *GMAT Registrant Survey Design Report*. Los Angeles Graduate Admission Council.
Jussaume, R. A., Yamada, Y. 1990. A comparison of the viability of mail surveys in Japan and the US. *Publ. Opin. Q.* 54:219–28
Kanuk, L., Berenson, C. 1975. Mail surveys and response rates: A literature review. *J. Marketing Res.* 12:440–53
Krosnick, J., Alwin, D. F. 1987. An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opin. Q.* 51:201–19
Linsky, A. S. 1975. Stimulating responses to mailed questionnaires: A review. *Public Opin. Q.* 39:82–101
Machlis, G., Dillman, D. L. 1990. Utilization of foot-in-the-door and social exchange principles to improve response to self-administered questionnaires. Pres. Ann. Meet. Am. Assoc. Publ. Opin. Res. Lancaster, Penn.
Mangione, T. W., Hingson, R., Barrett, J. 1982. Collecting sensitive data: a comparison to three survey strategies. *Sociol. Meth. Res.* 10:337–46
Means, B., Nigam, A., Zarrow, M., Loftus, E. F., Donaldson, M. S. 1989. Autobiographical memory for health related events: enhance memory for recurring incidence. Nat. Ctr. Health Statist. *Vital Health Statis.* 6(2):1989
Mullen, P. D., Easling, I., Nixon, S. A., Koester, D. R., Biddle, A. K. 1987. The cost-effectiveness of randomized incentive and follow-up contacts in a national mail survey of family physicians. *Eval. Health Profess.* 10:232–45
Nederhof, A. J. 1983a. Effects of repetition and consistency of personalization treat-

ments on response rate in mail surveys. *Soc. Sci. Res.* 12:1–9
Nederhof, A. J. 1983b. The effects of material incentives in mail surveys: two studies. *Public Opin. Q.* 47:103–11
Nederhof, A. J. 1985a. A survey on suicide—using a mail survey to study a highly threatening topic. *Qual. Quant.* 19:293–302
Nederhof, A. J. 1985b. A comparison of European and North American response patterns in mail survey. *J. Marketing Res. Soc.* 27:55–63
Nederhof, A. J. 1988. Effects of a final telephone reminder and questionnaire cover design in mail surveys. *Soc. Sci. Res.* 17(4):353–61
Nevin, J. R. Ford, N. M. 1976. Effects of a deadline and veiled threat on mail survey responses. *J. Appl. Psychol.* 61:116–18
Roberts, M. L., Hite, P. 1990. Unpublished report of a nation-wide survey of 900 households. Univ. Ala.
Schuman, H., Presser, S. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context.* New York: Academic
Schwarz, N., Hippler, H. L., Noelle-Neumann, E., Munkel, T. 1989. *Response order effects and long lists: primacy regency in asymmetric contrast effects.* Pap. pres. First Nags Head Conf. Cognition & Survey Res., Nags Head, NC
Scott, C. 1961. Research on mail surveys. *J. Royal Statist. Soc.* Series A, 124:143:205
Schwarz, N., Strack, F., Hippler, H–J., Bishop, G. 1991. The impact of administration mode on response effects in survey measurement. *Appl. Cognitive Psychol.* In press
Survey Sampling Inc. 1990. A survey researchers view of the U.S.
Talley, J. E., Barrow, J. C., Fulkerson, K. F., Moore, C. A. 1983. Conducting a needs assessment of University Psychological Services: a campaign of telephone and mail strategies. *J. Am. Coll. Health.* 32:101–3
Tarnai, J., Dillman, D. L. 1991. Questionnaire context as a source in departure differences in mail vs telephone surveys. In *Order Effects in Social and Psychological Research,* ed. N. Schwarz, S. Sudman. In press
Teske, H. C., Hazlett, M. A., Parker, M. L. 1983. *Technical Report.* Texas Crime Poll, 1982 Survey. Unpublished
US Office of Management and Budget. 1984. The role of telephone data collection in federal statistics. *Statis. Policy Work. Pap. No. 1.* USGPO
Walaker, A. H., Restuccia, J. D. 1984. Obtaining information on patient satisfaction with hospital care: mail versus telephone. *Health Serv. Res.* 19:291–306
Zapka, J. G., Stoddard, A. M., Lubin, H. 1988. A comparison of data from dental charts, client interview, and client mail survey. *Med. Care* 26:27–33