Mode of questionnaire administration can have serious effects on data quality

Ann Bowling

Abstract

Background One of the main primary data collection instruments in social, health and epidemiological research is the survey questionnaire. Modes of data collection by questionnaire differ in several ways, including the method of contacting respondents, the medium of delivering the questionnaire to respondents, and the administration of the questions. These are likely to have different effects on the quality of the data collected.

Methods This paper is based on a narrative review of systematic and non-systematic searches of the literature on the effects of mode of questionnaire administration on data quality.

Results Within different modes of questionnaire administration, there were many documented potential, biasing influences on the responses obtained. These were greatest between different types of mode (e.g. self-administered versus interview modes), rather than within modes. It can be difficult to separate out the effects of the different influences, at different levels.

Conclusions The biasing effects of mode of questionnaire administration has important implications for research methodology, the validity of the results of research, and for the soundness of public policy developed from evidence using questionnaire-based research. All users of questionnaires need to be aware of these potential effects on their data.

Keywords: mode of questionnaire administration, data collection bias

Introduction

One of the main primary data collection instruments in social, health and epidemiological research is the survey questionnaire. Modes of data collection by questionnaire vary in the method of contacting respondents, in the vehicle of delivering of the questionnaire, and in the way in which questions are administered. These variations can have different effects on the accuracy and quality of the data obtained. There are few publications on the topic, which are based on systematic, interdisciplinary searches of the literature. It is timely to attempt a focused search and report on the literature pertinent to population health, patients and health care.

Modes of collecting questionnaire data

Surveys can be conducted in different settings, and different questionnaire methods involve either paper and pencil, electronic (computer mouse/keyboard) or telephone key pad vehicles for collecting the data. These modes differ in several ways at different levels (see Box 1).

Thus, within any mode of administration, there are many potential influences on responses. These differences, at different levels, can make it difficult to separate out the effects of each on the quality of the data obtained. Even minor changes in question wording, question order or response format can result in differences in the type of response obtained, but can be difficult to separate out from other effects of different modes of administration.

In addition to the traditional range of paper and pencil methods, there is increasing academic interest in tools commonly used in market and public opinion research, e.g. the use of computer assisted face-to-face interviewing, computer-assisted telephone interviewing, self-administered computer methods, audio computer-assisted self-administered interviewing, and interactive voice response telephone methods (see Box 2). The range of electronic methods has been described by Tourangeau et al.

Burden on respondents

There are at least four steps involved in answering questionnaires, which make cognitive demands on respondents: comprehension of the question, recall of requested information from memory, evaluation of the link between the retrieved information and the question, and communication of the response. It is likely, then, that the channel of questionnaire presentation (e.g. auditory, oral, visual) affects the cognitive burden placed on respondents, especially the demand for literacy in the case of visual self-administration methods. And, as each mode inevitably imposes different cognitive requirements on respondents, and varies in the amount of privacy and anonymity they afford respondents, these can affect the process of responding to questions, and thus on the quality of the data.

Probably the least burdensome method is the personal, face-to-face interview (auditory channel) as this only requires the respondent to speak the same language in which the questions are asked, and to have basic verbal and listening skills. No
reading skills are required (unless written materials for the respondent are contained within the interview). A friendly, motivating interviewer can increase response and item response rates, maintain motivation with longer questionnaires, probe for responses, clarify ambiguous questions, help respondents with enlarged show cards of response choice options, use memory jogging techniques for aiding recall of events and behaviour, and control the order of the questions. Interviewers can also be trained to follow complex question routing and skipping instructions.

In contrast, telephone interviews make greater auditory demands and may be burdensome to respondents. Telephone interviews require basic verbal and language skills, and also require access to, or ownership of, a telephone. The most burdensome modes are likely to be visual and written methods of self-administration, as these demand that respondents are literate in reading the language/s of the survey, that they do not have visual impairments and have the dexterity (e.g. of wrist, fingers) to complete the questions. They require respondents to tick a box on a paper questionnaire, press an electronic key, or key on a touchtone telephone handset to indicate their response: respondents are required to read or listen, recognize numbers and write/key answers accurately. Respondents also need the ability to follow routing instructions.

Electronic methods require access to a computer and/or internet facilities (whether via an interviewer with a lap top personal computer (PC), or facilities in an office, clinic or home setting), basic computer literacy, and also familiarity with numbers and keyboards. They have literacy requirements in relation to reading the questions and replying, and can also have auditory requirements (ACASI). However, electronic programmes can be designed to require a limited range of keys, they have...
been documented in individual experiments and in reviews to have more complete item response rates than the various paper and pencil methods.\textsuperscript{5}

**Methods**

**Methods of systematic electronic literature search**

This paper is based on a narrative review of the literature on the effects of mode of questionnaire administration on data quality. The uniqueness of this review is that the search of the literature was conducted in a systematic fashion, including databases representing different disciplines (medicine, psychology and sociology), and supplemented with a wide non-systematic search of the grey literature. The databases searched were Medline, PsycINFO and Social Science Citation Index (via the Web of knowledge). All years were searched. The key terms were limited to ‘mode of administration’ and ‘data collection bias’ in order to focus the search. Both generic and health-specific literature was included (e.g. generic methodology papers and papers on the effects of mode of administering health status questionnaires, such as the Short Form-36).\textsuperscript{1} Abstracts of non-English references were included where translations were available online. Citations that included references were scanned and new references included where they met the review criteria. Studies were not included or excluded on grounds of methodological quality, as is the norm in systematic reviews of clinical trials. Although experimental techniques (e.g. randomization of respondents between modes of administration) are ideal methods, these are not frequently employed in methodological research. Full papers were obtained for each English language abstract included. Excluded papers related to standard comparisons of the validity of different questionnaire lengths (e.g. short versus long forms of health status questionnaires), and different quality of life measures; test–retest reliability checks, inter–interviewer bias, different question wording, or irrelevant papers (e.g. mode of drug administration).

**Methods of grey, non-electronic literature search**

In view of the difficulty of identifying relevant articles by standard methods of systematic reviews, and the cross-disciplinary nature of the topic, a search of the grey and non-electronic literature was conducted by: contacting published authors in the field; making personal contact with main government, independent and university social survey organizations in the UK, Europe and the USA; and checking the cited references in the literature elicited by the systematic review. Textbooks known to the investigator, with relevant sections on mode of administration, were also included. This exercise found that most of the key words used in electronic databases were inadequate to permit easy identification of specific papers on mode of administration effects. One key methodological experiment, comparing electronic and paper and pen self-completion methods, which were embedded in interview situations, used the key words ‘epidemiology, knowledge, attitude, practice studies, sexual behaviour, surveillance’, and none indicated mode of administration biases.\textsuperscript{6}

**Results**

The number of unduplicated publications identified and included in the review are displayed in Table 1. This yielded 73 out of 382 (19 per cent) potential papers that were relevant for review.

Much of the broader research on the effect of mode of questionnaire administration on data quality is from the USA, and published in internal reports of large survey organizations, in specialized books and social science journals. And most of this research consisted of comparisons of separate samples, rather than controlled, experimental designs. The most notable systematically conducted reviews include those by Tourangeau \textit{et al.},\textsuperscript{4} and by De Leeuw and van der Zouwen,\textsuperscript{5} based on meta-analyses of data quality in telephone and face-to-face surveys.

**Effects of mode of administration on data quality**

Data quality is a vague concept, and there is no agreed gold standard. It could be defined in terms of survey response rates, questionnaire item response rates, the accuracy of responses, absence of bias, and completeness of the information obtained from respondents. De Leeuw and van der Zouwen\textsuperscript{5} listed five main indicators for data quality, in addition to a sixth dependent variable (response rate – the number of completed interviews divided by the total number of eligible sample units) (see Box 3).

More broadly, these sources of error in surveys can be summarized as (i) non-measurement errors: survey design, sampling frame and sampling, non-response and item non-response; and (ii) measurement errors: survey instrument and data collection processes. Mode of questionnaire administration has effects on elements of both of these sources. The potential effects derived from the literature cited in this paper are summarized in Table 2, although this should be interpreted

### Table 1 Sources searched by number of elicited references

<table>
<thead>
<tr>
<th>Database*</th>
<th>Total papers found</th>
<th>Total papers included after removal of duplicates and exclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline*</td>
<td>333</td>
<td>30</td>
</tr>
<tr>
<td>PsycINFO and Social Science Citation Index†</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Other sources/citations†</td>
<td>46</td>
<td>41</td>
</tr>
<tr>
<td>Total</td>
<td>382</td>
<td>73</td>
</tr>
</tbody>
</table>

*Medline searched via National Library of Medicine, Social Science Citation Index searched via ISI Web of Knowledge, Social Science, PsycINFO searched via American Psychological Association.
†PsycINFO and Social Science totals relate to new references not already listed by Medline. Other source totals relate to new references not already listed by Medline or PsycINFO.
with caution, given that the literature is not always consistent, rarely based on experimental designs, and often varying by topic. A summary of the reviewed literature is presented next under the headings of non-measurement and measurement error.

### Non-measurement error

**Coverage for sampling**

All methods require up-to-date sampling or address lists prior to sampling, to ensure completeness of coverage of the target population, and each carries its own form of bias. Interview and postal surveys (outside market research) rely on complete and up-to-date lists of addresses (e.g. from post or zip code files); these may be incomplete and out of date, thus leading to sample bias. Telephone surveys generally use random digit dialling because of the number of people who choose not to be listed in telephone directories. These are also subject to sample selection bias because inclusion in the sample is limited to those with telephones and who answer them directly (i.e. rather than using answer phones and call screening facilities). Electronic surveys are limited to those with access to a personal computer, email and internet access, creating immediate sample bias, accentuated by the lack of complete lists of private email addresses. These standard issues are dealt with in most methodology textbooks.

### Response rates

Methodological research comparing different methods of administering questionnaires has focused on the issue of response rates, item response, and on methods of increasing these, in particular in relation to postal surveys. The main reasons for non-response include respondents’ unwillingness to participate in the study, the investigator’s inability to contact respondents (e.g. people who are out during home or telephone interview surveys) and communication barriers (e.g. literacy barriers, sensory impairments). Non-response is thus likely to be influenced by mode of questionnaire administration (e.g. people who have difficulty writing are unlikely to respond to a postal survey and hence differ in this respect to respondents).

The lower the response rate to a study, the greater the danger that the respondents may differ from non-respondents in their characteristics, which affects the precision (reliability) of the survey’s population estimates, resulting in study bias, and weakening the external validity (generalizability) of the survey results.

---

**Box 3. De Leeuw and van der Zouwen’s five main indicators for data quality.**

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Validity</th>
<th>Social acceptability</th>
<th>Item response</th>
<th>Amount of information</th>
<th>Similarity of response distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy or validity of response</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Absence of social desirability bias</td>
<td>Low</td>
<td>Great</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Item response</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Question order effects</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Response-choice order effects</td>
<td>Moderate</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Recall bias</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Social desirability bias</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>'Yes-saying' bias</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Interviewer bias</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Length of verbal response/amount of information</td>
<td>High</td>
<td>Low</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Willingness to disclose sensitive information</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Respondents’ preferences for mode of administration</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Potential for</th>
<th>Face-to-face interviews</th>
<th>Telephone interviews</th>
<th>Self-administered, postal</th>
<th>Self-administered, programmed, electronic</th>
</tr>
</thead>
<tbody>
<tr>
<td>More complete population</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>coverage for sampling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive burden</td>
<td>Low</td>
<td>Great</td>
<td>Great</td>
<td>Great</td>
</tr>
<tr>
<td>Survey response</td>
<td>High</td>
<td>Low</td>
<td>Medium – low</td>
<td>Low</td>
</tr>
<tr>
<td>Item response/completion</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>of questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question order effects</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Response-choice order effects</td>
<td>Moderate</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Recall bias</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Social desirability bias</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>'Yes-saying' bias</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Interviewer bias</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Length of verbal response/</td>
<td>High</td>
<td>Low</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>amount of information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willingness to disclose</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>sensitive information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respondents’ preferences</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Moderate</td>
</tr>
<tr>
<td>for mode of administration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Even if the quality of the data obtained is good, a biased sample is of little value in making population estimates that represent the target population. While there is much literature reporting on differences between respondents and non-respondents in relation to individual studies, overall reviews and systematic reviews of health-related literature on the differences between responders and non-responders are inconsistent or inconclusive.11,12

Face-to-face interview surveys have long been assumed to achieve higher response rates than postal and other types of surveys. A friendly interviewer on the doorstep can be motivating, and it may be easier to convince respondents of the legitimacy of the study in person, which should increase response rates. But narrative literature reviews, up until the 1990s, indicated that postal questionnaires, with at least two reminders and sponsorship by an official or respected body, could achieve response rates equal to interviews (e.g. 85 per cent) on appropriate topics.11,13,14 But response rates across all methods have declined over the past decade (this is evident when comparing the response rates over time to the British General Household Surveys).15 A number of studies have since reported differences between response rates to telephone, face-to-face and postal questionnaires.16 Sykes and Collins17 compared three large British population surveys conducted during the 1980s: two studies of social attitudes, and a study of lifestyle, which all directly compared telephone and face-to-face interviewing. The telephone method achieved substantially lower response rates than face-to-face methods, and yielded more childless couples than couples with children. Research has also shown higher response rates for self-administered postal questionnaires compared with self-administration questionnaires handed out to people (e.g. in hospitals) to complete and return.18

Different modes of administration by different sequencing can also produce different response rates. In a survey of hospital inpatient experiences by Harris et al.,19 patients were randomized to receive a postal questionnaire with telephone interview follow-up of non-respondents, or a telephone interview with postal follow-up of non-responders. The authors reported that the telephone first method had a higher response rate, and with higher item response. At an individual study level, the response rates to different modes of questionnaire administration are likely to vary by topic, and in particular for complex issues. There is less information on electronic methods, which also suffer from an inability to cover target populations adequately (see earlier).

Item response rates

Higher item non-response is generally reported in postal surveys compared with face-to-face interviews,20,21 and in postal questionnaires in comparison with telephone interviews.19 Consistent with this, De Leeuw and van der Zouwen’s8 meta-analysis found higher item response in face-to-face interviewers, which they explained by interviewers motivating people to respond, and greater control of the interviewer over the situation (e.g. ensuring that questions are answered and not missed, and recording responses correctly). And while electronic and automated telephone programmes can prevent respondents moving to the next question before they have completed the previous question [by keying in a recognized field (valid number)], they cannot prevent premature termination.

All methods can be prone to premature termination (the respondent does not wish to continue with the interview and terminates it), but this is probably less likely in the presence of a motivating interviewer. Telephone interview surveys, electronic and automated voice activated telephone surveys can be particularly prone to this22 as people become bored with lengthy schedules and discontinue them. However, electronic methods have been reported to have superior item response rates to traditional pen and paper methods. Johnson et al.,6 in a methodological experiment within a British survey on sexual behaviour, reported that while there were no differences in response between computer-assisted self-completion questionnaires (CASI) and pen and paper self-completion questionnaires (PAPI) (both administered within face-to-face interview situations), CASI led to fewer missing item responses than PAPI. Tourangeau et al.’s5 review supports this conclusion that electronic methods had more complete item response rates than the various paper and pencil methods.

Measurement error

The influence of the social setting and bias

While standardized surveys aim to include well-designed and tested questions that have the same meaning to all participants, cultural, social and language differences can all influence interpretations. The actual data collection process also involves an interaction between the questionnaire, the respondent and, in the case of face-to-face and telephone interviews, the interviewer. The nature of this interaction inevitably varies between interview settings and self-administration situations, as well as by individual interviewer. Differences in data by mode of questionnaire administration can therefore be hypothesized, simply by given differences in the structure of the social setting. In theory, these differences can also be compounded by differences in the setting the questionnaire is administered in, although not all investigations find differences in response by setting.23 Tourangeau et al.4 included the pace of the interview, as well as the control over the order of questions,1,2,9,24,25 as potential mediators to the effects of different modes of data collection on data quality (see Box 4).

Social desirability bias

Interviews, therefore, involve social interaction with another person, which can lead to respondents taking social norms into account when responding, resulting in social desirability bias (the desire of respondents to present themselves in the best possible light), resulting in the over-reporting of desirable behaviours, and under-reporting of undesirable behaviours (confounding associations between variables by attenuating, inflating or moderating relationships). Methods exist to reduce this problem, including
assurances of confidentiality and anonymity, although this can raise respondents’ suspicions about the sensitivity of the topic, and thereby reduce response.\textsuperscript{26} checking responses against known ‘facts’; indirect questioning; correlation of responses with social desirability measures; and randomized response techniques. With the latter technique, respondents are presented with pairs of questions, one of which is sensitive and one of which is not and they are asked to answer one question within the pair at random (e.g. by tossing a coin); the interviewer cannot see the outcome. In order to infer the response to the socially desirable question, the population distribution of responses to the non-socially desirable question needs to be known from other sources, a truly random procedure needs to be used in the selection of the question, and a large population must be sampled.\textsuperscript{27}

Respondents have been shown to give more positive and socially desirable responses in interview (face-to-face and telephone) surveys than in self-administration (e.g. postal) surveys,\textsuperscript{7,26,29} even when attempts have been made to take order and contextual effects into account.\textsuperscript{1,30} Thus estimates of positive health status, health related quality of life, engaging in desirable behaviours and activities, appear likely to be exaggerated when based on face-to-face or telephone interviews, rather than self-administration methods, and socially undesirable behaviours (e.g. smoking) are likely to be under-estimated.\textsuperscript{1,30–36} Sensitive health problems can also be under-reported in face-to-face or telephone interviews, compared with self-administered questionnaires (e.g. prostatic disease, urinary symptoms).\textsuperscript{37,38} But some research has reported no differences between interviewer versus self-completion modes and type of response,\textsuperscript{39–44} including a small number of studies in which respondents were randomized between various interview and self-administered questionnaire modes.\textsuperscript{44–47} Williams et al.\textsuperscript{48} in a survey of HIV risk behaviour questions used a cross-over design with random assignment of respondents to audio-CASI or face–to–face interview, and to different sequences of these at follow-up. They reported no effect of mode of administration on type of response.

Overall, the literature indicates that there appear to be fewer differences in type of response between the same administration modes, compared to between different modes,\textsuperscript{8,17,48,49} which has implications for the common research practice of using mixed mode designs (e.g. conducting face-to-face interviews at baseline, with postal or telephone interview at follow-up). Other non-randomized studies have reported no differences in type of responses between telephone and face-to-face interview surveys on voting and alcohol intake.\textsuperscript{21,50} Furthermore, De Leeuw and van der Zouwen’s\textsuperscript{8} meta-analysis found fairly small differences between telephone and face-to-face interviews in social desirability bias, amount of information provided, and similarity of response. But, inconsistent with these results, Evans et al.\textsuperscript{51} alternated the order of face-to-face and telephone interviews among consecutive general practice attenders, and reported that people aged over 60 were more likely to score as anxious or depressed in telephone than face-to-face interviews.

In relation to self-administration modes, Fouladi et al.\textsuperscript{52} compared responses to self-administered paper questionnaires and self-administered electronic (‘on-line’) questionnaires on emotional functioning, and only small differences in response patterns were detected. Similarly, no differences have been reported in response type between different self-administration modes (email, postal, handing out questionnaires to patients) in surveys of patient satisfaction and assessments of health services.\textsuperscript{18,53} Again, research is not all consistent,\textsuperscript{54} which is not unexpected given the predominance of non-experimental designs. Differences in type of responses could be attributable to selection and response bias, or even to real differences between samples.

**Acquiescence bias**

Any excess of positive responses in interview, compared to self-administration situations, could also be due to increased ‘yes-saying’ or acquiescence bias: a culturally based tendency to agree with others because it is perceived to be ‘easier’ to agree than disagree. Although ‘yes-saying’ can also be evident on self-administered questionnaires, it appears to be less pronounced than in interviews. The potential for a different type of reporting bias due to ‘ease’ exists in self-administered questionnaires, based on evidence from classic psychological experiments. These respondents tend to check the nearest response choices to the question. Commonly used attempts to control for this include switching the order of responses periodically in a measurement scale (e.g. from ‘Strongly agree – Strongly disagree’ to ‘Strongly disagree – Strongly agree’).\textsuperscript{9}

---

**Box 4. Pace of interview and control over order of question effects.**

1. Computer and self-administration methods possibly slow the pace of the interview down, giving respondents more time to think, thus yielding more accurate responses.
2. Interviews have the advantage that, while administering the questions, interviewers can control the order of the questions (face-to-face or telephone).
3. Electronic and automated telephone methods are usually programmed to prevent the respondent jumping ahead and previewing questions by not moving on to the next question until the previous questions have been answered, although they sometimes permit respondents to go back to check and correct answers.
4. Self-administered paper questionnaires have no interviewer/programme control over the order of the questions. Respondents can preview the questions and adjust their answers as a result.
While classic experiments have supported evidence of a cultural tendency towards ‘yes-saying’, research has yielded inconsistent results in relation to differences between mode of questionnaire administration and ‘yes-saying’. Nicholas et al.,21 again in their experiment of telephone and face-to-face interview methods, found little evidence of acquiescence bias operating more with one mode of administration than another. A meta-analysis on this issue by De Leeuw55 failed to detect any differences between postal, face-to-face and telephone interviews, although De Leeuw and van der Zouwen’s8 earlier meta-analysis found some evidence of more acquiescence (agreement) bias, more evasiveness (‘don’t know’ replies or no reply), and more extreme responses in telephone interviews compared with face-to-face interviews.

Interviewer bias

The presence of an interviewer can be distracting to respondents. If an excess of positive or socially desirable responses in interview situations is found, this could be due to interviewer bias (due to characteristics of the interviewer,9,24 or because people may be reluctant to reveal beliefs unlikely to be endorsed by the interviewer; see earlier, social desirability bias). In addition, interviewers can vary in their ability to appear or sound neutral, to listen, to probe adequately, to use techniques to aid recall and to record responses. Careful training and monitoring of interviewers can minimize this, and analysis of responses by interviewer (where more than one is used) can check for interviewer bias. Self-administration modes obviously avoid this source of bias.

Response-choice order: primacy and recency effects

Bias from the order of the response choices, by mode of administration, have also been reported. Ideally respondents listen to the question during interviews, keep all response options in mind, consider them and decide which is most important/applicable to themselves. This is a demanding task, especially when under time pressures, and in telephone interview situations where there can be no visual prompting (question-answer sequences rarely exceed 1 minute).56

Research has indicated that when questions are presented visually (as in self-administered questionnaires) respondents are likely to begin with the first response option presented (primacy effects). This is because in situations where respondents can consider response choices at their own pace, they may chose an early response alternative without much thought if it is agreeable, and then move on to the next question. In contrast, when questions are presented orally (as in face-to-face or telephone interviews) respondents tend to begin processing the final response option offered (while they still recall it) and, where agreeable, they select that option (recency effects). This tendency leads to response order effects.28,57–59 There is also some evidence that the selection of more extreme response choices (i.e. ‘very’ as opposed to ‘fairly’ satisfied), is more common in telephone than face-to-face interviews.21,59

Although the evidence on primacy and recency effects is mixed,16 a meta-analysis of response order experiments in large scale surveys showed that recency effects were most pronounced among people aged 65 and over,56,61 possibly due to cognitive changes that can occur with normal ageing.62 Other research has also reported that older and younger people are affected differently by features of the research instrument.63

Recall effects

Respondents need to recall the information in order to respond to a question, and this again can vary by mode of administration. The self-administered paper questionnaire is visual and the interview setting is aural, immediately resulting in different cognitive processes operating. In interviews, an interviewer can re-direct respondents back to the topic of relevance if they stray off it, probe to elicit relevant information, and utilize a range of techniques to prompt memory,9 but in self-administration settings this is not possible, although respondents themselves can consult diaries or other information sources to aid recall. In self-administration settings the respondent alone has to judge whether the information they have recalled is relevant to the question, and how best to respond. However, research is inconsistent for respondent recall by mode of administration mode64,65 partly reflecting bias from non-experimental designs.65

Length of verbal response

Telephone interviews tend to be shorter, the mean number of communication acts is shorter, and the length of utterances is shorter than with face-to-face interviews.8 Telephone methods have also been shown to yield more truncated responses, or no responses at all, to open ended questions than face-to-face interviews.17

De Leeuw and van der Zouwen’s8 meta-analysis confirmed that the amount of information given by respondents to open questions and checklists was greater in face-to-face than telephone interviews. This might be due to the need for feedback and pressure to maintain verbal flow in telephone conversations, resulting in respondents answering questions more quickly and interviewers allowing less time between questions. More sentences, of greater length, may also be more acceptable in the face-to-face interview situation.8

Sensitive information

Self-administration of questionnaires can increase respondents’ willingness to disclose sensitive information, compared with face-to-face or telephone interviews. The greater anonymity offered in postal survey, with its weak social presence, for example, has been reported to lead to high item response, and more accurate reporting on sensitive topics such as health and behaviour.4,66–68 There are, of course, established techniques for eliciting sensitive information and for checking for bias, which overlap with the methods of minimizing social desirability bias (see earlier). It is preferable to minimize the potential for such bias at the outset. Sensitive questions are best asked by more impersonal, self-administration methods.
as they lead to higher levels of reporting. The highest levels of reporting are for audio-computer self-administration questionnaires and computer-assisted self-completion interviews (which allow respondents to key their responses to questions directly into an interviewer’s laptop computer, permitting more confidentiality).

The literature is inconclusive about differences within modes (e.g. interview modes, such as telephone and face-to-face interview methods) and the validity of sensitive information. Some give evidence that telephone methods are more successful at eliciting frank responses about sensitive behaviour than face-to-face interviews. Others have reported that respondents are more willing to answer questions about race and income or illicit drug use in personal face-to-face rather than in conventional telephone interviews. Electronic self-completion questionnaires to paper self-completion questionnaires. Electronic self-completion modes, within face-to-face interview situations (i.e. on the interviewer’s lap top computer), have also been reported to be acceptable to people aged 65 and over. (And see Social desirability bias earlier).

**Respondent preferences**

While the biasing effects of mode of preference are unknown, studies examining respondents’ preferences report that people prefer face-to-face interviews to telephone interviews, and electronic self-completion questionnaires to paper self-completion questionnaires. Electronic self-completion modes, within face-to-face interview situations (i.e. on the interviewer’s lap top computer), have also been reported to be acceptable to people aged 65 and over. (And see Social desirability bias earlier).

**Discussion**

This paper was based on a narrative review of systematic and non-systematic searches of the literature on the effects of mode of questionnaire administration on data quality. The review showed that, while some studies were inconsistent or inconclusive, different modes of questionnaire administration are likely to affect the quality of the data collected. The effects appeared to be more marked between interview and self-administration modes, rather than within modes. It was often difficult to isolate the effects of the method of contact from the other differences between the data collection methods, and this limits knowledge about how the mode of administration alters the process of answering questions. A main problem with the literature elicited is that most of the studies did not use experimental or randomization methods to allocate the different questionnaire modes to participants. Thus, differences detected in responses between different modes could be due to differences between settings, or to genuine differences between respondents.

Explanatory models that have been proposed for the effects of data collection mode on data quality include the impersonality of the method of contacting respondents and in delivering and administering the questionnaire (highest in self-administration methods), the cognitive burden imposed on respondents by the method (greatest in self-administration methods), the legitimacy of the study (it is more difficult to establish the credentials of some surveys in telephone contacts), the control over the questionnaire (interviewers have the highest level of control over the order and completion of the questions), the rapport between respondent and interviewer (lowest in self-administration settings where there is no visual contact), and communication style (an interviewer can be motivating and clarify questions, but can lead to interviewer and social desirability bias) (see Box 5). These models need to be fully tested in experimental designs.

---

**Box 5. Explanations for effects of data collection mode on data quality.**

1. The impersonality of the method: while an interviewer can enhance motivation to respond as well as response accuracy, self-administration methods increase perceived impersonality and may encourage reporting of some sensitive information (e.g. in interview situations there may be fear of embarrassment with the exposure of weakness, failure or deviancy in the presence of a stranger).

2. The cognitive burden imposed by the method: different methods make different demands on respondents, including reading, listening, following instructions, recognising numbers and keying in responses. Face-to-face interviews make the least demands, while the lack of visual support in telephone interviews may make the task more complex.

3. The legitimacy of the study: this may be more difficult to establish with some methods than others. In contrast to paper or electronic communications, telephone contacts limit the possibilities for establishing the survey’s credentials. This might affect initial response and the importance respondents place on the study, and their motivation to answer questions accurately.

4. The control over the questionnaire varies: interviewers have the highest level of control over question order; in self-administered paper questionnaire modes there is little control over question order.

5. Rapport: rapport between respondent and interviewer may be more difficult to establish in self-administration and telephone interview than in face-to-face modes, as there is no visual contact. This can adversely affect motivation to respond, although social desirability bias may be reduced as there is less need for approval.

6. Communication style: more information may be obtained in interview than other situations, as interviewers can motivate respondents, pause to encourage (more, longer) responses, and clarify questions; interviewers can also lead to interviewer and social desirability bias.
This topic has important implications for research methodology, the validity of the results of research, the soundness of evidence-based public policy, and for clinicians who wish to screen their patients using questionnaires. All users of questionnaires need to be aware of the potential effects of mode of administration on their data. The validity of the common research practice of comparing data from dual modes of administration within studies is also called into question. While calls have been made for greater attention to questionnaire development in epidemiology, there has been less focus on the wide range of different biases, at different levels, stemming from the various modes of administering questionnaires.

References


25 Schwarz N, Hippler HJ. Subsequent questions may influence answers to preceding questions in mail surveys. Pub Opin Quart 1995; 59: 93–97.


McHorney CA, Kosinski M, Ware JE. Comparisons of the costs and quality of norms for the SF-36 Health Survey collected by mail versus telephone interview: results from a national survey. *Med Care* 1994; 32: 551–567.


de Leeuw ED. *Data quality in mail, telephone and face-to-face surveys*. Amsterdam: TT Publications, 1992.


77 Olsen J, on behalf of the IEA European Questionnaire Group. Epidemiology deserves better questionnaires. *Int J Epidemiol* 1998; **27**: 935.