PREVENTION IS THE BETTER CURE: HOW TO REDUCE MISSING DATA

Edith D. de Leeuw[1]

Abstract

Although item nonresponse can never be prevented totally, it can be reduced considerably by using an optimal data collection design. Thereby providing us not only with more data to use, but also with helpful auxiliary information for better imputation and adjustment. In this contribution a concise typology of missing data patterns and their sources of origin is presented. Based on this typology, sources of missing data are distinguished, followed by a discussion on how item nonresponse can be prevented.

Key words: item nonresponse, causes of missingness, ignorability, question wording, questionnaire development, data collection mode, survey

1. Introduction

Item nonresponse can pose serious problems to researchers, or in the words of Sherlock Holmes: 'Data! data! data! I can't make bricks without clay.'[2] When item nonresponse occurs, a unit (e.g., a person) provides data, but for some reason data on particular items or questions are not available for analysis (cf. Lessler & Kalsbeek, 1992). In other words, there are gaps in the data matrix. Not so long ago, researchers simply ignored the problem and restricted their analysis to observed values or to complete cases. However, cases with missing data and the easy solutions of 'pairwise' and 'listwise' deletion, result in loss of information; estimates will be less efficient, and statistical tests will have less power (cf. Huisman, 1999).

---

[1] MethodikA, Plantage Doklaan 40, NL-1018 CN Amsterdam, Holland

[2] Sir Arthur Conan Doyle, The copper beeches; The adventures of Sherlock Holmes, London, Penguin Books, 1981, p. 268.

Furthermore, the results of the analysis may be biased, because of the possibility of systematic differences between units that respond to a particular item and those that do not respond. Practical problems are that analyses may be performed on different data (sub)sets, and can therefore be inconsistent with each other (cf. Lessler & Kalsbeek, 1992).

More modern strategies to cope with missing data are imputation (Little & Rubin, 1987) or direct estimation (Arbuckle, 1996; Vermunt, 1996). The availability of modern and user friendly software will undoubtedly increase the use of imputation and direct estimation techniques. For an overview of several imputation techniques see Martin et al. (1996) and Huisman (this issue); for a comparison of available software see Hox (this issue). Still, prevention is the better cure. Reduction of item nonresponse will lead to less imputation in a data set and, ipso facto, to more data to base the imputation on. It will produce the clay for the bricks.

In this article, I will present an overview of missing data patterns and their sources of origin. Based on this typology, mechanisms that may cause missing data are distinguished. The main part of this article will be devoted to prevention of item nonresponse. I will discuss questionnaire construction and lay-out, pretesting and pilot studies, the role of the interviewer, computer assisted interviewing, and measures for special populations and topics.

2. Typology of Missing Data

When data are collected with questionnaires, be it in a social survey, in opinion polling, in school research, or in educational or psychological testing, missing data occur. There are several types of missing data, and each can by caused in different ways and by numerous factors. To successfully prevent missing data it is necessary to define and understand the various types.

First of all, there is *unit nonresponse*, data for a whole unit of analysis are not available for statistical analysis. This can be caused because a unit could not be contacted, refused to cooperate, or did cooperate but the questionnaire got lost during data editing or analysis (cf. Lessler & Kalsbeek, 1992). Unit nonresponse falls outside the scope of this article. For a concise description of causes and treatment of unit nonresponse, see De Leeuw & Hox (1998); for a detailed treatment of nonresponse in interviews, see Groves & Couper (1998). Unit nonresponse is often called first-level nonresponse. *Item nonresponse* is referred to as secondlevel nonresponse. Data on particular items are *unavailable* for analysis, but the unit has participated. The term 'unavailable' is used on purpose. Whether or not an answer is counted as missing depends on the goal of the study. For instance, 'do-not-know' can be seen as a meaningful response to a question about voting intentions in an election poll. For other questions (e.g., income) 'do-not-know' has no informational value and is counted as missing. Therefore, an item is missing if the researcher interprets it as such, and decides that imputation is required. Thus, item nonresponse is defined as the failure to obtain *information* for a question within an interview or questionnaire: data are missing (cf. Groves, 1989).

Item missing data can by divided in three main groups: (1) the data are missing systematically by design, (2) all the data are missing after a certain point in the questionnaire (partial nonresponse), and (3) data are missing for some items for some respondents (item-nonresponse).

(Ad 1) Missing by design. In this case the researcher decides that certain questions will not be posed to certain persons. This may be because certain questions are not applicable to all respondents, and the questionnaire routing skips these questions. Another example is the use of a specific (factorial) design to administer different subsets of questions or stimuli to different persons. These designs are well known in experimental psychology and have their roots in early agricultural experiments. They have many practical advantages (e.g., less time or resources are needed, the respondent's task is shorter and less stressful, the data collection is more efficient). A prime illustration is the study into equality of income by Hermkens (1983). In this study, respondents were presented with 'vignettes' or small stories about persons and had to rate whether the reported income was fair. The descriptions contained in the vignettes varied systematically. Using a greco-latin square a total of 720 vignettes was designed. Each respondent had to rate a subsample of the vignettes. A third example of missing by design is the use of incomplete testing designs in educational testing, for instance adaptive or tailored testing (cf. Huisman, 1999).

In the case of missing by design, the decision to create a missing by not posing the question is controlled by the researcher. Using the known properties of the design, the resulting missing data can be handled statistically. For instance, in the case of vignettes in a factorial survey, multilevel analysis may be used (Hox, Kreft, & Hermkens, 1991). Huisman (1999: 79) points out that commonly used incomplete testing designs have known missing data mechanisms that are ignorable for inferences about the ability parameter (see Jansen, 1997).

(Ad 2) Partial nonresponse. After a certain point in time all items are missing. Two

well-known examples are panel mortality and breaking-off during an interview. In panel research and also in repeated measurement designs, part of the initial sample will not respond to subsequent questionnaires or interviews. Researchers are well aware of this phenomenon and methods have been developed to reduce non response and drop-out in panels (Freedman et al, 1980; Kasprzyk et al., 1989) and adjust for it (Van der Pol, 1989; Engel & Reinecke, 1994). Break-offs mostly occur in telephone interviews. At a certain point, the respondent disconnects and the remainder of the questions is not answered. When the break-off occurs early in the interview and only a few questions are answered, it is usually treated as a unit nonresponse. When the break-off occurs at the end of the interview and most questions are answered, the remaining unanswered questions are usually treated as item nonresponse.

(Ad 3) Item nonresponse. There are gaps in the data set for some persons for some individual items. Item missing data occur because the information is not provided, because the information provided for a certain item is not usable, or because usable information is lost. For example, a question may be skipped on purpose or by mistake, an answer is not known, an answer is not possible, an answer falls outside the range of permissible responses, an answer cannot be coded, or an answer is unreadable.

Data can be missing in several ways. They can be missing completely at random (MCAR), when the missingness is unrelated to the unknown value of the question in case, and is also unrelated to the values of other variables. For example, a question is skipped by mistake. When the missingness is related to the value of observed data, but not to the value of the question itself, it is said that the data are missing at random (MAR). For example, an older respondent has difficulty remembering an event. Finally, when the missingness is related to the answer to the question itself, the data are nonrandomly missing (NMAR). For example, a respondent says 'do-not-know', refuses, or skips a question, because his/her answer is socially undesirable (e.g., drinking a lot). These three missingness mechanisms are quite different and require specific data analysis strategies (cf. Little & Rubin, 1987). Whether or not missing data will bias the conclusion, depends on the missing data mechanism. If data are missing completely at random (MCAR), or even when data are missing at random (MAR), one may assume ignorability of the missing data mechanism. In that case relatively simple imputation procedures will work well. However, when the data are *not* missing at random, the mechanism is nonignorable and serious bias may occur. In that case, non-standard imputation techniques are called for that take into account the variables that are related to the missingness-mechanism.

In the remainder of this article, I will concentrate on item nonresponse, its causes, and ways to prevent it.

3. Sources of Item Nonresponse

3.1. Empirical findings

There are four potential sources of item-nonresponse: the method of data collection (mode), the questionnaire, the respondent, and the interviewer. Of course, interaction between sources can occur, for instance questionnaire and respondent, data collection mode and interviewer (cf. Groves, 1989).

Correlates of item nonresponse vary over studies. However, research into correlates of item nonresponse has shown several consistent patterns. Important characteristics of the questionnaire that influence the item nonresponse, are questionnaire lay-out (Jenkins & Dillman, 1997), and the inclusion of 'do-not-know'-options (Sudman & Bradburn, 1974).

The method of data collection does have a clear influence on item nonresponse. An extensive meta-analysis (De Leeuw, 1992) showed that interview surveys (both face-to-face and telephone) generally result in less item nonresponse than mail surveys. The exception is when sensitive questions are asked; in that case self-administered questionnaires (e.g., mail survey) perform better. Furthermore, both computer assisted interviewing and computer assisted self-administered questionnaires result in less item nonresponse than paper and pencil surveys (De Leeuw, Hox & Snijkers, 1998).

Respondent's age and education correlate consistently with item nonresponse. Elderly respondents and less educated respondents tend to have a higher number of missing data (Colsher & Wallace, 1989; Dillman, 1978; Herzog & Rodgers, 1992; Huisman, 1999; Sudman & Bradburn, 1974).

Finally, as the influence of data collection mode on missing data mentioned above shows, the role of the interviewer is important. However, training and supervision of the interviewer are crucial. Only well-trained interviewers have a low item-nonresponse; also interviewers who are monitored strictly, produce less missing data than interviewers who are not (cf. Fowler, 1991).

3.2. The question-answer process

To understand why missing data occur, it is important to completely understand what happens during the question-answer process (Tourangeau, 1984; Strack & Martin, 1987; Schwarz, 1997). First, the respondents have to *understand* the question; they have to determine the intended meaning.

In the next step they have to *recall* relevant information from memory: a sometimes difficult cognitive task. For an opinion question, they may either retrieve a previously formed opinion from memory or they may 'compute' an opinion on the spot. For knowledge questions and test items, the relevant knowledge has to be retrieved and combined. For behavioral questions, the respondent has to recall or reconstruct relevant instances of the behaviour from memory, and determine whether the behaviour occurred during the reference period (e.g., last three months). If the question asks for 'usual' behaviour, the respondent has to decide whether the recalled behaviour is representative or whether it reflects a deviation from the usual behaviour. So, the next step after retrieval from memory is 'computing' a *judgement*.

After a private judgement is formed in the mind of the respondent, she/he has to communicate the answer to the researcher. In case of a 'closed' question, the respondent has to *format* the judgement to fit the response categories. When 'open-ended' questions are used, the judgement has to be verbalized into a preliminary answer.

Before the answer is finally communicated, the respondent may want to *'edit'* the response, due to social desirability and situational adequacy. Especially with sensitive topics, or face-to-face interviews, this may be the case.

The question-answer process starts with the respondent's exposure to a question and ends with an overt answer of the respondent. In between, five important tasks have to be performed and during each task something may go wrong, creating a 'missing'. During the designing and planning phase of a study, a researcher can take many steps to ensure that the question-answer process proceeds as smoothly as possible. Thereby avoiding errors and improving data quality. A well designed questionnaire is the first step in preventing errors. This reduces not only respondent errors, but also interviewer errors (Fowler, 1991). Furthermore, an appropriate data collection mode should be chosen, respondents should be instructed well, and interviewers should be trained. But even if the question-answer process is followed through without any problem, errors may occur in noting down responses, and coding and editing the answer by interviewers or researchers during data processing and analysis. Quality control during coding and analysis is important too. The key to the reduction of item nonresponse, is Total Quality Management (cf. Hox, 1998; Dippo, 1997).

4. Sources and Prevention

4.1. The mode of data collection

There are three major methods of data collection (modes) in surveys: the self-administered questionnaire, the face-to-face interview, and the telephone interview. Each mode has advantages and disadvantages regarding data quality and nonresponse (For a meta-analysis, see De Leeuw, 1992). Often methods are mixed within one study. For example, a panel survey where the first wave is conducted face-to-face, and subsequent waves either by phone or through a mail (self-administered) questionnaire. Computer assisted variants have been successfully developed for each data collection mode.

In general, self-administered questionnaires are experienced as more confidential, and respondents give more honest answers to sensitive questions. Furthermore, in the presence of an interviewer respondents may be reluctant to answer these questions at all. When the topic is sensitive or can be threatening to the respondent, the use of self-administered questionnaires reduces the number of missing data (De Leeuw, 1992). But, respondents can make mistakes using a self-administered questionnaire, and skip a question. Therefore, the lay-out of questionnaires is extremely important. For lay-out rules and examples, see Jenkins & Dillman (1997).

A method to completely avoid mistakes, is the use of computer-assisted questionnaires (Nicholls et al., 1997; Saris, 1998). In a well-tested computer-assisted interview, all intended questions are asked and routing mistakes are avoided (cf. Van Hattum & De Leeuw, 1999). In addition, range and consistency checks during the (self-administered) interview take the place of post-editing. This makes it possible to rephrase and pose a question again, and there is no need to edit an inconsistency into a missing value. Again it should be emphasized that a 'do-not-know' response should not appear explicitly on the screen, but a special escape key should accept a do-not-know when necessary. That computer-assisted self-interviewing can be successfully implemented for special groups, such as children, has been demonstrated by

De Leeuw et al (1997).

4.2. The questionnaire

A well-designed questionnaire helps to avoid mistakes of both interviewers and respondents. The importance of questionnaire design for data quality in self-administered questionnaires has been recognized for a long time (cf. Dillman, 1978). The same principles govern data quality in interviewer-administered questionnaires. Even good interviewers make mistakes and routing errors will occur when instructions are not clear. Routing and branching errors should be avoided by a transparent lay-out, which uses principles of visual perception and graphical design (cf. Jenkins & Dillman, 1997) and guides the interviewer or the respondent error-free from question to question. The same principles can be used to make the structure of a test as transparent as possible.

Furthermore, the question-answer process should be completed successfully. The question and the question wording should be easily understood, and the response categories should fit and be exhaustive. In other words, the question should be simple and understandable. Besides question wording, the number of response categories (the item format) is important. In general, a larger number is better than just two response categories. People often feel uncomfortable with only two forced choices; their intended answer does not fit this limited forced choice and they escape in a do-not-know, or cannot answer. Four to seven categories are optimal (Krosnick & Fabrigar, 1997; Leigh & Martin, 1987), but all response categories should be meaningful to the respondent! For instance, a neutral mid-category is only meaningful when a bi-polar response scale is used (e.g., an agree-disagree response scale). Therefore, a neutral mid-category should only be used in a bi-polar response scale. It should be noted that mode of data collection influences the number of optimal response categories too; in a telephone survey five is a workable maximum (cf. De Leeuw, 1992).

There has been much debate about whether or not to include a 'do-not-know' response category (cf. Schuman & Presser, 1981). From a practical point of view one should never explicitly offer an easy way out through a 'do-not-know'-category or a 'no opinion filter'. There is no solid evidence that *explicitly* offering them is improving the data quality, and they do increase the item nonresponse (Krosnick & Fabrigar, 1997). However, a respondent should always be able to answer 'do-no-know' if necessary. For instance, although the interviewer does not read out aloud 'do-not-know' when reading the response possibilities, on the interview form or the computer a 'do-not-know'-option should be available when needed. Also, 'not applicable' should always be possible.

Of course, when 'do-not-know' is a meaningful answer, for instance in reaction to a question on voting intention, it should be offered explicitly. In this case 'do-not-know' has informational value; it does not represent 'missing data' (cf. section 2).

Short and clear instructions to the respondent should be embedded at appropriate places in the questionnaire (Dillman, 1978). This helps to avoid unnecessary missing data. For example, when asking how often something occurred, 'zero' can be a meaningful answer. In this case, the instructions should state clearly that one should explicitly fill in 'zero' and not skip the question. If this is not done, many unnecessary missing values are assigned during coding and editing (Skinner, 1999).

Even the best researcher does not write perfect questions. It is therefore essential that questions should be tested. Two forms of testing are recommended: the pretest and the pilot- or field test. The pretest is an intensive small scale test in which all steps of the question-answer process are checked. Focus groups or individual intensive cognitive interviews are used to discover problems and to learn how respondents interpret the question. This is usually done, after the semi-final version of the questionnaire is ready. A small number of potential respondents is invited to the office and intensive in-depth interviews are conducted, during which the researcher checks if all steps in the question-answer process are completed successfully. Usually, after the pretest, the questionnaire is adapted to facilitate the understanding of key-terms. For an introduction into cognitive pretests, see Forsyth & Lessler (1991) and Campanelli (1997). When preparing for large-scale surveys, a pilot or field-test usually is planned before the main data collection. A pilot is a small scale realistic test in the field of the *total* survey, and includes sampling, approaching respondents, data collection, coding and editing. I will discuss this further in section 5.

4.3. The respondent

Answering questions is a difficult task and respondents can fail to provide adequate responses. Respondents can skip questions by mistake, they may refuse to answer, or they may not be able to provide a correct answer. This may be caused by a problem in the question-answer process (e.g., not understanding the question, not able to retrieve needed information), by lack of motivation of the respondent, by the topic of the question (e.g., sensitive issues), or by badly designed questionnaires.

There are several ways, in which a researcher, can attempt to minimize respondent failures. First and most important of all, special attention should be paid to questionnaire, question writing and question testing. This has been extensively discussed above, and I will summarize those aspects that are of special importance to the respondent's role in the occurrence of missing data. When testing questionnaires, special attention should be paid to question comprehension and the inclusion of all relevant response categories. A well tested questionnaire is the basis of good data quality. Secondly, mistakes should be avoided as far as possible. An ergonomic lay-out of the questionnaire and instructions embedded in the questionnaire will help the respondent in providing the right answer in the right place (cf. Jenkins & Dillman, 1997). A well-written introduction letter, and interesting questions can keep a respondent motivated (Dillman, 1978), just as a well trained and attentive interviewer can stimulate a respondent to go carefully through the question-answer process.

Thirdly, computer assisted self interviewing methods can successfully help to avoid respondent mistakes or refusals. The interview program takes over the interviewer role and handles the questionnaire logic and questionnaire flow, making it easy for the respondent to answer. While, the respondent remains the locus of control and determines the pacing of the interview. This gives the respondent more time to understand the question and retrieve and compose an answer, which will improve the quality of the answer (Schwarz et al, 1991). Computer-assisted interviewing also leads to more self-disclosure when sensitive questions are asked (for an overview see Weisband & Kiesler, 1996).

When respondents are not very willing to part with sensitive information, there are several methods to stimulate a respondent to give a valid answer. These methods are not exclusive and can and should be used in combination. The underlying idea in all these measures is raising respondents' trust. This can be done by making the method as confidential as possible, for instance by combining an interview with a self-administered questionnaire that the respondent seals in an envelope, by using computer-assisted techniques, or by using special techniques like randomized response. In addition, one should always give reassurance and explain briefly how information will be handled and what the reason is for asking the questions. However, one should take care not to 'overdo' this and give lengthy reassurances; this can have the opposite effect (cf. Hippler et al., 1990). One should avoid appearances of censure and word questions in a non-judgmental style, and make response categories as broad as possible. For an overview of techniques, see De Leeuw (1999); Malow et al (1998).

When using retrospective questions, selective memory can play a role, and respondents in fact 'do-not-know' the answer. There are several ways in which the respondent can be stimulated to give an adequate response. For instance, one can encourage the respondent to use personal records (e.g., diaries, banking slips). One can also stimulate a more thorough question-answer sequence by using longer introduction or questions (cf. Scherpenzeel & Saris, 1996). The researcher should take care that the respondent understands the introduction and question; for instance, use several clear short sentences to build a longer introduction.

Special techniques to prod the respondent's memory and to improve recall, provide the respondent with memory cues, such as calendars on which special dates (e.g., birthdays, anniversaries, holidays) are noted down. This is also known as 'time-line follow back methodology.' Other techniques make use of the 'domain-dependent encoding' of memory. By using extra, introductory questions, the respondent is brought back to the situation in which the researcher is interested. The auxiliary questions are followed by the 'real' question. For example: the respondent is first asked about the last car trip: when was that, was it business or pleasure, where did you go to, etc. Finally, the central, key question is asked " did you use your car phone?".

4.4. The interviewer

Interviewers can have a very positive role in reducing nonresponse. They can guide the respondent through the questionnaire, explain a question, and adequately 'probe' a respondent. That is, ask the question again, when a respondent is not quite sure of the answer. In standard interview training, interviewers are often instructed to probe once after an initial 'do-not-know' or when a respondent hesitates in choosing the best fitting response category.

But, interviewers may also induce nonresponse. There are several ways in which an interviewer may cause missing data. The interviewer can fail to ask the question, or probe a respondent. Interviewers can also fail to record the answer, or record the answer incorrectly or illegible. In the latter case, the answer will often be coded as missing during post-interview editing.

There are two causes for the interviewer failures mentioned above. First of all, the interviewer can make a genuine mistake. For instance, take a wrong routing by mistake, skip two pages at once, et cetera. Secondly, the interviewer 'fails' on purpose. For instance, the interviewer wants to end the interview quickly or does not want to go through too much trouble. By taking a wrong routing on purpose, he/she can avoid some long and tedious questioning. By, not probing and just noting down 'do-not-know' the interview will take less time.

There are several ways in which genuine mistakes can be reduced or even avoided. First of all, interviewers should be trained intensively in the correct procedures. In addition, they should receive specific instructions about the questionnaire at the beginning of each new survey (Billiet & Loosveldt, 1988; Carton, 1999; McCrossan, 1991). Secondly, an ergonomic lay-out of the questionnaire or interviewer schedule reduces skipping and routing errors and the use of computer-assisted interviewing may even avoid these completely (see section 4.2).

Purposeful 'mistakes' can be reduced by strict interviewer supervision and control procedures (e.g., re-contacting a small sample of the interviewed), and by the use of computer log-files (Couper et al., 1997).

5. Conclusion and Discussion

With extra attention to each phase of the survey, it is possible to reduce the amount of item nonreponse. Well-trained interviewers, appropriate data collection methods, and a well designed questionnaire, all help to reduce item nonresponse. Crucial is an extensive pretest of the questionnaire to detect problems in question wording or presentation of the questions that lead to errors in the question-answer process.

When a large scale survey is planned, or when an existing survey is redesigned, field tests usually precede the implementation. In a field test or pilot all procedures necessary for a survey are followed through on a smaller scale. This gives researchers an opportunity for a last check on missing data. Statistical analysis of the collected data can provide information on patterns of missingness, which can be useful for a last redesign of the questionnaire. Similarly, interviewer debriefing studies, and follow-up interviews with respondents can provide the researcher with valuable information on problems during the data collection and suggest possible solutions. More elaborate and time consuming, but very useful procedures include interviewer-respondent interaction coding (for an introduction and application of the method see Dykema et al., 1997; Van der Zouwen et al., 1991).

Item nonresponse cannot be totally prevented, and during the initial data analysis imputation is still necessary. Knowing how item missing data occur may help to choose the most appropriate imputation method. For instance, respondents' age and education consistently correlate with item nonresponse. Therefore, the missing data are not missing completely at random. By using age and education in the adjustment model this is accounted for. Other lessons that can be learned from missing data patterns, concern the nonignorable aspect of missings on sensitive questions. In general, refusal to respond is not easy to impute for. The missingness mechanism is almost certainly *not* at random, and a more elaborate imputation technique is needed, that takes the missingness mechanism into account (cf. Huisman, 1999). As 'do-not-know'-answers are often a polite way to refuse, these also pose a serious imputation problem. Only in those cases where it is likely that a respondent really does not know the answer, and where the mechanism can be described, is relatively strait-forward imputation possible. For example, when an older respondent has difficulty remembering an event. Other cases of item nonresponse (e.g., missing by design, skipped by mistake, edited decision) may be *not* missing at random and nonignorable. However, either the mechanism is controlled by the researcher or knowledge about the mechanism is available and therefore can be adjusted for (Huisman, 1999). The knowledge about the missing mechanism can be included in the statistical analysis, and be taken into account.

One can successfully impute missing data, but it is not always easy or strait-forward. Not imputing, and simply using the 'default' option of the statistical program may sound as an easy solution, but one should realize that this technique is based on the very strong assumption of missing completely at random (cf. Hox, this issue). As discussed above, this is only rarely the case. Therefore, the best policy is first to prevent missing data as far as possible, and in the second stage use all available information to investigate the missing data patterns and adjust for missing data.

It is recommended to try and collect as many data as possible, and in addition to collect auxiliary information. By doing this researchers collect helpful information that can be used to optimize imputation procedures. For instance, on income questions the item nonresponse is often as high as 20-30%. A follow-up question asking for a categorical indication of income, will get answers from about half the nonresponders. This information can be used to impute an answer on the income question (Lavrakas, 1999). Interviewers are a valuable tool for collecting information that can be used in imputation. Researchers often forget, that interviewers can do far more than asking questions. They can observe with intelligence, and report on the question-answer process (cf. Snijkers et al. 1996). These observation can provide valuable material for the imputation of missing data!

References

Arbuckle, J. (1996). Full information estimation in the presence of missing data. In: G.A. Marcoulides & R.E. Schumacker (eds.). *Advanced structural equation modeling*. Mahwah, N.Y.: Erlbaum.

Billiet, J., & Loosveldt, G. (1988). Improvement of the quality of responses to factual survey questions by interviewer training. *Public Opinion Quarterly, 52*, 190-211.

Campanelli, P. (1997). Testing survey questions: New directions in cognitive interviewing. *BMS, 55 (special issue on the cognitive interview)*, 5-17.

Carton, A. (1999). *Selectie, training en evaluatie van interviewers binnen een interviewernetwerk* [In Dutch: An interviewer network: Contructing a procedure to evaluate interviewers]. Leuven: Garant.

Colsher, P.L. & Wallace, R.B. (1989). Data quality and age: Health and psychobehavioral correlates of item nonresponse and inconsistent responses. *Journal of Gerontology, 44*, 45-52.

Couper, M.P., Hansen, S.E., Sadovsky, S.A. (1997). Evaluating interviewer use of CAPI technology. In: L. Lyberg et al (eds). *Survey Measurement and Process Quality*. New York: Wiley.

Dillman, D.A. (1978). Mail and telephone surveys; The total design method. New York: Wiley (A totally revised and updated version is now in press).

Dippo, C.S. (1997). Survey measurement and process improvement: Concepts and integration. In: L. Lyberg et al (eds). *Survey Measurement and Process Quality*. New York: Wiley.

Dykema, J., Lepkowski, J.M., Blixt, S. (1997). The effect of interviewer and respondent behavior on data quality: An analysis of interaction coding in a validation study.In: L. Lyberg et al (eds). *Survey Measurement and Process Quality*. New York: Wiley.

Engel, U., & Reinecke, J. (1994). *Panelanalyse: Grundlagen, Techniken, Beispiele*. Berlin: Walter de Gruyter.

Fowler, F.J. jr. (1991). Reducing interviewer related error through interviewer training, supervision and other means. In: P. Biemer et al (eds). *Measurement errors in surveys*. New York: Wiley.

Freedman, D.S., Thornton, A., Camburn, D. (1980). Maintaining response rates in longitudinal studies. *Sociological Methods & Research, 9*, 87-98.

Forsyth, B.H., & Lessler, J.T. (1991). Cognitive laboratory methods: A taxonomy. In: P. Biemer et al (eds). *Measurement errors in surveys*. New York: Wiley.

Groves, R.M. (1989). *Survey errors and survey costs*. New York: Wiley.

Groves, R.M., & Couper, M.P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

Hermkens, P.L.J. (1983). *Oordelen over de rechtvaardigheid van inkomens [In Dutch: Judgements on the fairness of income*]. Amsterdam: Kobra.

Herzog, A.R., & Rodgers, W.L. (1992). The use of survey methods in research on older Americans. In: R.B. Wallace & R.F. Woolson (eds). *The epidemiological study of the elderly*. Oxford: Oxford University Press.

Hattum, M. van, & Leeuw, E.D. (1999). A Disk-by-Mail survey of teachers and pupils in Dutch primary schools: Logistics and data quality. *Journal of Official Statistics, 3 (in press)*.

Hippler, H-J., Schwarz, N, & Singer, E. (1990). Der influess von Datenschutzzusagen auf die teilnamebereitschaft an Umfragen [In German: The influence of dataprotection reassurance on the willingness to participate in a survey]. *ZUMA-nachrichten, 27*, 54-67.

Hox, J.J. (1998) Er is nieuws onder de zon: Nieuwe oplossingen voor oude problemen [In Dutch: New solutions to old problems]. *Kwantitatieve Methoden, 98*, 95-118.

Hox, J.J., Kreft, I.G.G., & Hermkens, P.L.J. (1991). The analysis of factorial surveys. *Sociological Methods & Research, 19,* 493-510.

Huisman, M.(1999). *Item nonresponse: occurrence, causes, and imputation of missing answers to test items*. Leiden: DSWO-press.

Jansen, M.G.H. (1997). The rasch model for speed tests and some extensions with applications to incomplete designs. *Journal of Educational and Behavioral Statistics, 22*, 125-140.

Jenkins, C.R., & Dillman, D.A. (1997). Towards a theory of self-administered questionnaire design. In: L. Lyberg et al (eds). *Survey Measurement and Process Quality*. New York: Wiley.

Kasprzyk, D., Duncan, G.J., Kalton, G., Singh, M.P. (1989). *Panel surveys*. New York:

Wiley.

Krosnick, J.A., & Fabrigar, L.R. (1997). Designing rating scales for effective measurement in surveys. In: L. Lyberg et al (eds). *Survey Measurement and Process Quality*. New York: Wiley.

Lavrakas, P. (1999). Personal communication to AAPOR-net, June 11.

Leigh, J.H., & Martin, C.R. (1987). Do-not-know item nonresponse in telephone surveys: Effects of question form and respondent characteristics. *Journal of Marketing Research, 24*, 418-424.

Lessler, J.T., & Kalsbeek, W.D. (1992). *Nonsampling error in surveys*. New York: Wiley.

Leeuw, E.D., (1999). Vragen naar gevoelige informatie: Een overzicht. [In Dutch: Asking sensitive information: A review of methods]. *Pedagogische Studien, 76*,92-103.

Leeuw, E.D. de (1992). *Data quality in mail, telephone, and face to face surveys.* Amsterdam: TT-publikaties.

Leeuw, E.D., de, & Hox, J.J. (1998) Nonresponse in surveys: een overzicht [In Dutch: Non response in surveys: An overview]. *Kwantitatieve Methoden, 57*, 31-53.

Leeuw, E.D. de, Hox, J.J., Snijkers, G. (1998). The effect of computer-assisted interviewing on data quality. In: B. Blyth (ed). *Market research and information technology. Application and innovation.* Esomar monograph 6 (pp.173-198). Amsterdam: Esomar.

Leeuw, E.D., Hox, J., Kef, S., & Hattum, M. van (1997). Overcoming the problems of special interviews on sensitive topics: Computer assisted self-interviewing tailored for young children and adolescents. *Sawtooth Software Conference Proceedings*. Sequim: WA: Sawtooth.

Little, R.J.A., & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Malow, R., Gustman, S., Ziskind, D., McMahon, R., St. Lawrence, J. (1998). Evaluating HIV prevention interventions among drug abusers: Validity issues. *Journal of HIV/AIDs Prevention & Education for Adolescents and Children, 2,* 21-40.

Martin, J. et al. (1996). *Task Force on Imputation, Report on Imputation*, Government Statistical services, Methodology series, # 3, U.K.: GSS/ ONS.

McCrossan, L. (1991). *A handbook for interviewers*. London: HMSO

Nicholls, W.L. II., Baker, R.P., & Martin, J. (1997). The effect of new data collection technologies on survey data quality. In: L. Lyberg et al (eds). *Survey Measurement and Process Quality*. New York: Wiley.

Pol, F.J.R. van de (1989). *Issues of design and analysis of panels*. Amsterdam: Sociometric

Research Foundation.

Saris, W.E. (1998). Ten years of interviewing without interviewers. In: M.P. Couper et al (eds.). *Computer Assisted information Collection*. New York: Wiley.

Skinner, C. (1999). Developing an imputation strategy, with illustrations from a self-completion survey of local authorities. Lecture presented at the Survey Methods Centre seminar on item-nonresponse in surveys, London, Royal Statistical Society, March 1999.

Strack, F., & Martin, L. (1987). Thinking judging and communicating: A process account of context effects in attitude surveys. In: H.J. Hippler et al (eds). *Social information processing and survey methodology*. New York: Springer.

Sudman, S., & Bradburn, N.M. (1974). *Response effects in surveys*. Chicago: Aldine.

Schuman, H.H., & Presser, S. (1981). *Questions & answers in attitude surveys*. New York: Academic Press.

Scherpenzeel, A., & Saris, W. (1997). The validity and reliability of survey questions: A meta-analysis of MTMM studies. *Sociological Methods and Research, 25*, 341-383.

Schwarz, N. (1997). Questionnaire design: The rocky road from concepts to answers. In: L. Lyberg et al (eds). *Survey Measurement and Process Quality*. New York: Wiley.

Snijkers, G., Akkerboom, H., Kuijpers, I, Leeuw, E.de. (1996). Computer-assisted qualitative interviewing: An intermediate technology of quality assessment. Paper presented at INTERCASIC '96, San Antonio, Texas, 11-12 dec. 1996.

Tourangeau, R (1984). Cognitive science and survey methods: A cognitive perspective. In T. Jabine et al (eds). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington DC: National Academy Press.

Vermunt, J. (1996). Causal log-linear modeling with latent variables and missing data. In: U. Engel & J. Reinecke (eds). *Analysis of change. Advanced techniques in panel data analysis*. New York: De Gruyter.

Weisband, S., & Kiesler, S. (1996). Self disclosure on computer forms: Meta-analysis and implications. Tucson: University of Arizona.
(http://www.al.arizona.edu~weisband/chi/chi96.html)

Zouwen, J. van der, Dijkstra, W., & Smith, J. (1991). Studying respondent-interviewer interaction; The relationship between interviewer style, interviewer behavior and response behavior. In: P. Biemer et al (eds). *Measurement errors in surveys*. New York: Wiley.

Ontvangen: 23-03-1999
Geaccepteerd: 12-07-1999