KM 31(1989)
pag 123-146

Data Quality in Mail Surveys as Compared to Face to Face and Telephone Interviews: a Meta-Analysis of the Research Literature

E.D. De Leeuw*, J.J. Hox**, and Johannes Van der Zouwen*

Key words: mail survey, telephone survey, face to face survey, interview, mode effect, data quality, meta analysis, quantitative review.

Abstract

In order to investigate the data quality in mail surveys a meta analysis of existing research literature was conducted. In total, 28 articles were reviewed in which mail and face to face surveys were compared, and 19 articles in which mail and telephone surveys were compared. It was found that both face to face and telephone surveys give higher response rates and less item non response than mail surveys. However, mail surveys result in more accurate answers. And, in general, mail surveys perform better when sensitive or embarrassing questions are being asked.

---

* Department of Research Methods, Free University, Koningslaan 22-24, 1075 AD Amsterdam, The Netherlands. (0)20-6647131, EARN: V70UEDIT at HASARA11.
** Department of Education, University of Amsterdam, Prinsengracht 227, 1015 DT Amsterdam, The Netherlands. (0)20-5253315, EARN: A716HOX at HASARA11.
This study was (partly) made possible by a grant from the Social Cultural Sciences Foundation, which is subsidized by the Netherlands Organization for the Advancement of Scientific Research (NWO)

Introduction

The potential high response rate of mail surveys, together with their low costs and geographical flexibility (Houston & Ford, 1976; Kanuk & Berenson, 1975), make mail surveys an attractive alternative for telephone and face-to-face surveys (cf. Thornberry, Nicholls & Kulpinsky, 1982).

That high response rates for mail surveys are possible, is the result of much effort of researchers over the last three decades. The large number of empirical studies on response increasing techniques for mail surveys has been summarized in several comprehensive review articles and quantitative reviews (Baumgartner & Heberlein, 1984; Goyder, 1982; Heberlein & Baumgartner, 1978; Kanuk & Berenson, 1975; Linsky, 1975; Yu & Cooper, 1983).

The utility of a data collection technique does not **only** depend on costs and response rate, but also on the quality of the data collected (Herman, 1977, see also Jones & Lang, 1982, and De Leeuw & Hox, 1988). Unfortunately the issue of data quality, or response error has received little attention in methodological studies on mail surveys (Houston & Ford, 1976). In empirical studies on mode effects however, mail surveys are often included as one of the methods compared. This makes it possible to investigate differences between the data resulting from mail surveys and the data resulting from face to face and telephone surveys on several indicators of data quality.

To integrate research findings on mode differences and to provide a comprehensive review on this subject, we used principles of meta-analysis (Glass, McGaw, and Smith, 1981). This method makes it possible to present both an overview of mode differences found with respect to data quality, and an estimate of the (effect) size of those differences. After a short introduction on meta-analysis, we will give a systematic overview of empirical findings on differences between mail surveys and face to face interviews, and between mail surveys and telephone interviews. We will end with a short summary and a discussion of the findings.

Method

On Meta-Analysis

Though the name "meta-analysis" deceptively suggests otherwise, meta-analysis is not one method or one type of analysis. Meta-analysis or integrative analysis, as it is often called, is a coherent set of quantitative methods for reviewing research literature (Glass, McGaw and Smith, 1981; Light and Pillemer, 1984). The primary aim of meta-analysis is inferring (non-causal) generalizations about substantive issues from a set of studies directly bearing on those issues (Jackson, 1980). In meta-analysis quantitative study outcomes from known research on a particular, well defined, question are statistically combined.

The methods used in meta-analysis are not new. The principles of meta-analysis are the same as those governing "ordinary" survey research. Typical steps taken in meta-analysis are: a precise definition of the research problem, data collection (i.e. collection of relevant articles or papers), coding of the variables of interest, and statistical analysis (Wolf, 1986). In general, the dependent variable "study-outcome" is operationalized in two ways: both the significance level (p-value) and an effect-size measure are coded. Furthermore, background variables such as year of publication, and source of publication are routinely coded, just as age and sex are routinely asked in a survey. Also, several research design characteristics of each study are coded (e.g. sampling method, type of subjects). This coding process results in a data matrix in which the cases (or rows) are the research studies of interest for the meta-analysis. Standard statistical procedures can then be used. In other words, the basic idea is to apply statistical methods, with the published statistics in previous studies of interest as the data (Walberg and Haertel, 1980). This distinguishes meta-analysis from the more traditional, narrative forms of literature review (Bangert-Drowns, 1986).

Retrieval and selection of the studies reviewed

We started our meta-analysis with an on-line computer search of the relevant literature. The abstracting services used were: Psychological Abstracts (1967 - 1986), Sociological Abstracts (1963 - 1986), Dissertation Abstracts (1861 - 1986), and Dialog/SSCI (Social Sciences Citation Index, 1972 - 1986). The following key words were used, both single and in combination: artifact, bias, comparison, data collection method, face to face, interview, mail, personal, postal, response, response bias, response effect, response style, social desirability, survey, and telephone. In addition the abstracts of SRM, a Dutch documentation center in the field of social research methodology, were searched for the period 1979 to 1986. The reference lists of the studies found in this way were then searched to uncover additional studies.

Most studies found (81%), were done in the U.S.A. This could partly be a result of the data bases available for the computer search. In order to avoid "retrieval bias", we published an appeal for research articles in three European Newsletters.

Studies were included in the meta-analysis when they empirically compared telephone and face to face interviews with mail surveys or self-administered questionnaires. Articles that only reviewed past literature, without presenting any new material, were not included. A second important criterion for inclusion was that the study reported at least one indicator for "data quality". Therefore, papers reporting response rates only, were not included.

In total 28 articles and reports were found in which mail and face to face surveys were compared on the quantity and the quality of the data, and 19 articles and reports in which mail and telephone surveys were compared. Fourteen different journals in the domains of psychology, sociology, marketing and opinion research, medicine, and criminology, provided the relevant literature. The oldest reference was published in 1947, the most recent one in 1987. Some studies

were (partly) reported in more than one article or paper. In order to avoid dependencies between the cases in the statistical analyses, the unit of analysis or case in this meta-analysis is **a study**, and not an article or paper (Bangert-Drowns, 1986; Rosenthal and Rubin, 1986). This explains why the number of cases reported in the analysis is less than respectively 28 and 19.

The coding of the studies

Data quality is a complex and fuzzy concept (Bailar, 1984; O'Toole, Battistuta, Long, and Crouch, 1986). Especially in a study of subjective phenomena (i.e. attitudes, beliefs or other attributes, which cannot be observed directly), it is difficult, if not impossible, to assess the correctness of the answers (cf. Turner and Martin, 1984). In those cases various proxy variables, or indicators for the quality of the data have been used (Groves, 1978). As a result, a large variety of different indicators of data quality can be found in empirical comparisons of face to face and telephone surveys. Only those indicators for data quality used in at least three studies, were coded for the meta-analysis.

These indicators are:

(1) Accuracy or response validity; for this indicator the answer of the respondent is checked against the "true" value as found in official records (e.g. the possession of a drivers license). This indicator is only applicable when validating information is available (cf. Sudman and Bradburn, 1974)

(2) Absence of social desirability bias (SD bias); inversely proportional to the number of socially desirable answers on a particular question. An answer is said to be socially desirable, when that specific answer is more determined by what is acceptable in society than by the real situation (cf. DeMaio, 1984).

(3) Item response, inversely proportional to the number of no answer or missing data per question (excluding do not know-responses).

(4) Similarity of response distributions obtained by different modes. Indicated by the absence of a significant difference between the proportions obtained under the different modes. This indicator, though often used, is only a very rough indicator for data quality.

Besides these four indicators of data quality, a fifth dependent variable was used: response rate, defined by the number of completed interviews divided by the total number of eligible sample units (cf. Groves and Kahn, 1979; Kviz, 1977).

A coding schedule, partially based on Sudman and Bradburn (1974), was used. We included background variables relating to the research report (e.g. journal, year and country of publication), and the study itself (e.g. type and size of sample, subject of the research and its saliency for respondents, equivalence of samples and questionnaires used in the study).

For each indicator of data quality the statistical significance level (p-value) of the differences between the modes was coded, and the direction of this difference (i.e. which mode offers data of better quality). For the indicator similarity coding of the direction of the difference was only done when the authors of the original study gave a convincing justification and a decision in terms of better quality could safely be made (e.g. more reporting of undesirable behavior such as using contraceptives as a Roman catholic).

To estimate the size of the mode effect, Cohen's (1969) effect size **e** was computed. Since Cohen's **e**, just as other commonly accepted effect size indices for cross-classifications such as Cramer's generalized phi, partly depends on the marginal distributions and the number or rows and columns of the table, another index was also computed. This index, whch we call the Standard Norm for Effect Size (SNES), is based on Cohen's (1969) definitions for small, medium, and large effect size (cf. Hox and De Leeuw, 1988; De Leeuw and Van der Zouwen, 1988). Key points on the SNES scale are 0 (no effect), 2 (small effect), 4 (medium

effect), 6 (large effect), and 8 (maximum effect).

The unit of analysis is a study (Bangert-Drowns, 1986; Wolf, 1986); when a particular study used more than one measure of the same construct (e.g. item-(non)response for several questions), effect sizes were combined by taking the mean effect size prior to the coding (Rosenthal and Rubin, 1986). All the studies were coded by one of the authors; problem cases were discussed with a statistician and with the co-author.

RESULTS

Mail and face to face surveys compared

In total, 27 comparisons between mail and face to face interviews were coded. In five cases relevant information was not available and/or reinterviews of the same respondents were used; those cases were excluded from further analysis. The remaining 22 cases have been published between 1947 and 1986. For a short description of the studies see Table 1.

The most frequently used outlets for research reports were the Journal of the American Statistical Association (3 times), and Public Opinion Quarterly (3).

A variety of subjects was covered in the questionnaires, with a dominant role for questions about health (7 cases). In nine cases the surveys studied dealt primarily with questions about (biographical) facts, in eight cases with questions about behavior, in four with questions on attitudes, and in one case with questions about emotions.

The samples used varied strongly with respect to the number of respondents (standard deviation: 1714), with an average sample size of 1457. In most cases (13), a random sample was drawn; in seven cases a convenience sample was used, and in two cases a panel.

Table 1. Comparisons of questionnaires with face to face interviews. Major characteristics of the studies reviewed. Author, year of publication, subject, indicators coded for meta-analysis and summary conclusion as given in the original article a) Some studies are partly reported in more than one article, the first author and year of publication is then given in parentheses.

1st author	Year	Subject	Indicator	Conclusion
Aakster	1968	Comparison selfadministered questionnaire, interviewer present with mail survey on health		Mail survey more item nonresponse on complex questions, SAQ more on embarrassing questions
Assael	1982	consumer/ business	Accuracy	Mail most effective reducing resp.error
Cannel (Cannel 1964)	1963	health	Accuracy S.D. bias	When respondent has records selfenumerative more accurate, no diff. in S.D.bias
Dillman	1984	housing	Item resp Similarity	Mail surveys less extremeness
Ellis	1947	Reinvestigation of previously interviewed on emotions, relations		Answers on questionnaire more incriminating
Hinkle	1978	health/ mental	Similarity	Mail higher SES, more neutral & negative answers
Hochstim (& Hochstim 1962)	1967 b)	health/ general	Item resp Similarity	Data collection strategies used,
Hochstim	1967	health/ cervical cytology	Accuracy Item resp Similarity	proved to be practically interchangeable
Knudsen	1967	attitude/ relations/ sex	Similarity	Questionnaire lower proportion women with restrictive norms
Krohn	1975	selfreport/ delinquency	Similarity	No reason to assume one technique is any more valid than other
Locander	1976	personal facts	Accuracy	None of the methods differed significant
Mangione	1982	drinking	Similarity S.D. bias	No significant diff.; in person somewhat more drinking
McDonagh	1965	General reinterview after mailed quest.		No statistically significant diff.
Nederhof	1984	equity	Similarity	More altruistic in face to face interv.

(Table 1 continues on next page)

Table 1. Continuation.

1st author	Year	Subject	Indicator	Conclusion
Nuckols	1964	finances/ insurance	Accuracy S.D. bias Itemresp	Mail panel showed up well, more accurate, less bias
O'Dell	1962	consumer (panels)	Itemresp Similarity	Selection of method is decision based on optimum allocation of research dollar
O'Toole	1986	health/ veterans	Accuracy Item resp	Overall no mode differences, mail less complete
Siemiatycki	1979	health/ community needs	Accuracy Itemresp	Mail greater validity & willingness to answer
Sudman	1965	religion/ education	Similarity	No large differences SAQ seems to give better measure of feelings
Van Amstel	1981	Comparing SAQ, interviewer present with mail survey on health		In mail survey more personal problems were reported
Van Sonsbeek	1983	health	Accuracy Itemresp Similarity	Results on medical consumption are very similar
Walsh	1967	personal/		No method elicits
Walsh	1968	education	Similarity	more accurate self-
Walsh	1969	(three replications)		reports than another
Wierdsma	1985	health	Similarity	Mail questionnaires are not second to the interview method
Wiseman	1972	several topics	Similarity	Responses not always independent of method
Zeiner-Henrikson	1972	Mail survey on cardiac pain of earlier interviewed patients		Two methods yield much variety, not interchangeable

a) Country of origin of the studies was the U.S.A., with the exception of Aakster, Nederhof, Van Amstel, Van Sonsbeek, and Wierdsma (The Netherlands), O'Toole (Australia), Siemiatycki (Canada), and Zeiner-Henrikson (Norway).
b) Two separate studies are reported in one article.

The response rate of the face to face interview is higher than that of the mail surveys: for the face to face interview a mean response rate of 75 percent is reported (standard deviation is 11%) versus a mean response rate of 70 percent for the mail survey (standard deviation 7%).

For the data quality indicators accuracy, item response, absence of social desirability bias, and similarity the p-values were combined over the cases, using the z-transformation (Cooper 1979; Rosenthal, 1978). For each indicator the combined p-value was less than 0.01.

Table 2 summarizes the results for the sizes of the mode effects. In most comparisons only one or two indicators of data quality were used. As a consequence, the data points for each indicator are limited and differ in number.

Table 2. Comparison of data quality in mail and face to face Surveys; mean effect size, mean, standard deviation, minimum and maximum SNES-value, number of comparisons in analysis

Indicator	Mean e	SNES a)				N of cases
		Mean	SD	Min	Max	
Accuracy	0.01	1.0	0.9	0	2	6
Abs. S.D. bias	0.01	1.3	1.2	0	2	3
Item (non)resp	0.01	1.0	0.9	0	2	8
Similarity	0.02	1.3	0.9	0	3	15

a) SNES (Standard Norm Effect Size) is measured on a nine-point scale: (0) no effect, (4) medium effect, and (8) maximum effect.

When we inspect Table 2 we see that the mode differences found, though statistically significant, are small. Even the largest effect size found (similarity: mean SNES = 1.3), is according to Cohen's definitions of effect size small indeed.

When possible the direction of the differences was estimated. For the indicators accuracy, absence of social desirability bias, and similarity the overall effect was in favor of the mail questionnaire, with one exception. In one

study (Cannell & Fowler, 1963) a (non significant) difference was found with slightly more socially undesirable and embarrassing answers in the face to face condition. In this specific experiment the respondents had to sign their name on the questionnaire in the self-enumerative condition. The authors use the concept anonymity in order to explain this unpredicted effect. They argue that anonymity rather then the presence or absence of an interviewer causes a reduction of social desirability bias when using self-enumerative procedures (Cannel & Fowler, 1963).

For the indicator item response the differences were in general in favor of the face to face interview. However, when respondents were asked about their income, a different result was found. In that case the mail survey produced less item non response (Nuckols, 1964; Siemiatycky, 1979; Van Sonsbeek & Stronkhorst, 1983). This is in accordance with the findings of Aakster (1968) who reports that in a mail survey less item nonresponse was found on embarrassing questions than when a questionnaire was handed over by an interviewer.

In two of the articles coded, some interesting additional information was given concerning the extremeness of responses. Both Dillman and Mason (1984) in the United States and Van Sonsbeek and Stronkhorst 1983) in the Netherlands, found that in the face to face interview respondents are more likely to use the extreme positive end of a scale.

Mail and telephone surveys compared

In total, 17 comparisons between mail and telephone interviews were coded. In four cases relevant information was not available; those cases were excluded from further analysis. The remaining 13 cases have been published between 1967 and 1987. For a short description of the cases see Table 3.

Table 3. Comparisons of questionnaires with telephone interviews. Major characteristics of the studies reviewed. Author, year of publication, subject, indicators coded for meta-analysis and summary conclusion as given in the original article a). Some studies are partly reported in more than one article, the first author and year of publication is then given in parentheses.

1st author	Year	Subject	Indicator	Conclusion
Assael	1982	consumer/ business	Accuracy	Mail most effective reducing resp.error
Bishop	1987	Comparison of response effects, two cross-cultural experiments		Order effects less likely in mail, form effects as likely
Dillman	1984	housing	Item resp Similarity	Mail surveys less extremeness
Hinkle	1978	health/ mental	Similarity	Mail higher SES, more neutral & negative answers
Hochstim (& Hochstim	1967 1962)	health/ general	Item resp Similarity	Data collection strategies used,
Hochstim	1967	health/ cervical cytology	Accuracy Item resp Similarity	proved to be practically interchangeable
Locander	1976	personal facts	Accuracy	None of the methods differed significant
Mangione	1982	drinking	Similarity	Results equivalent, mail less complete
McGuire	1977	Media habits, but non-equivalent groups		Combination mail & telephone is best
O'Toole	1986	health/ veterans	Accuracy Item resp	Overall no mode differences, mail less complete
Prawl	1976	Educational evaluation, no mode comparison made		Telephone data seem highly credible
San-Augustine	1978	attitudes on blacks	Similarity	Mail low response, more liberal; tel. preferable
Siemiatycki (Siemiatycki	1979 1984)	health/ community needs	Accuracy Item resp	Mail greater validity & willingness to answer
Sudman	1974	consumer/ expenditure	Similarity	Daily telephone not as complete as diary
Wheatly	1973	consumer/ marketing	Similarity	No difference in nature of response

(Table 3 continues on next page)

Table 3. Continuation.

1st author	Year	Subject	Indicator	Conclusion
Williams	1976	media habits	Similarity	Mail more likely premeditated resp.
Wiseman	1972	several topics	Similarity	Responses not always independent of method

a) Country of origin of the studies was the U.S.A., with the exception of Bishop (America/Germany), O'Toole (Australia), and Siemiatycki (Canada).
b) Two separate studies are reported in one article.

The most frequently used outlets for research reports were the Journal of the American Statistical Association (3), and the Journal of Marketing Research (2).

A variety of subjects was covered in the questionnaires, again with a predominant role for questions about health (4). In six cases the surveys studied dealt primarily with questions about behavior, in four cases with questions about attitudes, and in three cases with questions about (biographical) facts.

The samples used varied strongly with respect to the number of respondents (standard deviation: 1155), with an average sample size of 1254. In most cases (10), a random sample was drawn; in the remaining three cases a convenience sample was used.

The response rate of the telephone interview is higher than that of the mail surveys: for the telephone interview a mean response rate of 74 percent is reported (standard deviation 11%) versus a mean response rate of 70 percent for the mail survey (standard deviation 7%).

For the indicator absence of social desirability bias no data were available. For each of the remaining indicators - accuracy, item response, and similarity - the p-values were combined over the cases, using the z-transformation (Cooper 1979; Rosenthal, 1978). For each indicator the combined p-value was less than 0.01. Table 4 summarizes the results for the mode effects on data quality. In most comparisons only

one or two indicators of data quality were used. As a consequence, the data points for each indicator are limited and differ in number.

Table 4. Comparison of data quality in mail and telephone surveys; mean effect size, mean, standard deviation, minimum and maximum SNES-value, number of comparisons in analysis

Indicator	Mean e	SNES a)				N of cases
		Mean	SD	Min	Max	
Accuracy	0.01	0.8	1.0	0	2	4
Abs. S.D. bias	--	--	--	-	-	0
Item (non)resp	0.01	0.8	1.1	0	2	5
Similarity	0.02	1.1	1.2	0	3	8

a) SNES (Standard Norm Effect Size) is measured on a nine-point scale: (0) no effect, (4) medium effect, and (8) maximum effect.

The mode differences found, though statistically significant, are again small.

When possible the direction of the differences was estimated. For the indicators accuracy and similarity the differences found were in favor of the mail survey. For the indicator item response the differences were in favor of the telephone survey. An exception was the study of Siemiatycki (1979), who found that so called sensitive questions (e.g. income) had less item non response in the mail strategy. In general, when compared with telephone interviews, mail surveys tend to have more item non response and tend to do better with sensitive questions.

Going back to the individual studies we see that sometimes additional indicators for data quality were reported. When we take these in consideration, an interesting pattern emerges. It is harder to have people answer questions in a mail survey. Both the overall non response and the item non response tend to be higher in mail surveys. But when the questions are answered, the resulting data seem to have a higher quality and well known response effects seem to be less influential in mail surveys. For instance Bishop and

Hippler (1988) found in two cross-culturally replicated experiments that order effects are significantly less likely to occur in a self-administered questionnaire than in a telephone survey; question wording and question form effects were as likely to occur in both methods. Furthermore, Dillman and Mason (1984), in a methodological study on survey modes, found that mail respondents are less likely than telephone (and face to face) respondents to use the extreme response category on the positive end of the scale.

Summary and Discussion

When we look at the results presented in this paper, a clear picture emerges. First of all, both face to face and telephone interviews give higher response rates and less item nonresponse than mail surveys. Secondly, mail surveys result in more accurate answers. And, in general, mail questionnaires perform "better" with more embarrassing questions (e.g. drinking behavior, health questions, personal feelings). Furthermore, in a review of comparisons between face to face and telephone surveys De Leeuw and Van der Zouwen (1988) found that on all indicators the face to face interview scored slightly better than the telephone interview.

In summary: when regarding the quality of the data, both mail and face to face surveys have certain strong and weak points. Using a telephone interview one is always slightly worse off. This does not mean that one should never use telephone interviews. There are more factors (e.g. costs, completion time, ease of integration into a highly computerized research environment) relevant for choosing the optimum data collection technique (cf. O'Dell, 1962).

One should bear in mind that the studies analyzed in this review were all methodological experiments on the influence of the data collection method utilized on data quality. In general, these studies paid much attention to the design and implementation of these methods. This shows up in the high

mean response rate (± 70%). Exploring the data, a tendency can be noted for studies with a relatively low quality (as indicated by size and type of sample used, and number of methodological details presented) to report somewhat larger effects than studies of a relatively high quality. This suggests that under field conditions, where the pressure for high methodological standards may be less predominant, the effects of data collection method on data quality may be stronger.

In this article we have given a tentative and conditional answer to the question which method is best. We have not answered the question why a specific method is better. In the literature on survey methodology several hypotheses concerning the origin of certain mode differences can be found.

For instance, Siemiatycki (1979) suggests that self-administered questionnaires are in general completed in a more relaxed atmosphere and without any time pressure. This allows the respondents enough time to check records or to consult others. Galtung (1967) remarked that in a self-enumerative questionnaire the respondent notes the answers down and not an interviewer. This provides the respondent with an extra visual check on the correctness of the answer. Both arguments imply a better performance on the indicator accuracy for the mail survey.

Apart from being better in accuracy, mail surveys also give better results when sensitive questions are used. Both the results on social desirability and on similarity point in this direction. Cannel and Fowler (1963) suggest that it is a feeling of anonymity rather than the mere presence or absence of the interviewer, which causes this effect. Bradburn and Sudman (1974) further elaborate this idea and hypothesize that self presentation of the respondent is an important factor in how respondents answer sensitive questions. In their opinion self-administered questionnaires are more private and do not require a direct revelation of the self to other persons who are present. The presence of an interviewer on the telephone or in person may also

influence the responses in a positive way. A skilled interviewer can always get the refusal rate down (Galtung, 1967). Not only the refusal rate of the interview as a whole, but also the refusal rate of separate questions. A skilled interviewer can use probes, explain ambiguous questions and by doing this keep the item nonresponse low. In addition, during a face to face interview nonverbal ways of communication can be used to motivate and reinforce respondents (Sykes and Hoinville, 1985; Dillman and Mason 1984), resulting in a lower amount of (item) nonresponse.

Finally, reviews are always one step behind the actual state of the art. In the articles reviewed here, no explicit reference was made to the use of computer assisted data collection techniques. As the acceptance and the use of the computer is increasing (for instance see Spaeth, 1987 and Berry, 1988 on the use of CATI) it is of paramount importance that further research in this field is done, and that controlled experiments and comparisons of paper and pencil with computer assisted techniques are performed.

References

Aakster, C.W. "Vergelijking van Schriftelijke en Mondelinge Enquete (Comparison of Mail and Home Survey)". **Sociologische Gids**, 1968, 15, 322-326.

Bailar, B., "The Quality of Survey Data", **Proceedings of the section on survey research methods of the American Statistical Association**, A.S.A., Washington D.C., 1984.

Bangert-Drowns, R.L., "Review of Developments in Meta-Analytic Method", **Psychological Bulletin**, 1986, 99, 388-399.

Baumgartner, R.M. and Heberlein, T.A. "Recent Research on Mailed Questionnaire Response Rates". In: D.C. Lockhart (ed) **Making Effective Use of Mailed Questionnaires**, Jossey-Bass, San Francisco, 1984.

Berry, S. "The Use of Cati". In: R.M. Groves et al (eds) **Telephone Survey Methodology**, Wiley, New York, 1988.

Bishop, G.F. and Hippler, H-J. "A Comparison of Response Effects in Self-administered and Telephone Surveys". In: R. M. Groves et al (eds) **Telephone Survey Methodology**, New York: Wiley, 1988.

Cannel, Ch.F. and Fowler, F.J. jr. "Comparison of a Self Enumerative Procedure and a Personal Interview: a Validity

Study". **Public Opinion Quarterly**, 1963, 27, 250-264.
Cohen, J., **Statistical Power Analysis for the Behavioral Sciences.** Academic Press, New York, 1969.
Cooper, H.M. "Statistically combining independent studies: a meta-analysis of sex differences in conformity research". **Journal of Personality and Social Psychology**, 1979, 37, 131-146.
De Leeuw, E.D. and Hox, J.J., "Artifacts in Mail Surveys: The Influence of Dillman's Total Design Method on the Quality of Responses". In: W.E. Saris and I.N. Gallhofer (eds.), **Sociometric Research**, Vol. II, McMillan, London, 1988, 61-73.
De Leeuw, E.D. and Van der Zouwen, J. "Data Quality in Telephone and Face to face Surveys: a Comparative Meta-analysis". In: R.M. Groves et al (eds) **Telephone Survey Methodology**, Wiley, New York, 1988.
De Maio, T.J. "Social Desirability and Survey Measurement: a Review". In: Turner, Ch.F. and Martin, E. **Surveying Subjective Phenomena**, Vol II. Russel Sage Foundation, New York, 1984.
Dillman, D. A. and Mason, R. G. **The influence of survey method on question response.** Paper presented at the annual meeting of the American Association for Public Opinion Research, Delavan, Wisconsin, 1984.
Galtung, J. **Theory and Methods of Social Research.** Allen & Unwin, London, 1967.
Glass, G.V., McGaw, B., and Smith, M.L., **Meta-Analysis in Social Research.** Sage, Beverley Hills, 1981.
Goyder, J.C. "Further Evidence on Factors Affecting Response Rates to Mailed Questionnaires". **American Sociological Review**, 1982, 47, 550-553.
Groves, R.M., "On the Mode of Administering a Questionnaire and Response to Open-Ended Items", **Social Science Research**, 1978, 7, 257-271.
Groves, R.M. and Kahn, R.L. **Surveys by Telephone; a National Comparison with Personal Interviews.** Academic Press, New York, 1979.
Heberlein, T.A. and Baumgartner, R.M. "Factors Affecting Response Rates to Mailed Questionnaires: a Quantitative Analysis of the Published Literature". **American Sociological Review**, 1978,43, 447-462.
Herman, J.B., "Mixed-Mode Data Collection: Telephone and Personal Interviewing", **Journal of Applied Psychology**, 1977, 62, 399-404.
Houstom, M.J. and Ford, N.M. "Broadening the Scope of Methodological Research on Mail Surveys". **Journal of Marketing Research**, 1976, 13, 397-403.
Hox, J.J. and De Leeuw, E.D. "Meta-analysis on Studies using Categorical Data". In: G.G.H. Jansen & W. van Schuur (eds), 1988, **The many faces of multivariate analysis.** Groningen:

RION.
Jackson, G.B. "Methods for Integrative Reviews", **Review of Educational Research**, 1980, 50, 428-460.
Jones, W.H. and Lang, J.R., "Reliability and Validity Effects under Mail Survey Conditions", **Journal of Business Research**, 1982, 10, 339-353.
Kanuk, L. and Berenson, C. "Mail Surveys and Response Rates: a Literature Review". Journal of Marketing Research, 1975, 12, 440-453.
Kviz, F.J. "Towards a standard definition of response rate". **Public Opinion Quarterly**, 1977, 41, 265-267.
Light, R.J.; and Pillemer, D.B., **Summing Up; The Science of Reviewing Research.** Harvard University Press, Cambridge MA, 1984.
Linsky, A.S. "Stimulating Responses to Mailed Questionnaires: a Review". **Public Opinion Quarterly**, 1975, 39, 82-101.
Locander, W.; Sudman, S.; and Bradburn, N. "An investigation of interview method, threat and response distortion". **Journal of the American Statistical Association**, 1976, 71, 269-275.
Nuckols, R.C. "Personal Interview versus Mail Panel Study". **Journal of Marketing Research**, 1964, 1, 11-16.
O'Toole, B.I., Battistuta, D., Long, A., and Crouch, K., "A Comparison of Costs and Data Quality of Three Health Survey Methods: Mail, Telephone, and Personal Home Interview", **American Journal of Epidemiology**, 1986, 124, 317-328.
Rosenthal, R. "Combining Results of Independent Studies". **Psychological Bulletin**, 1978, 85, 185-193.
Rosenthal, R and Rubin, D.B., "Meta-Analytic Procedures for Combining Studies with Multiple Effect Sizes", **Psychological Bulletin**, 1986, 99, 400-406.
Siemiatycki, J. "A Comparison of Mail, Telephone, and Home Interview Strategies for Household Health Surveys". **American Journal of Public Health**, 1979, 69, 238-245.
Spaeth, M.A. "Cati Facilities at Survey Research Organizations". **Survey Research**, 1987, 3-4, 18-22.
Sudman, S. and Bradburn, N.M., **Response Effects in Surveys; A Review and Synthesis.** Aldine, Chicago, 1974.
Sykes, W. and Hoinville, G. **Telephone Interviewing on a Survey of Social Attitudes.** Social and Community Planning Research, London, 1985.
Thornberry, O., Nicholls, W.L., and Kulpinsky, S. **Data Collection Methods in Federal Statistical Surveys.** American Statistical Association, Proceedings of the 1982 conference, section on research methods.
Turner, Ch.F. and Martin, E. **Surveying Subjective Phenomena**, Vol I. Russel Sage Foundation, New York, 1984.

Waelberg, H.J. and Haertel, E.H. "Research integration: an introduction and overview". **Evaluation in Education**, 1980, 4, 5-10.

Van Amstel, R. "Postenquete of Bezoekenquete (Mail Survey or Personal Delivered Selfadministerd Questionnaire)". **Tijdschrift voor Sociale Geneeskunde**, 1981, 59, 164-169.

Van Sonsbeek, J.L.A. and Stronkhorst L.H. **Vergelijking van drie Waarnemingsvarianten bij de Meeting van Medische Consumptie (A Comparison of Data Collection Methods in the Measurement of Medical Consumption)**. Centraal Bureau voor de Statistiek, 1983.

Wolf, F.M., **Meta-analysis; quantitative methods for research synthesis**. Sage, Beverly Hills, 1986.

Yu, J. and Cooper, H. "A Quantitative Review of Research Design Effects on Response Rates to Questionnaires". **Journal of Marketing Research**, 1983, 20, 36-44.

Appendix

Articles in the meta-analysis: Comparisons of mail and face to face surveys. Some studies were partly reported in more than one article or paper. These are marked with an asterisk.

Aakster, C.W. Vergelijking van schriftelijke en mondelinge enquete (Comparison of mail and home survey). **Sociologische Gids**, 1968, 15, 322-326.

Assael, Henry and Keon, John. Non sampling versus sampling errors in survey research. **Journal of Marketing**, 1982, 46, 114-123.

*Cannel, Charles F. and Fowler, Floyd J. jr. A note on interviewer effect in self-enumerative procedures. **American Sociological Review**, 1964, 24, 270.

*Cannel, Charles F. and Fowler, Floyd J. jr. Comparison of a self enumerative procedure and a personal interview: a validity study. **Public Opinion Quarterly**, 1963, 27, 250-264.

Dillman, Don A. and Mason, Robert G. **The influence of survey method on question response.** Paper presented at the annual meeting of the American Association for Public Opinion Research, Delavan, Wisconsin, 1984.

Ellis, Albert. Questionnaire versus interview methods in the study of human love relationships. **American Sociological Review**, 1947, 12, 541-553.

Hinkle, Andrew Lee and King, Glen D. A comparison of three survey methods to obtain data for community health program planning. **American Journal of Community Psychology**, 1978, 6, 389-397.

*Hochstim, Joseph R. **Comparison of three information gathering strategies in a population study of sociometrical variables.** Proceedings of the Social Statistics Section of the American Statistical Association, 1962, 154-159.

*Hochstim, Joseph R. A critical comparison of three strategies for collecting data from households. **Journal of the American Statistical Association**, 1967, 62, 976-989.

Knudsen, Dean D., Pope, Hollowel, and Irish, Donald P. Response differences to questions on sexual standards: an interview-questionnaire comparison. **Public Opinion Quarterly**, 1967, 31, 290-297.

Krohn, Marvin.,Waldo, Gordon P., and Chiricos, Theodore G. Self reported delinquency: a comparison of structured interviews and self-administerd checklists. **The Journal of Criminal Law and Criminology**, 1975, 65, 545-553.

Locander, William, Sudman, Seymour and Bradburn, Norman. An investigation of interview method, threat and response distortion **Journal of the American Statistical Association,**

1976, 71, 269-275.
Mangione, Thomas W., Hingson, Ralph and Barrett, Jane. Collecting sensitive data. **Sociological Methods and Research**, 1982, 10, 337-346.
McDonagh, Edward C. and Rosenblum, A. Leon. A comparison of mailed questionnaires and subsequent structured interviews. **Public Opinion Quarterly**, 1965, 29, 131-136.
Nederhof, Anton J. Visibility of response as a mediating factor in equity research. **The Journal of Social Psychology**, 1984, 122,211-215.
Nuckols, Robert C. Personal interview versus mail panel study. **Journal of Marketing Research**, 1964, 1, 11-16.
O'Dell, William F. Personal interviews or mail panels. **Journal of Marketing**, 1962, 26, 34-39.
O'Toole, Brian I., Battistuta, Diana, Long, Ann and Crouch, Keitha. A comparison of costs and data quality of three health survey methods: mail, telephone and personal home interview. **American Journal of Epidemiology**, 1986, 124, 317-328.
Siemiatycki, Jack. A comparison of mail, telephone, and home interview strategies for household health surveys. **American Journal of Public Health**, 1979, 69, 238-245.
Sudman, Seymour, Greely, Andrew, and Pinto Leonard. The effectiveness of self-administerd questionnaires. **Journal of Marketing Research**, 1965, 2, 293-297.
Van Amstel, R. Postenquete of bezoek enquete (Mail survey or personal delivered selfadministerd questionnaire). **Tijdschrift voor Sociale Geneeskunde**, 1981, 59, 164-169.
Van Sonsbeek, J.L.A. and Stronkhorst L.H. **Vergelijking van drie waarnemingsvarianten bij de meeting van medische consumptie (A comparison of data collection methods in the measurement of medical consumption)**. Centraal Bureau voor de Statistiek, 1983.
Walsh, W. Bruce. Validity of self report. **Journal of Counseling Psychology**, 1967, 14, 18-23.
Walsh, W. Bruce. Validity of self report: another look. **Journal of Counseling Psychology**, 1968, 15, 180-188.
Walsh, W. Bruce. Self report under socially undesirable and distortion conditions. **Journal of Counseling Psychology**, 1969, 16, 569-574.
Wierdsma, A.I. and Garretsen, H.F.L. Gezondheids enquete per post of op bezoek (health surveys: by mail or home interview). **Tijdschrift voor Sociale Gezondheidszorg**, 1985, 63, 592-595.
Wiseman, Frederich, Methodological bias in public opinion surveys. **Public Opinion Quarterly**, 1972, 36, 105-108.
Zeiner-Henriksen, Truls. Comparison of personal interview and postal inquiry methods for assessing prevalence of angina and possible infarction. **Journal of Chronical Diseases**, 1972, 25, 433-440.

Articles for the meta-analysis: Comparisons of mail and telephone surveys. Some studies were partly reported in more than one article or paper. These are marked with an asterisk.

Assael, Henry and Keon, John. Non sampling versus sampling errors in survey research. **Journal of Marketing**, 1982, 46, 114-123.

Bishop, George F. and Hippler, Hans-J. **A comparison of response effects in self-administered and telephone surveys.** Paper presented at the international conference on telephone methodology, Nov. 1987, North Carolina. To appear in: R. M. Groves et al (eds) Telephone Survey Methodology, New York: Wiley.

Dillman, Don A. and Mason, Robert G. **The influence of survey method on question response.** Paper presented at the annual meeting of the American Association for Public Opinion Research, Delavan, Wisconsin, 1984.

Hinkle, Andrew Lee and King, Glen D. A comparison of three survey methods to obtain data for community health program planning. **American Journal of Community Psychology**, 1978, 6, 389-397.

*Hochstim, Joseph R. **Comparison of three information gathering strategies in a population study of sociometrical variables.** Proceedings of the Social Statistics Section of the American Statistical Association, 1962, 154-159.

*Hochstim, Joseph R. A critical comparison of three strategies for collecting data from households. **Journal of the American Statistical Association**, 1967, 62, 976-989.

Locander, William, Sudman, Seymour and Bradburn, Norman. An investigation of interview method, threat and response distortion **Journal of the American Statistical Association**, 1976, 71, 269-275.

Mangione, Thomas W., Hingson, Ralph and Barrett, Jane. Collecting sensitive data. **Sociological Methods and Research**, 1982, 10, 337-346.

McGuire, Bernadette and Leroy, David J. Comparison of mail and telephone methods of studying media contactors. **Journal of Broadcasting**, 1977, 21, 4, 391-400.

O'Toole, Brian I., Battistuta, Diana, Long, Ann and Crouch, Keitha. A comparison of costs and data quality of three health survey methods: mail, telephone and personal home interview. **American Journal of Epidemiology**, 1986, 124, 317-328.

Prawl, Warren L. and Jorns, William J. Reviewing county extension programs. **Journal of Extension**, 1976, 14, 11-17.

San Augustine, Andre J. and Friedman, Hershey H. **The use of the telephone interview in obtaining information of a sensitive nature: a comparative study.** Proceedings of the

section on survey research methods of the American Statistical Association, 1978, 559-561.
*Siemiatycki, Jack. A comparison of mail, telephone, and home interview strategies for household health surveys. **American Journal of Public Health**, 1979, 69, 238-245.
*Siemiatycki, Jack and Campbell, Sally. Nonresponse bias and early versus all responders in mail and telephone surveys. **American Journal of Epidemilogy**, 1984, 120, 2, 291-301.
*Siemiatycki, Jack, Campbell, Sally, Richardson, Lesley and Aubert, Dennis. Quality of response in different population groups in mail and telephone surveys. **American Journal of Epidemilogy**, 1984, 120, 2, 302-314.
Sudman, Seymour and Ferber, Robert. A comparison of alternative procedures for collecting consumer expenditure data for frequently purchased products. **Journal of Marketing Research**, 1974, 11, 128-135.
Wheatley, John J. Self administered written questionnaires or telephone interviews. **Journal of Marketing Research**, 1973, 10, 94-96.
Williams, Wenmouth jr. and Leroy, David J. Alternative methods of measuring public radio audiences: a pilot project. **Journalism Quarterly**, 1976, 53, 3, 516-521.
Wiseman, Frederich, Methodological bias in public opinion surveys. **Public Opinion Quarterly**, 1972, 36, 105-108.

Ontvangen: xx-07-1988
Geaccepteerd: 20-04-1989