6

The design of questionnaires

> Words are the building blocks for all question structures, but deciding which words to be used and in what order is far from simple. The wrong choice of words can create any number of problems, from excessive vagueness to too much precision, from being misunderstood to not being understood at all, and from being too objectionable to being uninteresting and irrelevant. *(Dillman 2000, p. 50)*

Introduction

For our purposes, a *questionnaire* can be defined as a tool designed to elicit and record, or guide the elicitation and recording of, recalled exposures from subjects of an epidemiological study. It contains questions to be put to the subject, and may also include answers to those questions from which the subject must choose those that are appropriate to him/her.

The objectives of questionnaire design are:

- to obtain measurements of exposure variables essential to the objectives of the study
- to minimize error in these measurements
- to create an instrument that is easy for the interviewer and subject to use, and for the investigator to process and analyse.

These objectives are potentially in conflict, and any questionnaire usually represents a compromise among them. For example, it may be necessary to trade off some ease in processing and analysis against ease in completion by the interviewer or the subject. Similarly, the addition of some questions essential to the objectives of the study, for example questions about sexual behaviour in a study of the aetiology of cancer of the cervix, may make a questionnaire more difficult for an interviewer to administer and more threatening to the respondent. Judgement must be exercised in making decisions about the content and

Principles of Exposure Measurement in Epidemiology:Collecting, Evaluating, and Improving Measures of Disease Risk Factors. Second Edition. Emily White, Bruce K. Armstrong and Rodolfo Saracci, Oxford University Press (2008).
DOI: 10.1093/acprof:oso/9780198509851.003.0006

structure of questionnaires. Where compromise is necessary, the designer should favour decisions that maximize the usefulness of the questionnaire to the objectives of the study and minimize error in measurement.

In this chapter, we cover the major topics of importance in the design of questionnaires:

- choice of the items of data to be covered by the questionnaire
- the types of question that can be used
- the material covered by each question
- the wording of questions
- question order
- physical format and structure of questionnaires
- the problem of collecting information on behaviours that vary with time
- aids to recall
- pre-testing questionnaires.

More comprehensive accounts of the design of questionnaires can be found in Bennett and Ritchie (1975), Sudman and Bradburn (1983), Schwarz and Sudman (1996), Aday (1996), Dillman (2000), and Tourangeau *et al.* (2000). Decisions about choice of mode of administration of the questionnaire (i.e. telephone interview, in-person interview, self-administered questionnaire, or various types of computer-assisted interviews) are discussed in Chapter 2. Chapter 7 covers issues specific to administering questionnaires by in-person and telephone interviewing.

Choice of items to be covered

Questionnaire design usually begins with selection of the items of data that must be translated into questions. The ground to be covered is determined by two main factors: the objectives of the study, and the limitations imposed by the burden that can be placed on respondents, including the feasible length of the questionnaire.

Objectives of the study

'The content of a questionnaire is generally designed to investigate the minimum amount of an individual's total experience that will provide sufficient information concerning the problem under study' (Bennett and Ritchie 1975). Just as the objectives of the study determine the variables to be measured as a whole, they also determine the specific items to be covered in the questionnaire. If a question does not contribute to the achievement of the objectives, it has no place in the questionnaire.

Adequately detailed data should be sought for each essential exposure variable, as described in Chapter 1. For example, for exposure that happens frequently, it is usual to ask about the time exposure began, the time it ended, and the frequency and intensity of the exposure and their variation over time. A comprehensive list of potential confounders and effect modifiers should also be developed. A well thought out plan for data analysis, and a description of the algorithms that will be used to create exposure dose variables and covariate variables, is essential in determining the items and detail required. Developing the exposure and covariate algorithms that will be used at the end of the study *before* questionnaire development at the beginning of a study is important to avoid a surprising common problem—that at the time of data analysis, the researcher realizes that an item needed to compute an exposure dose variable had not been collected!

Length of the questionnaire

The topics to be covered in a questionnaire and the detail in which they are covered are limited first and foremost by the length of time that subjects are willing to spend on the questioning process. While there are inevitable exceptions, it may be taken as a general rule that the maximum time that can be spent administering a questionnaire is 1–2 hours by face-to-face interview and 30–40 minutes by telephone interview. Self-administered questionnaires are at an added disadvantage in that the subject can gain an impression of the size of the response task before deciding whether to embark on it. Response rates are reduced with longer mailed questionnaires (Edwards *et al*. 2002).

Other aspects of respondent burden

'Respondent burden concerns the level of demand placed on the respondent necessary to answer the survey instrument questions' (Sudman and Andersen 1977). Length of the questionnaire is one aspect of respondent burden. Additional contributors to respondent burden are:

- length and distance (from the present) of the period of time over which recall is requested
- salience (or impact) to the subject of the topic of questioning, including its sensitivity
- frequency of the event
- complexity or detail of the data sought.

In general, recall over a long period of time or from the distant past, topics of low salience or impact, questioning regarding frequent events (such as eating), and complex questions (e.g. a full occupational history with details of exposure

to hazards in each occupation) will all add to respondent burden. Being a proxy respondent for someone else is generally a further burden.

Increased burden on the respondent has several consequences:

- risk increases for termination of the interview or non-completion of a self-administered questionnaire
- quality of data obtained is reduced
- response rate is threatened
- the population may become alienated from survey research and cooperation in future studies may be reduced (Sudman and Andersen 1977).

The last effect is a particular problem for longitudinal studies requiring recurrent surveys in one population.

There is often a conflict between collecting information considered to be necessary to the objectives of the study, keeping the questionnaire to an acceptable length, and minimizing other aspects of respondent burden. In resolving this conflict, it is important not only to collect the amount of information necessary to the objectives of the study, but also to ensure that questionnaire length and respondent burden are kept to levels that do not threaten subject participation or cause a material increase in measurement error.

Types of question

Questions are generally classified as either 'open-ended' or 'closed-ended'. *Open-ended questions* are questions to which no answers are provided by the investigator. Only the question is asked, and the respondent's answer is recorded verbatim. In an interview, extensive probing may be used to ensure that all relevant aspects of the topic are covered by the answer. *Closed-ended questions* are questions for which the range of possible answers is specified by the investigator and the respondent is asked to make a choice from among the answers provided.

Open-ended questions

Open-ended questions should be used in epidemiology for eliciting and recording *simple* facts to which there are a large number of possible answers, for example, age, occupation, country of birth, number of cigarettes smoked a day, amount of alcohol drunk in a particular period of time, etc. The use of closed-ended questions for these topics leads to loss of information and, when asking about a socially undesirable behaviour, a greater degree of error. For example,

reporting of intake of beer was nearly 50 per cent less when a closed-ended rather than an open-ended question was used (Blair *et al.* 1977).

One advantage of open-ended questions is that subjects cannot be influenced by the response options, as they may be for closed-ended questions. Schwarz *et al.* (1985) found that when offered a range of categories from 'up to half an hour' to 'more than 2½ hours' for daily TV watching, 16.2 per cent of subjects estimated that they watched more than 2½ hours. When offered the range 'up to 2½ hours' to 'more than 4½ hours', 37.5 per cent estimated that they watched more than 2½ hours. It appears that the response categories offered are seen as normative by some respondents, and their responses are influenced away from the extremes, particularly if one or other extreme is viewed as socially undesirable.

The collection of data on income (used in epidemiology as a measure of socio-economic status) may be an exception to the rule that simple factual data are best sought through open-ended questions. While there is probably no particular tendency for subjects to over- or understate their income, they are sensitive about disclosing the exact amount. Thus a closed-ended question with income categories might be acceptable, where an open-ended question would not. It is also usual in self-administered questionnaires to ask questions with a limited range of categorical responses (such as 'What is your sex?' or 'What is your marital status?') as closed-ended questions to permit maximum use of self-coding responses.

When the likely answer to an open-ended question is neither simple nor factual, the use of such a question increases the burden on both respondent and interviewer and produces answers that are difficult both to code and to analyse.

Closed-ended questions

The use of closed-ended questions is comparatively uncommon in epidemiology for exposure measurement, and appropriately so for the reasons given above. Most data sought are both simple and factual, and are best dealt with by open-ended questions. However, closed-ended questions are needed more often on self-administered questionnaires than on interviews, for at least two reasons. First, self-administered questionnaires need to be simpler to complete than interviews. For example, an interview may ask an open-ended question on frequency of a behaviour, for which the interviewer records both the frequency and the period: for example, for a response of 'three times per week', the interviewer records the number '3' and the time period 'week'. This would be too complex for respondents on a self-administered questionnaire, and so a list of responses ranging from, say, 'more than twice a day' to 'never or less

than once a week' would be needed. Secondly, self-administered questionnaires are often used for large studies which use automated data capture systems (e.g. optical scanning), many of which can only read closed-ended responses (e.g. marks in boxes).

The alternative answers offered in a closed-ended question should be simple and brief, and mutually exclusive if only one is to be selected. If more than one response could be selected, it may be best to seek explicit 'yes/no' responses for each of the categories. If the response categories provided are not exhaustive of all possible responses, a final open category (e.g. 'Other. Please give details _____') should be given.

As noted above, respondents are reluctant to place themselves in the extreme category of an undesirable behaviour. When categories are needed, for example for alcohol intake, it may be appropriate to add an extra extreme category (e.g. 6+ a day) to encourage response to the penultimate category.

One decision in selecting responses to closed-ended questions is whether to include 'don't know' as an option. A 'don't know' response should be offered only if the possibility exists that some respondents would truly not know the answer. For example, in asking about family history, a 'don't know' response is needed for those who are adopted and do not know their biological family. However, generally a 'don't know' response category should not be given, because a greater percentage of subjects will answer the question if an explicit 'don't know' is not given. For example, a study by Poe *et al.* (1988) has shown that, in approaching the next of kin of dead subjects for data on demographic variables and aspects of health history and health-related behaviour, exclusion of the 'don't know' option gave an appreciably higher proportion of usable responses for many items without adversely affecting response rate or intramethod reliability.

The response list should be in a logical sequence, because subjects will often stop reading the answers after they find the appropriate one. For the same reason, specific definitions should be given before general ones. For example, for a question on type of spread used at the table, the response 'Low-fat spread or diet margarine' should be before 'Regular margarine'. Then subjects would read the 'low-fat' response before deciding whether the 'regular margarine' response applied to them.

It is also important in closed-ended questions to limit the number of response categories as far as is reasonably possible. A large number of alternative responses increases respondent burden, increases the probability of non-response to the question (Leigh and Martin 1987), and may increase the probability that one of the response options listed first will be selected (Krosnick and Alwin 1987).

Question content

Questions may be about:

- *knowledge* (what people know)
- *attitudes* (what people say they want or think)
- *beliefs* (what people say is true)
- *experiences* (what has happened to people)
- *behaviours* (what people do, have done, or will do)
- *attributes* (what people are).

Generally, only experiences, behaviours, and attributes are relevant to exposure measurement in epidemiology. However, most of the research on question content and questionnaires has related to questions about knowledge, attitudes, and beliefs. This research will be drawn on and, as far as possible, described in terms of its relevance to exposure measurement through questionnaires.

Like the subject matter as a whole, the content of individual questions is largely determined by the objectives of the study. Therefore it is not possible to be prescriptive on this subject, but some general advice can be given.

Before developing questions on a particular topic, the investigator should become thoroughly familiar with it by reading and by discussions with experts. It is advisable to obtain copies of questionnaires that have been used previously by experts to cover the subject matter of interest and to make prudent use of them, within the limits of any copyright restrictions that may apply and provided that due acknowledgement is given.

The use of standard questions has a number of advantages.

- The questions will usually have been used extensively and proved satisfactory in use.
- The questions may have been assessed for reliability and/or validity and, even if they have not, sufficient results may be available from their use to permit validity to be inferred.
- Their use will permit comparison among datasets and possibly the combining of datasets.
- It is an easy way of drawing on the expertise of others, and it can substantially facilitate the task of questionnaire design.

All that said, it is important for the investigator to evaluate questions obtained from other sources in terms of the adequacy of their design, their appropriateness to the objectives of the current study, and their suitability for use in the population on which the study will be conducted. It should also be noted that questions developed for use in a face-to-face interview may require

modification in their wording or format if they are to be used in a telephone interview or a mailed self-administered questionnaire.

One source of questionnaires is a website developed by the US National Cancer Institute (http://dceg.cancer.gov/QMOD/qmod_categories.htm). This includes questionnaires for a range of exposures that have been used in studies of cancer.

Question wording

There are two important issues to be considered in question wording.

- How does one arrive at a suitable wording in the first place?
- Are small changes in wording likely to lead to differences in response?

The latter issue is particularly relevant to comparisons between populations and over time when the same basic data have been sought by slightly different questions.

Table 6.1 gives a list of questions that should be asked about the wording of each question in a questionnaire.

The words

The words used in a questionnaire should be the usual 'working tools' of the respondents. They should be neither too difficult (be suspicious of words

Table 6.1 Questions that should be asked about the wording of each question in a questionnaire[a]

- Will the words be uniformly understood by the subject population?
- Does the question contain abbreviations, unconventional phrases, or jargon?
- Is the question vague?
- Is the question too precise?
- Is the question biased?
- Is the topic sensitive?
- Does the question contain a double negative?
- Are the answer options mutually exclusive (or if not, is '*Answer all that apply*' stated)?
- Are the answer options exhaustive, i.e. is there an appropriate response for all respondents?
- Is an unambiguous time reference provided?
- Does the question contain more than one concept?
- Does the question require a calculation?
- Is the concept too complex?

[a] Adapted from Dillman (2000).

more than seven letters in length) nor too simple. Difficult words may not be understood, and simple words (where better but more difficult words could have been used) may appear condescending, may not convey the right meaning, and may needlessly lengthen the questionnaire. Where doubt exists, however, there is a virtue in simplicity. Abbreviations, unconventional phrases, and jargon present the same problems as difficult words; they may not be understood or, perhaps worse, they may be misunderstood.

These general principles may be illustrated by some examples. The question

HAVE YOU EVER HAD AN ECG?

includes an abbreviation that is also technical jargon. In addition, the abbreviation used in some English-speaking countries is EKG, not ECG. Nonetheless, the abbreviation is likely to be familiar to many subjects, and perhaps more familiar than any alternative terms. The solution here is to offer some alternative terms in the question, for example

HAVE YOU EVER HAD AN ECG, THAT IS, A 'HEART TRACING', EKG, OR ELECTROCARDIOGRAPH?

In the particular circumstance of quantifying sensitive behaviour (e.g. sexual activity or the use of alcohol or illicit drugs), asking the respondent to supply words that he/she is accustomed to using to describe the behaviour may increase the accuracy of reporting (Blair *et al.* 1977). Two forms of question were asked about drunkenness:

IN THE PAST YEAR, HOW OFTEN DID YOU BECOME INTOXICATED WHILE DRINKING ANY KIND OF ALCOHOLIC BEVERAGE?

and

SOMETIMES PEOPLE DRINK A LITTLE TOO MUCH BEER, WINE, OR WHISKY SO THAT THEY ACT DIFFERENTLY FROM USUAL. WHAT WORD DO YOU THINK WE SHOULD USE TO DESCRIBE PEOPLE WHEN THEY GET THAT WAY, SO THAT YOU WILL KNOW WHAT WE MEAN AND FEEL COMFORT ABLE TALKING ABOUT IT?
IN THE PAST YEAR, HOW OFTEN DID YOU BECOME (*RESPONDENT'S WORD*) WHILE DRINKING ANY KIND OF ALCOHOLIC BEVERAGE?

The second form of the question consistently increased the reporting of this and other sensitive behaviours. Because socially undesirable behaviours tend to be under-reported, an increase in reporting is assumed to mean reduction in error.

There are no data to suggest that particular terms familiar to the subjects should be used for other than sensitive behaviours and, in general, slang or

vulgar terms should not be used in questionnaires. Their use may lead to misunderstanding, the appearance of condescension, and, for some subjects, offence.

Vague questions

Questions may contain vague words—words that vary substantially in their meaning among different people. 'Usually', 'normally', and 'regularly' are three commonly used vague descriptors of frequency. In many circumstances they can be replaced by more precise quantifiers. For example,

HOW OLD WERE YOU WHEN YOU FIRST BEGAN TO SMOKE CIGARETTES REGULARLY?

could be made more precise by asking

HOW OLD WERE YOU WHEN YOU FIRST SMOKED ONE OR MORE CIGARETTES A DAY FOR ONE MONTH OR LONGER?

The latter wording eliminates uncertainty about the meaning of 'regularly'.

Similarly, it would be better to ask subjects about their 'usual' intake of alcoholic beverages over a specific period of time (e.g. the past 12 months) than simply to ask about their 'usual' intake.

Questions that are too precise

While precision is desirable, particularly when estimating amount or duration of exposure, respondent burden may be increased unduly if too much precision is requested. For example, in precisely quantifying dose rate and cumulative exposure to cigarette smoke, it might be tempting to ask smokers to estimate their average daily cigarette consumption for each year of their smoking life. This would be unreasonably burdensome and prone to substantial error in recall. A better approach would be to ask subjects about major changes in daily cigarette intake (e.g. an increase or decrease of 10 or more cigarettes a day), and document the time of each of these changes.

Biased questions

Biased questions are questions that suggest to the respondent that a particular answer is preferred from among all possible answers. 'Leading' questions are well known, and should be easily avoided. 'Loaded' questions are more likely to turn up in questionnaires seeking attitudes or beliefs than more factual data. The question

DO YOU THINK THAT SMOKING SHOULD BE BANNED IN PLANES?

would be more likely to bias responses because of use of the strong negative word 'banned' than would the more neutral, and balanced,

DO YOU THINK THAT SMOKING SHOULD BE PERMITTED OR NOT PERMITTED IN PLANES?

A significant influence of question wording has been shown in surveys of attitudes to public assistance to the poor in the USA (Smith 1987). In six surveys spanning 17 years, the proportion of Americans in favour of more spending on welfare was substantially less when the word 'welfare' was used in the question than when the question referred to 'assistance to the poor'. Smith (1987) concluded that this difference was due to a connotation of 'welfare' with waste and bureaucracy that 'assistance to the poor' did not have, or at least not to the same degree.

Sensitive questions

Sensitive or threatening questions are questions that '. . . ask respondents about behaviours that are illegal, contra-normative [deviant] or generally not discussed in public without tension, or relate to issues of self-preservation' (Blair *et al*. 1977). They fall into two distinct classes: those that ask about behaviours or attributes that are socially desirable, and those that ask about socially undesirable behaviours or attributes. The threat of questions about socially desirable behaviours arises from the possibility that a person does not wish to admit that he/she does not practise the behaviour (e.g. giving to charity or, of more direct relevance to epidemiology, exercising regularly). These behaviours tend to be over-reported. Socially undesirable behaviours or attributes (such as past history of sexually transmitted disease and alcohol drinking) tend to be under-reported. Questions about income, savings, and assets are threatening, although they are not readily categorized as either socially desirable or undesirable. A question can be considered as potentially threatening if subjects can possibly feel that there is a right or wrong answer to it (Sudman and Bradburn 1983).

Eliciting accurate answers to sensitive questions can be enhanced by selection of a more impersonal mode of administration (see Chapter 2), by interviewer training if a personal interview is to be used (see Chapter 7), and by the wording of questions (covered below).

A number of techniques have been used to maximize reporting of socially undesirable behaviours, two of which have already been mentioned: use of words familiar to the respondent, and open-ended questions. Reporting of undesirable behaviours was also increased by use of a long introduction to

the question (Blair *et al.* 1977). For example, the following introduction was added to the question on drunkenness:

OCCASIONALLY PEOPLE DRINK ON AN EMPTY STOMACH OR DRINK A LITTLE TOO MUCH AND BECOME (INTOXICATED *OR RESPONDENT'S WORD*). IN THE PAST YEAR, HOW OFTEN . . .

This introduction exemplifies another technique for dealing with question threat—deliberately loading the question. In this case the use of 'occasionally', 'drink on an empty stomach', and 'a little too much' all tend to minimize the significance of the behaviour so that the respondent will be more willing to report it. The threat may also be reduced by *embedding* the sensitive question in a list of questions on related topics, some of which are more threatening and some less so. Questions on past history of sexually transmitted disease, for example, can be asked in a series of questions on past disease history.

Reporting of socially undesirable behaviours increases if explicit assurances of confidentiality are given before the interview or before the sensitive question (Frey 1986; Singer 1985). In an in-person interview, socially undesirable questions can often be phrased in a straightforward way, with reliance on the accepting demeanour of the interviewer to yield accurate answers.

There is a difference between strategies appropriate to reducing over-reporting of socially desirable behaviour and those appropriate to increasing reporting of socially undesirable behaviour. Use of a long introductory statement, for example, is likely to encourage rather than discourage over-reporting of desirable behaviours, unless it loads the question against reporting the behaviour. A question on walking might be asked as follows:

MANY PEOPLE FIND IT DIFFICULT TO FIND TIME TO GET REGULAR EXERCISE, LIKE WALKING. IN THE PAST YEAR DID YOU WALK FOR EXERCISE AT LEAST ONCE A WEEK?

The significance of the behaviour may also be minimized by use of the phrase: 'Did you happen to . . .'

IN THE LAST MONTH, DID YOU EVER HAPPEN TO FORGET TO USE A CONDOM?

This might be best asked after the question 'Do you usually use a condom?' to allow a socially desirable response first, which might help to obtain more accurate answers to the question of actual interest above.

A final approach appropriate for questions about socially desirable behaviours, as with socially undesirable behaviours, is to reduce the threat by embedding them in a sequence of questions about non-threatening behaviours.

For example, intake of fat-free milk could be included in a list of questions on frequency of drinking types of milk.

Double negatives

A double negative can arise whenever a question that is phrased negatively can have a negative answer. Dillman (2000) gave the following example:

SHOULD THE CITY MANAGER NOT BE RESPONSIBLE TO THE MAYOR?

Yes ☐
No ☐

The question was asked because the city council was contemplating a change from the city manager being responsible to the mayor to him/her not being responsible to the mayor. It is unnatural to say 'yes' when the answer really means 'no' (that the city manager should *not* be responsible to the mayor), and so answers to this question would be ambiguous. The solution in this case was to ask to whom the city manager should be responsible—the mayor or the city council. Double negatives are probably more common in questions on attitudes and beliefs than in questions likely to be asked in epidemiological studies.

Mutually exclusive answers

A subject could reasonably select more than one answer to the following question:

WHAT SPREAD DO YOU USE ON BREAD?

BUTTER ☐
LOW-FAT SPREAD
OR DIET MARGARINE ☐
REGULAR MARGARINE ☐
OTHER SPREAD ☐
PLEASE GIVE DETAILS
_____

Therefore the subject is uncertain about which alternative to choose, and there is a risk of non-response. The problem could be solved by asking 'What spread do you *usually* eat on bread?', by instructing respondents to '*Mark all that apply*', or, preferably, by asking about the frequency of use of all the spreads individually.

Response for all participants

Response lists should also be exhaustive, i.e. all respondents should be able to find an appropriate response. Response categories are sometimes omitted because the researcher has made an assumption that would not apply to all

respondents. For example, the following question assumes that all respondents eats steak:

HOW DO YOU LIKE YOUR STEAK TO BE COOKED?

RARE ☐
MEDIUM ☐
WELL DONE ☐

If this assumption is wrong, at best the respondent will not answer the question (in a self-administered questionnaire), and at worst the respondent may find the assumption about his/her behaviour to be offensive. The solution is to add the response alternative '*I don't eat steak*', or to ascertain first whether or not the respondent eats steak and skip to a succeeding question if he/she does not.

An unambiguous time reference

Any period referred to in a question should be clear and unambiguous. The question described above on walking 'in the past year' might have been asked as follows:

DO YOU WALK FOR EXERCISE AT LEAST ONCE PER WEEK?

This question, while most likely referring to the present, has no clear time reference. Different subjects would be likely to refer to different periods of time in the past in answering it.

In a case–control study, it is usually desirable to specify a time for each case somewhat before onset of disease beyond which the exposure will not be recorded to avoid eliciting behaviour which has been influenced by the onset of disease, and a similar date for controls (see Chapter 1). For example:

IN 2002, DID YOU WALK FOR EXERCISE AT LEAST ONCE PER WEEK?

A specific *reference date*, usually the date of diagnosis of the case (or diagnosis date minus a year or two to account for a time when symptoms may have influenced behaviours), is assigned to each case and his/her matched controls. The use of this reference date is usually explained at the beginning of the interview or questionnaire, and questions relating to it usually begin: 'Before [reference date], did you . . .', etc.

One potential solution to problems of ambiguity in the time reference of a question is for the interviewer to have, or the questionnaire to include, a calendar on which the period referred to in that question, or a number of related questions, is marked (see the section below on aids to recall).

More than one concept

A question that has more than one concept should not be posed. For example:

THINK ABOUT YOUR DIET OVER THE PAST YEAR. HOW OFTEN DID YOU EAT A SERVING OF FRUITS OR VEGETABLES? *DO NOT INCLUDE JUICES, SALADS, POTATOES, OR BEANS*.

While fruit and vegetable intake may form a single exposure in a study, most subjects would consider fruits and vegetables as separate categories. Subjects would typically think through the answer by adding the number of times they eat vegetables to the number of times they eat fruit. This is an example of *decomposition*, which is a method used by subjects in formulating an answer to a question by breaking it down into more manageable parts (Bradburn *et al.* 1987). Thus the above question should be asked as two questions, one on fruits and one on vegetables. One can learn which questions subjects decompose and how they decompose them through 'think aloud' interviews, described below in the section on pre-testing questionnaires. The researcher can use this information to break down a complex question into simpler components.

Questions that require calculations

Certain questions which may appear to be a simple concept may actually require the subject to perform a calculation to derive an answer. For example:

IN THE PAST YEAR, ABOUT HOW MANY HOURS PER WEEK DID YOU WALK FOR EXERCISE?

To answer this question, most subjects would need to break it down into the number of times they walked each week and the minutes they walked per session, and then multiply these together (and then convert to hours!). In this case, it would be better to ask separate questions:

IN THE PAST YEAR, ABOUT HOW MANY TIMES PER WEEK DID YOU WALK FOR EXERCISE?

and

HOW MANY MINUTES DID YOU WALK EACH SESSION?

Complex concepts

Some exposures have complex definitions, which would be burdensome for the subject to comprehend. For example, to be classified as 'exposed' in a study, a subject might need to have both a sufficient dose and duration of exposure. For example, the researcher may want to define walking for exercise

as walking at least once a week for at least 20 minutes per session and at least at a moderate pace. Then the question might be phrased as:

OVER THE PAST YEAR, DID YOU WALK FOR EXERCISE AT LEAST ONCE A WEEK FOR 20 MINUTES OR MORE PER SESSION? *DO NOT INCLUDE CASUAL WALKING*.

If the question were phrased this way, it would certainly annoy some respondents with its complexity. Instead, part of the concept in the question can be incorporated into the answers to sub-questions. The above question could be phrased: '*Over the past year, did you walk for exercise at least once a week?*' Then, sub-questions would ask about sessions per week, minutes per session (with the lowest response category 'less than 20 minutes'), and pace (with the lowest category 'casual or strolling'). Then, when the algorithm for energy expenditure from walking is developed, those who walked for less than 20 minutes per session or who walked at a casual pace could be coded as not walking for exercise.

Question order

General principles

Beliefs about the order in which questions should appear in a questionnaire have been summarized by Sudman and Bradburn (1983). Questions about a particular topic should be grouped together, and proceed from the general to the particular within a group. This approach, by focusing first in a general way on a particular behaviour or experience, assists and allows more time for recall of the specific details. It is sometimes suggested that questions using a particular response scale should be grouped together, but this is probably undesirable because it may tend to promote a *response set*—a tendency to give the same response to each question regardless of what the correct response should be.

The order in which questions are asked can influence the responses obtained. The evidence for this statement derives mainly from research into questions on attitudes and beliefs (e.g. Helsing and Comstock 1975; McFarland 1981; Schuman *et al*. 1981), but the possibility exists that similar effects could be observed for some questions relating to experiences, behaviours, or attributes. Thus for comparative studies between populations and over time, the order of the questions in the questionnaire should be kept constant as far as possible. Any new questions that are added should go at the end of the questionnaire. Alternatively, any change in question order might be evaluated by comparison of the old and new questionnaires in a single population.

The first question

It is common practice to place the demographic questions at the beginning, but this is not a good idea. These questions are of comparatively low interest to the respondents, and some of them are threatening (see below). Instead, the questionnaire should begin with a question or questions which relate directly to the topic of the research and will command the subject's interest. For example, in a study of sun exposure in relation to skin cancer, it is appropriate to begin with questions on recreational pursuits involving sun exposure.

Sensitive questions

Sensitive questions should be placed towards the end of the questionnaire, in order of increasing threat. This reduces the likelihood that they will precipitate early termination of the interview, or failure to complete the questionnaire. The degree of threat presented by particular questions can be determined empirically by asking subjects 'how uneasy most people' would feel about particular topics of questioning in a questionnaire. In a questionnaire mainly about leisure and sporting activities, the degree of threat of particular topics, from least to greatest, was (Blair *et al.* 1977):

- general leisure and sporting activity
- alcohol drinking
- gambling with friends
- income
- petting and kissing
- drunkenness
- use of stimulants and depressants
- sexual intercourse
- use of marijuana
- masturbation.

Although sensitive questions should be placed near the end of a questionnaire, they should not appear on the last page of a self-administered questionnaire. Questions on the last page are highly visible if the subject peruses the questionnaire.

Demographic questions (e.g. about race, religion, or income) are sometimes threatening and usually not interesting, and therefore are appropriately asked at the end of the questionnaire. If, for some reason (e.g. to select particular respondents), it is necessary to place them at the beginning, some explanation for their position should be given to the respondent.

Logical sequence

Within any topic, questions should follow a logical sequence—the sequence that the respondents might be expected to follow in thinking about the topic. Thus, for example, in collecting residential or job histories, it is usual to proceed chronologically beginning with the present residence or occupation and proceeding to successively earlier ones. The 'backwards' chronological approach also gives the respondent more time to recall the events of the more distant past (Bradburn *et al.* 1987). Other topics, such as pregnancy outcomes, might be more easily recalled in chronological order.

Questionnaire structure

In addition to the questions, every questionnaire should contain:

- an introduction
- instructions (for self-administered questionnaires), including skip patterns
- linking phrases between topics
- a conclusion.

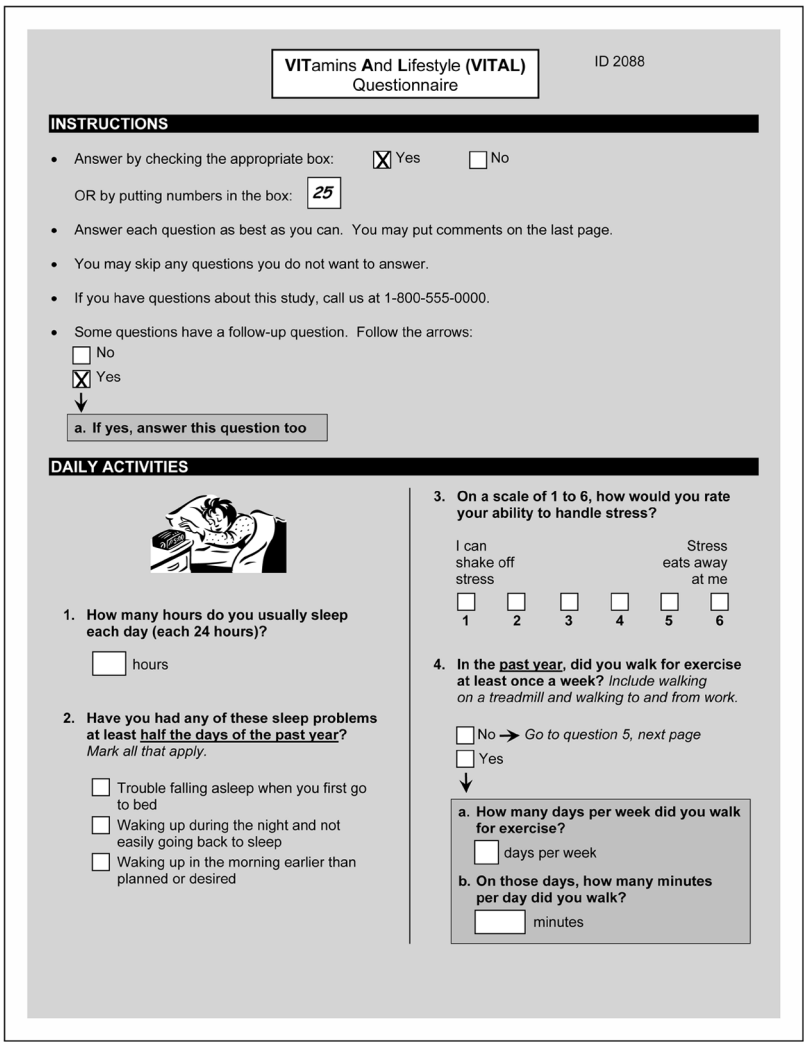
Introduction and instructions

In an interview, the introduction takes the form of a standard statement read by the interviewer. For a self-administered questionnaire it is usually part of the letter soliciting their cooperation. The introduction serves both to elicit participation and to discharge the investigator's ethical obligations to the subjects. The items to be included in the introduction are discussed in Chapter 11 in the section on response rates and their maximization.

General instructions will usually be only necessary in a self-administered questionnaire. Such instructions should be short and simple; an example is given in Figure 6.1. General instructions for interviewers will form part of interviewer training and an interviewer's manual rather than the questionnaire. Instructions relating to specific questions should appear with those questions in the body of the questionnaire, whether it is for self-administration or interview. Figure 6.1 provides examples of question-specific instructions given in italics after some questions. These include instructions to 'mark all that apply' when more then one response may be appropriate, skip instructions (discussed below), and instructions about the meaning of specific questions.

Linking statements

Linking statements break the subject's concentration on a particular topic, provide a brief pause, and establish his/her concentration on a new topic. They may also be used to break the monotony of a long series of questions on



VITamins And Lifestyle (VITAL) Questionnaire

ID 2088

INSTRUCTIONS

- Answer by checking the appropriate box: ☒ Yes ☐ No
  OR by putting numbers in the box: 25
- Answer each question as best as you can. You may put comments on the last page.
- You may skip any questions you do not want to answer.
- If you have questions about this study, call us at 1-800-555-0000.
- Some questions have a follow-up question. Follow the arrows:
  ☐ No
  ☒ Yes
  ↓
  a. If yes, answer this question too

DAILY ACTIVITIES

Figure 6.1 Example of instructions and format of a self-administered questionnaire.

one topic. They should not be unnecessarily long, appear to be demanding, or give unwarranted importance to the succeeding questions. Nor should they contain words or phrases that may bias the succeeding responses. The following are some examples.

- To signify a major change in questioning:

  THE FOOD WE EAT IS AN IMPORTANT PART OF OUR EVERYDAY LIVES. I WOULD NOW LIKE TO ASK SOME QUESTIONS ABOUT THE FOODS THAT YOU USUALLY EAT AND THE AMOUNTS OF THEM THAT YOU EAT.

- To break the monotony of a series of food frequency questions:

  NEXT, I WOULD LIKE TO ASK ABOUT BREAD AND BREAKFAST CEREALS.

- To introduce the demographic questions at the end of the questionnaire:

  FINALLY, I WOULD LIKE TO ASK A FEW QUESTIONS ABOUT YOU FOR STATISTICAL PURPOSES.

Skips or branches

It is commonly necessary to make 'skips' or branch points in a questionnaire where some succeeding questions are not applicable to all respondents. Computer-assisted interviewing has greatly reduced missing data due to failure to follow the proper skip patterns (Nicholls *et al.*1997). However, paper and pencil questionnaires must still rely on good instructions and navigational aids for the interviewer or subject to follow the skip patterns. This is especially true for self-administered questionnaires, where failure to follow the skip patterns correctly is a major source of missing data (De Leeuw 2001).

Probably the most important requirement of skip instructions is that they are placed immediately after the *answer* that leads to the branch point in the questionnaire. Skip instructions should always be worded positively ('Go to question 4') rather than negatively ('Skip question 3'). Figure 6.1 (question 4) provides an example of skip pattern instructions for a self-administered questionnaire, using arrows and instructions to show the skip pattern and using arrows and boxes to show the sub-questions. It is more important to make the path clear for those who are to complete the sub-questions, as was done in the example, than for those who should skip them. Complex branching designs are usually only possible in interviewer-administered questionnaires.

A device that may be used to avoid skips in a self-administered questionnaire is the inclusion of a specific 'inapplicable' category among the alternative answers, as in the following example:

HOW OFTEN DO YOU CUT THE FAT OFF MEAT BEFORE YOU COOK OR EAT IT?

NEVER	☐
LESS THAN HALF THE TIME	☐
MORE THAN HALF THE TIME	☐
ALWAYS	☐
I NEVER EAT MEAT	☐

Here, use of 'I never eat meat' as a response category provides an alternative answer for everyone and eliminates the need for a skip.

Conclusion

The conclusion of either a self-administered questionnaire or an interview should include an expression of appreciation and should provide an opportunity for the respondent to make comments. A self-administered questionnaire could also include a request for the respondent to check that all questions have been answered (or that all pages have been completed, for a long questionnaire). The conclusion should also contain the address for the return of a mailed questionnaire; while an addressed return envelope will usually be included, it may have become separated from the questionnaire. Interviewer-administered questionnaires should also provide for the entry of the interviewer's comments.

Questionnaire format

Formatting questions

The principles of formatting individual questions are summarized in Table 6.2. Examples of formatting interviewer-administered questions have been given in some of the questions shown in the preceding pages. Figure 6.1 gives examples of formatting self-administered questions.

Table 6.2 Principles of formatting individual questions (for self-administered questionnaires and interviews)

- Identify separate parts of questions with different typefaces

  For interviewer-administered questionnaires:

  CAPITAL LETTERS for the question
  Bold face for the alternative responses that are not to be read to respondent
  BOLD CAPITAL letters for the alternative responses to be read
  Italics for instructions not to be read
  CAPITAL ITALICS for instructions to be read

  For self-administered questionnaires:

  Bold face for questions
  Regular typeface for alternative responses
  Italics for instructions
- Include specific instructions and prompts (for interviewers) with the question as needed
- Record responses to closed-ended questions by check boxes
- Use vertical answer formats (except for scales)
- Provide spaces or boxes for coding open-ended questions
- Consider data capture methods when making design decisions, e.g. if key entry is to be used, precode closed-ended responses

The principles require little explanation. The different typefaces for questions, responses, and instructions help to lead the interviewer or respondent to the correct parts of the question. The use of capital letters (of the appropriate typeface) provides a consistent cue to the interviewer of what is to be read aloud. Putting any specific instructions or prompts (for interviewers) next to the question eliminates the need for complex initial instructions and serves as a reminder at the time the instruction is needed. The use of checkboxes for closed-ended questions is easy to understand, and a vertical format for the alternative responses makes it clear which box goes with which response. An exception to this is response scales or ratings (e.g. a scale for an attitude or for degree of pain). These categories may be better visualized on self-administered questionnaires if they are on a horizontal line (e.g. question 3 in Figure 6.1). Coding boxes or spaces are also needed for open-ended questions, including those that will be coded after the interview (e.g. occupational coding).

Finally, the method of data-capture needs to be considered when making formatting decisions. When the responses to a questionnaire are to be key-entered, the response options need to be precoded into numbers that are printed on the form to facilitate key-entry. For example:

Yes $\square_1$
No $\square_2$

It is also helpful to the key entry personnel if the responses to be entered are consistently placed on the right hand margin, rather than throughout each page of the questionnaire. Formats need to be tailored to the specific requirements of other data-capture methods, including computer-assisted interviews, web surveys, and self-administered questionnaires that will be processed by optical mark or character recognition. Dillman (2000) provides a detailed description of additional format considerations when designing web surveys and surveys to be processed by optical recognition.

Formatting self-administered questionnaires

Presentation of the questionnaire is particularly important if it is to be self-administered, both for ease of use and to give an authoritative appearance that will encourage response. Dillman (2002) proposed a Tailored Design Method, based on decades of experience and on numerous studies conducted by him and others. This is outlined in Table 6.3.

The questionnaire should be printed in booklet form so that will open flat on a table. The first page should have the title of the project and instructions for completing the questionnaire. A graphic illustration may make it more attractive to the target population. Use of coloured paper may increase

Table 6.3 Formatting self-administered questionnaires[a]

- Use a booklet format
- Put title of project and instructions on first page
- Include a graphic and/or colour for visual interest
- Consider coloured background with white response boxes
- Use a two-column format
- Consecutively number pages and questions
- Make skip patterns clear, e.g. through use of arrows after response and instructions

[a] Adapted from Dillman (2000).

response rates by a small degree (Edwards *et al.* 2002). Missing data may be reduced if subjects are asked to write their answers, either open-ended or check marks, in white boxes which stand out from a coloured background. A two-column format for the questions is easier to read because participants may skip words when reading longer lines of text. Two columns also allows more questions per page. Use of the vertical answer format will generally increase the amount of space between questions, and it is important that the questionnaire is not too congested. An example of many of these principles is given in Figure 6.1.

It is important for the respondent to be able to see the path of questions easily. Pages and questions should be numbered consecutively, and subsections of questions should be indented and identified with letters rather than numbers. Question numbers should extend to the left of questions to stand out. As far as possible, questions should not extend over more than one page. As discussed above, clear skip patterns are particularly important in self-administered questionnaires.

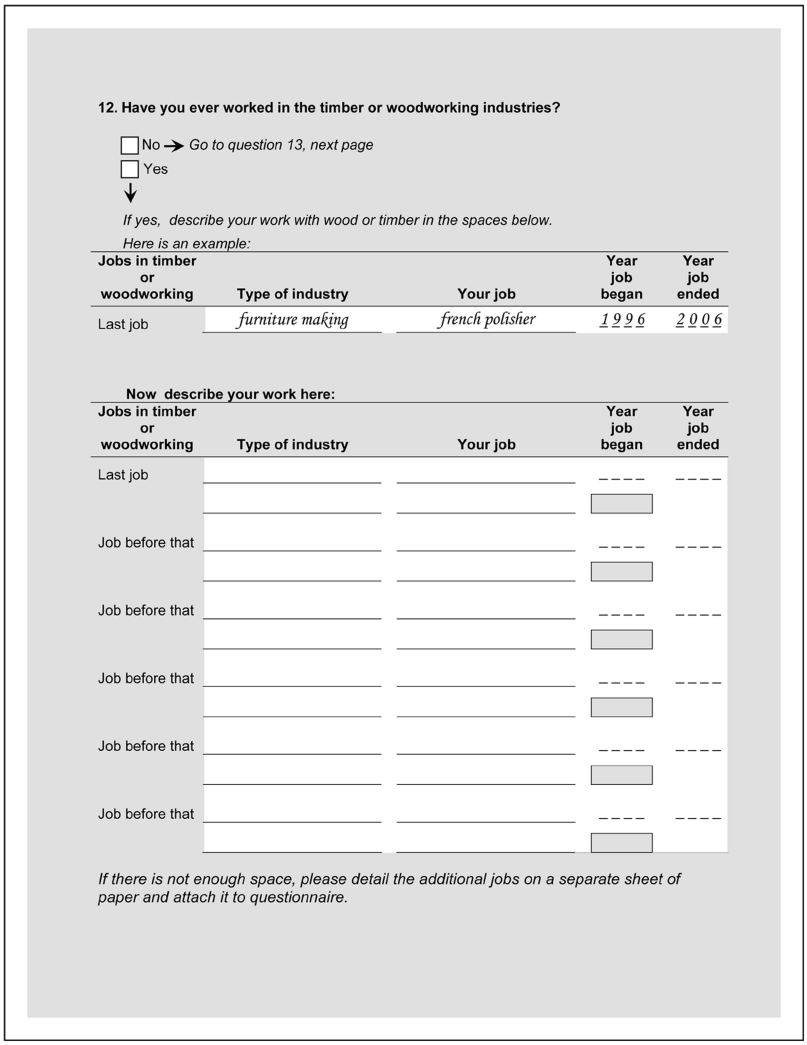
As noted in Table 6.2, the questions should be in bold typeface, the closed-ended response list should be in regular typeface, and italics are often used for instructions. Certain visual changes, such as underlining words or using capital letters, can be used to emphasize a change in concept, for example when the time frame differs between questions. However, such visual changes should not be over-used.

Matrix formats

The collection of data about a number of exposures of interest to epidemiology is done most efficiently in the form of a matrix. For example, all places of residence, all occupations, or all uses of a particular medication can be recorded on one axis, and a number of specific details regarding each episode or period of exposure can be recorded on the other. The conversion of such

matrices to a vertical single-column format increases the length of questionnaires substantially and is tedious to the interviewer or respondent. Therefore it is desirable to retain the matrix form in the questionnaire if possible. However, a matrix adds complexity to a self-administered questionnaire, and this may increase respondent burden and lead to missing data.

Figure 6.2 shows a matrix that was used to collect occupational information in a self-administered questionnaire. The data sought were limited in scope,



12. Have you ever worked in the timber or woodworking industries?

☐ No → *Go to question 13, next page*
☐ Yes
↓

If yes, describe your work with wood or timber in the spaces below.
Here is an example:

Jobs in timber or woodworking	Type of industry	Your job	Year job began	Year job ended
Last job	*furniture making*	*french polisher*	*1996*	*2006*

Now describe your work here:

Jobs in timber or woodworking	Type of industry	Your job	Year job began	Year job ended
Last job			____	____
Job before that			____	____
Job before that			____	____
Job before that			____	____
Job before that			____	____
Job before that			____	____

If there is not enough space, please detail the additional jobs on a separate sheet of paper and attach it to questionnaire.

Figure 6.2 An open-ended question with matrix response intended for use in a self-administered questionnaire.

and the matrix was kept small enough to fit on one page. The example included in the question was intended to show the respondent the form and detail of response that was desired, in the hope that instructions for the question could be kept to a minimum.

Asking about behaviours that vary over time

Epidemiologists are often interested in exposures occurring over many years or decades. A behaviour which varies appreciably over the time period of interest often presents difficulty in the design of questionnaires. It will rarely be sufficient to summarize it by way of a simple question on its 'usual' frequency, both because this summary is likely to introduce error and because the pattern of variation of the behaviour with time may be important (see Chapter 1). It may be unreasonably burdensome on the respondent to attempt to elicit a usual pattern for the behaviour (e.g. smoking) for each year of what is considered to be the aetiologically relevant time period. One solution is division of that period into larger but arbitrarily determined intervals of time (e.g. 0–4 years before diagnosis, 5–9 years before diagnosis, etc., or decades of age).

Another approach to measuring time-variable behaviour is to focus on obtaining details of the behaviour in periods of life in which it has been reasonably stable. Because this approach is complex, it would only be suitable for interviewer-administered questionnaires. Essentially, data are collected in a matrix format, except that reported changes in the subject's pattern of exposure over time determine when a new row is to be completed. This approach has been used in the Lifetime Drinking History described by Skinner and Sheu (1982), in which subjects are first asked about their alcohol consumption in the first year that they drank on a regular basis. Major changes in their drinking patterns are then identified chronologically. For each pattern, questions are asked about the frequency of drinking, typical and maximum quantity consumed per occasion, types of beverage, and when this pattern of drinking changed.

This approach has also been used to document exposure to the sun in a case–control study of skin cancer conducted following a survey of the prevalence of skin cancer in the town of Geraldton, Western Australia (Kricker *et al.* 1991). Subjects first completed, on their own, a life calendar covering each change of residence, school, occupation, and the number of days they worked each week. At interview, they were asked to identify, on the calendar, years in which major changes in their outdoor activity took place. They were prompted with the suggestion that these changes '. . . could've been due to things like changes in where you lived, changes in the sport you played, changes of school, different jobs, changes in the number of days you worked, and by marriage'. Having identified these years of change and, by inference, periods of comparatively

stable outdoor activity, subjects were asked detailed questions about outdoor activity in each period. These questions included the typical numbers of hours outdoors between 9 a.m. and 5 p.m. and 10 a.m. and 2 p.m., on working and non-working days, in both cooler and warmer months of the year, and during summer holidays, frequency of sunbathing, extent of suntan, frequency and severity of sunburn, use of sunscreen preparations, and wearing a hat.

Aids to recall

Disease may be influenced by exposures which occurred many years or decades before diagnosis. Therefore epidemiologists have long used a variety of approaches to improve recall of past exposures by subjects. These include use of lists of alternative answers, use of a life events calendar, and lead-in questions to help the participant recall a specific time period or event. These approaches are now supported by an increasingly strong body of research on memory (Bradburn *et al.* 1987; Jobe and Mingay 1989; Schwarz and Sudman 1996; Tourangeau *et al.* 2000).

The most commonly used aid to recall is the list of alternative answers supplied as part of a self-administered questionnaire. In an interview, this list may be provided in the form of a card given to the respondent. A list of possible answers may also be given for open-ended questions (e.g. a list of recreational physical activities), but it is essential that it be as comprehensive as possible or it may bias response. Where the list of possible answers to an open-ended question is very long, a list of headings under which all responses might fall can be used instead to stimulate recall.

In questions about use of the oral contraceptive pill, it is common to show the respondent photographs of all present and past formulations to assist her in eliciting the brand actually used. The photographs include the pill itself, the packaging, and the name. Similarly, Beresford and Coker (1989) found that six out of 14 women who had used post-menopausal hormone replacement therapy for 6 months or more were assisted in their recall of the name and dose by reference to photographs of preparations which had been in use over the preceding 15 years.

Another effective aid to recall is the *life events calendar*, on which subjects are asked to place personal landmarks (e.g. marriage, birth of children) and/or job and residential histories. As a means of improving recall of use of birth control methods, Daling and colleagues (Holt *et al.* 1989; J. Daling, personal communication) have used a simple month-by-year matrix calendar on which marriages and periods of 'living as married' were first recorded by use of a black marker pen, pregnancies were recorded by use of a blue marker, and periods of birth control use were recorded, with reference to the other details

on the calendar, by use of a red marker. Details for each of the periods of birth control use were then sought and recorded on the main data form. Engel *et al.* (2001) were able to elicit more jobs from study participants, particularly jobs from the distant past, by use of a life events calendar than from questions on recall of job history alone. Use of a calendar of residences, schools, and jobs to assist in recall of outdoor activities has been described earlier in this chapter.

Use of a life events calendar also improves the accuracy of recall of the *dates* of events, as well as recall of the events (Means and Loftus 1991). Research on memory has shown that most events are not stored with dates. Instead, people attach dates to events by relating them to datable events (e.g. marriage) and time periods (college, jobs, places of residence).

The use of a life events calendar is one example of the use of autobiographical sequences to aid recall (Bradburn *et al.* 1987). *Autobiographical sequences* are groups of events clustered in time, and often organized within some wider framework (e.g. a job or an illness), within which memory appears to be organized. Thus any means of entry into an autobiographical sequence, whether by way of calendar time, place lived, or through some highly salient event, such as childbirth or an illness, may assist in recall of events or behaviours of low salience. Thus, for example, asking first about illnesses which may have been indications for the use of particular medications may assist in recall of those medications. In two studies it was found that open-ended questions about prescription drug use during pregnancy without aids to recall only yielded 40–65 per cent of drugs eventually recalled. Additionally, drugs were recalled when questions were then cited on use of drugs for specific indications, and another 15–40 per cent of drugs were recalled only when the specific drug was named (Mitchell *et al.* 1986; de Jong-van den Berg *et al.* 1993).

A simple aid to recall is simply allowing the subject some time to think about the question. Thus longer introductions or redundant wording would allow a few more seconds for retrieval of information. Telephone interviews are at a disadvantage with regard to recall because silent pauses are more awkward on the telephone, and subjects might respond too quickly to retrieve the memory fully.

Recall may also be aided by asking the subject to refer to his/her personal records where they may be relevant. This is more easily done in a mailed questionnaire, where the respondent can refer to the records at leisure, than it is in a personal interview. However, these records may be made available in an interview by notifying the subjects in advance of the proposed lines of questioning and asking that they collect together whatever records they may have.

Pre-testing questionnaires

Overview

Pre-testing is an essential part of the development of all questionnaires, regardless of whether or not they have been substantially based on previous questionnaires. The objectives of pre-testing are to identify questions that are poorly understood, are ambiguous, or evoke hostile or other undesirable responses. Some of the questions that a pre-test should answer are (Dillman 2000) as follows.

- Are all the words understood?
- Are the questions interpreted similarly by all respondents?
- Does each closed-ended question have an answer that applies to each respondent?
- Are some questions not answered?
- Do some questions elicit uninterpretable answers?

Table 6.4 lists techniques of pre-testing questionnaires These include traditional methods such as expert reviews, debriefing meetings with interviewers after they have conducted a number of interviews, and analysis of the distribution of item responses, including percentage non-response. Also listed in the table are newer cognitive and behavioural approaches to questionnaire design and pre-testing that have been developed in the field of survey research (Schwarz and Sudman 1996; Esposito and Rothgeb 1997; Jobe 2003). These include cognitive or intensive interviews with respondents, interactive coding of interviewer and respondent behaviours during the interview, and testing different questionnaire versions in small experiments. Each of these methods is described in more detail below.

Pre-testing typically involves several of these methods applied to different versions of the questionnaire as it is revised and gets closer to its final form. Some of the techniques, such as expert reviews, cognitive interviews, and experiments, are more appropriate for earlier stages of pre-testing, while others, such as item response distribution analysis and validity studies, would be done on questionnaires which are closer to finalization. (Some authors use the word *pre-test* for the early testing of the questionnaire and the phrase *pilot test* for later testing of the study field methods, including selecting subjects, recruitment, and data collection). Some of the methods, such as intensive interviews and interactive coding, might be too costly in terms of time and money for most epidemiological studies. However, even when time and budget are restricted, one should at a minimum seek expert review of the questionnaire (e.g. from colleagues), conduct a pre-test on at least 20 test subjects followed by debriefings (of respondents for self-administered

Table 6.4 Methods of pre-testing questionnaires (for self-administered questionnaires and interviews)

EXPERT/PEER REVIEW for content, wording, and format
COGNITIVE/INTENSIVE INTERVIEWS WITH RESPONDENTS to see how items were understood and answered
Think-aloud techniques
Paraphrasing the question by respondent
Specific probes about how question was answered
INTERVIEWER DEBRIEFINGS to identify items with problems
Debriefing questionnaires
Group sessions
RESPONDENT DEBRIEFINGS to identify items with problems
Debriefing questionnaires or interviews
Focus groups
OBSERVATION OF INTERVIEWS and POSSIBLE INTERACTION CODING to identify interviewer and respondent behaviours for each question, such as rewording of question by interviewer or uncertainty about the answer expressed by respondent
ITEM RESPONSE DISTRIBUTION ANALYSIS
Distribution of valid responses
Percentage 'don't know"
Percentage missing
EXPERIMENTS, e.g. comparison of two versions of question wording
RELIABILITY AND VALIDITY STUDIES to quantify repeatability and accuracy

questionnaires or of interviewers for interviewer-administered questionnaires), and perform an item distribution analysis. Interviewer debriefing sessions and monitoring of the distribution of item responses should also be done at the beginning of actual data collection and, if necessary, further changes made to the questionnaire or protocol to resolve problems.

Ideally, the subjects in questionnaire pre-tests should be similar to the target population of the parent epidemiological study in terms of age, education, and study eligibility criteria. However, for a first pre-test, one could use a small sample of convenience such as co-workers, friends, or relatives of the interviewers. The mode of administration (self-administration, telephone, or in-person interview) should be the same as it will be in the parent epidemiological study.

Expert reviews

A good first step in pre-testing is to ask experts in the content areas or in questionnaire development to review the questionnaire and make comments.

Often one's professional colleagues can fill this role. They can help determine whether all necessary items have been included to meet the aims of the study, as well as providing advice on wording, format, etc. The study statistician and data processing supervisor also need to review the content and format of questionnaire to identify problems, before data collection begins, which might otherwise arise at the data processing and analysis phase.

Cognitive/intensive interviews

Cognitive interviews are a useful tool in refining questions on new or complex exposure areas. Cognitive or intensive interviews with participants gather detailed information about how the respondents formulate their answers to key questions (Belson 1981; Willis *et al*. 1991; Schwarz and Sudman 1996; Willis *et al*. 1999). Responding to a question involves several cognitive steps: understanding the question, retrieving the information from memory, judgement (such as estimation when an exact number cannot be recalled), censoring the response (typically to be more socially desirable), and providing an answer in the response format requested (Tourangeau *et al*. 2000). In *cognitive interview*s, the interviewer asks respondents to 'think out loud' while they are coming up with their response, or to explain how they came to their response after the answer has been given. This can aid the researcher in understanding whether a question is misunderstood by some subjects or whether it is complex and needs to be decomposed into simpler questions. Respondents can also be instructed to state when any question is difficult, annoying, or embarrassing to answer. *Intensive interviews* have the same elements as cognitive interviews, but also ask the subject to paraphrase the question after it is asked and give a confidence rating of the accuracy of their answer.

Cognitive interviews can also use item-specific probes to understand how specific concepts are understood. For example, for the question, 'How many people are there in your household?', the following questions (among others) would be appropriate.

- Did the respondent include him/herself in the count?
- To what period did the respondent think the question related, i.e. if a household member had been temporarily away, would he/she have been included?
- How did the respondent interpret the term household?

This is a demanding form of pre-test; Belson (1981) suggested that each respondent should be asked in detail about only three or four questions. However, this method has the potential to uncover problems in questions that might otherwise be missed.

Interviewer and respondent debriefings

Debriefings can be interviews, questionnaires, or group meetings conducted immediately or soon after the questionnaire is completed. Debriefings can be with the interviewer or respondent.

Interviewers should always be debriefed after they have pre-tested the questionnaire on a number of participants. They can be asked which items caused the most problems in terms of obtaining adequate answers from participants, and for those items, about what percentage of the time there was a problem. Experienced interviewers are very knowledgeable about which questions are misunderstood by respondents or are too taxing for the respondent to answer. Interviewer debriefing can take the form of a debriefing questionnaire or a group session with several interviewers where problem questions and possible modifications are elicited.

Respondent debriefings can be interviews, questionnaires, or focus groups. Respondents should be told that they are pilot participants before they complete the questionnaire or interview, and that their help is needed to identify any problems with the questionnaire. In the debriefing, they should be asked about which questions were confusing or hardest to answer and why, for which questions the answer they wanted to give was not an alternate response, whether any questions were offensive to them, and whether they found it easy to follow the skip patterns. Questions should also be asked about specific items about which the researcher has concerns; for example, did the participant notice that the time reference changed for a particular question? If the debriefing is an interview, the respondent should be asked why he/she skipped any questions that the interviewer notes should have been answered. Finally, respondents should be asked for any additional comments and suggestions.

Interaction or behaviour coding

In *interaction or behaviour coding*, a monitor listens to the interview (usually a tape recording) and codes specific behaviours of the interviewer and the respondent for each question (Schwarz and Sudman 1996; Dykema *et al*. 1997). An example of a behaviour coding scheme is given in Table 6.5. Length of pause between the end of the question and the answer (reaction time or response latency) is sometimes coded as an indicator of difficulty in answering the question. By analysing the frequencies of behaviours for each question, one can determine problem questions, for example which questions interviewers do not read as worded (often to attempt to improve the meaning of the question). Questions with a moderate proportion (e.g. 10 per cent or more) of respondent codes for uncertainty, uncodeable answers,

Table 6.5 Interviewer–respondent interaction coding scheme[a]

INTERVIEWER BEHAVIOURS in question-asking (assign one code only)	
Substantive change	Makes a substantive change in reading question as written
Incorrect prompt	Repeats question not as written or suggests an answer
Skips question	Skips applicable questions
Reads wrong question	Reads question that was not supposed to be read
RESPONDENT BEHAVIOURS (code all that apply)	
Interrupt	Interrupts question with an answer
Uncertain	Expresses uncertainty about question, requests clarification
Qualified	Qualifies answer
Uncodeable	Response does not meet question objectives, uncodeable
Don't know	Offers a 'don't know' response
Refusal	Refuses to answer

[a] Adapted from Dykema *et al*. (1997).

or interruptions can also be used to identify problematic questions which could be phrased better.

Dykema *et al*. (1997) tested the usefulness of behaviour coding in predicting questions with large measurement error. They recorded interviewer–respondent interaction codes (as shown in Table 6.5) for each question in a health-related questionnaire. They then evaluated the accuracy of the subjects' answers by comparing the responses in the health interview with medical records to establish whether the behaviour codes indicated inaccuracies. The interviewer behaviours were not associated with inaccuracy (see Chapter 7). However, if any of the respondent behaviours occurred for a question, that question was more likely to be inaccurately answered by that respondent. In particular, initial 'don't know' and qualified answers were associated with inaccuracy. Thus if these codes occur frequently for a question, it may be an indication that the question as phrased is difficult to answer accurately.

Item non-response and response distributions

After a reasonable number of representative pilot participants have completed the pre-test questionnaire/interview, the distribution of responses to each question should be reviewed. The percentage of non-response for each question is a particular concern. There are two types of non-response: missing/refusal and explicit 'don't know'. A major source of missing data in self-administered questionnaires is subjects not following the skip patterns. Improving the format of the questionnaire to make the skip pattern clearer

could reduce the amount of missing data. Another reason for missing/refusal is the sensitivity of the question. The techniques presented earlier in this chapter for asking sensitive questions might reduce the amount of non-response.

The second purpose of reviewing item response distributions is to review the frequency of each response category when questions are closed-ended. Response categories should be changed if some responses were selected by a very low or high percentage of respondents. An exception to this is when certain extreme response categories are given to encourage reporting of socially undesirable behaviours.

Experiments

As part of pre-testing, two (or more) versions of each question can be tested, usually with each questionnaire version given to a different group of subjects. The versions of each question can be different phrasing of the question or different categories for the answers for closed-ended questions. The two versions can be compared based on interviewer debriefings, respondent debriefings, item analysis, behaviour coding, or comparison with more accurate measures of the exposure (e.g. records or diaries). Interviewer debriefing questions could include: Which form of the questionnaire was easier to administer? Why was this form easier and the other more difficult? From this information, the researcher can select the best alternative of each question for the final questionnaire.

Reliability and validity studies

Conducting a validity or reliability study to evaluate the accuracy of an exposure measurement can be an important part of pre-testing. If one determines before an epidemiological study begins that the exposure is not measured with reasonable accuracy, then the researcher can continue to search for or develop a more accurate method. Chapter 4 covers the design, analysis, and interpretation of validity and reliability studies.

Revising the questionnaire

Typically a questionnaire is tested and revised several times before being used in the field. Once problems are identified during a pre-test, they can be resolved through changes in questionnaire wording, questionnaire format, or interviewer training. However, often a revised question which solves problems identified by some respondents will lead to problems for other respondents. For many questions, there will be at least a small percentage of respondents who have trouble understanding the question, or who have personal circumstances which make the question difficult for them to answer. Often adding explanations to make the question clearer to some respondents

will make the question longer, more burdensome, or even confusing to others. Thus it seems reasonable to modify only questions that are problematic for a moderate proportion of subjects (e.g. 10 per cent or more) and/or have a simple solution. For example, in pre-testing the self-administered question 'How many flights of stairs do you climb each day at home, work, or elsewhere?', some respondents questioned whether to count going up and down a flight of stairs as one or two flights. Others asked questions such as whether climbing up two floors three times a day counts as two, three or six flights per day, although most seemed to understand that the answer in this case should be six. Rather than adding lengthy explanations to solve these problems for a small percentage of subjects, this question was simply changed to 'How many flights of stairs do you climb *up* each day?'

Translating questionnaires

The conduct of international multicentre epidemiological studies often necessitates the translation of a questionnaire into a language other than that in which it was first developed. Translation may also be required when a population contains ethnic minority groups. There is evidence that the intramethod reliability of some questions in a questionnaire is greater when they are administered in the respondent's mother tongue, even when the respondent is multilingual (Becklake *et al.* 1987).

There are four phases in translation of a questionnaire (Del Greco *et al.* 1987):

- preliminary translation
- evaluation of the preliminary translation
- ascertainment of cross-language equivalence
- assessment of validity and reliability.

The preliminary translation aims at producing a translated questionnaire which is as near as possible in meaning to the original. It is best done by someone who understands both the overall objective of the questionnaire and the intent of each question, as well as being expert in both the original language and the language into which the translation is being made. The usual method of evaluating the preliminary translation is to have it translated back into the first language by someone who has not seen the original version. The back-translated version is then compared with the original version, and further work done on questions that have changed their meaning. Some questions may go through the process of re-translation and back-translation several times before they are considered to have been translated correctly. A complementary approach to back-translation in evaluating the preliminary translation is

to have bilingual experts evaluate the translation of each question in terms of its content, meaning, clarity of expression, and comparability to the original question.

Cross-language equivalence is determined by administering both the original and the translated versions of the questionnaire to bilingual subjects, and comparing their responses to each. It is usual to give half the subjects the original questionnaire first and half the translated questionnaire first to minimize order effects. A high correlation between the responses is taken to indicate cross-language equivalence.

Del Greco *et al.* (1987) noted that the reliability and validity of the questionnaire may not be maintained after translation, and it should be re-evaluated in the translated form. However, any change in validity and reliability could be due as much to cultural differences between the two populations as to any problems with the translation.

Where only a few interviews need to be conducted in a foreign language, it may be convenient to use a bilingual interviewer who translates from the original questionnaire as he/she interviews. Alternatively, an interpreter may be used. In either of these approaches, it is much more likely than with a carefully translated questionnaire that the questions will not be translated correctly and erroneous responses will be obtained.

Summary

The objectives of questionnaire design are to obtain, with minimum error, measurements of exposure variables essential to the objectives of the study and to create an instrument that is easy for both the interviewer and subject to use, and is easy to process and analyse.

Open-ended questions should be used as far as possible to seek the simple factual information that is most commonly needed in epidemiological studies. If closed-ended questions are used, the alternative answers offered should be simple, brief, mutually exclusive, and exhaustive.

The words used in questions should be the usual working tools of the respondents; jargon and complex, vague, and loaded words should be avoided. Questions should:

- contain only one concept
- not require any calculations
- have an unambiguous time reference.

Questions that ask about behaviour or attributes which are socially desirable or undesirable present a particular threat to subjects and require special care in wording.

The way in which questions are formatted, the order in which they are presented, the structure of the questionnaire as a whole, and the way it is printed are all important in facilitating its use by interviewers and respondents, in ensuring ease and accuracy of processing data from the questionnaire, and in minimizing error.

Aids may be used to assist subjects in recalling information. They include:

- lists of alternative answers
- photographs of specific agents to which the subject may have been exposed
- a calendar on which key dates are marked
- reference to personal records.

All questionnaires should be pre-tested before their use in the research study begins. The particular objectives of pre-testing are to see whether the questions are understood and elicit appropriate responses and to ensure that, where alternative answers have been provided, they cover the full range of relevant answers. Pre-testing should include, at least, initial evaluation by peers and testing on a sample of subjects from the population to be studied.

References

Aday, L.A. (1996). *Designing and Conducting Health Surveys: A Comprehensive Guide* (2nd edn). Jossey-Bass, San Francisco, CA.

Becklake, M.R., Freeman, S., Goldsmith, C., *et al*. (1987). Respiratory questionnaires in occupational studies: their use in multilingual work forces on the Witwatersrand. *International Journal of Epidemiology*, **16**, 606–11.

Belson, W.A. (1981). *The Design and Understanding of Survey Questions*. Gower, Aldershot, Hampshire.

Bennett, A.E. and Ritchie, K. (1975). *Questionnaires in Medicine*. Oxford University Press, London.

Beresford, S.A.A. and Coker, A.L. (1989). Pictorially assisted recall of past hormone use in case–control studies. *American Journal of Epidemiology*, **130**, 202–5.

Blair, E., Sudman, S., Bradburn, N.M., and Stocking, C. (1977). How to ask questions about drinking and sex: response effects in measuring consumer behavior. *Journal of Marketing Research*, **14**, 316–21.

Bradburn, N.M., Rips, L.J., and Shevell, S. K. (1987). Answering autobiographical questions: the impact of memory and inference on surveys. *Science*, **236**, 157–61.

de Jong-van den Berg, L.T.W., Waardenburg, C.M., Haaijer-Ruskamp, F.M., Dukes, M.N.G., and Wesseling, H. (1993). Drug use in pregnancy: a comparative appraisal of data collecting methods. *European Journal of Clinical Pharmacology*, **45**, 9–14.

De Leeuw, E.D. (2001). Reducing missing data in surveys: an overview of methods. *Quality & Quantity*, **35**, 147–60.

Del Greco, L., Walop, W., and Eastridge, L. (1987). Questionnaire development. III: Translation. *Canadian Medical Association Journal*, **136**, 817–18.

Dillman, D.A. (2000). *Mail and Internet Surveys: The Tailored Design Method* (2nd edn). Wiley, New York.

Dykema, J., Lepkowski, J.M., and Blixt, S. (1997). The effect of interviewer and respondent behavior on data quality: analysis of interaction coding in a validation study. In *Survey Measurement and Process Quality* (ed. L. Lyberg, P. Biemer, M. Collins, *et al.*), pp. 287–310. Wiley, New York.

Edwards, P., Roberts, I., Clarke, M., *et al.* (2002). Increasing response rates to postal questionnaires: systematic review. *British Medical Journal*, **324**, 1183–91.

Engel, L.S., Keifer, M.C., and Zahm, S.H. (2001). Comparison of a traditional questionnaire with an icon/calendar-based questionnaire to assess occupational history. *American Journal of Industrial Medicine*, **40**, 502–11.

Esposito, J.L. and Rothgeb, J.M. (1997). Evaluating survey data: making the transition from pretesting to quality assessment. In *Survey Measurement and Process Quality* (ed. L. Lyberg, P. Biemer, M. Collins, *et al.*), pp. 541–573. Wiley, New York.

Frey, J.H. (1986). An experiment with a confidentiality reminder in a telephone survey. *Public Opinion Quarterly*, **50**, 267–9.

Helsing, K.J. and Comstock, G.W. (1975). Response variation and location of questions within a questionnaire. *International Journal of Epidemiology*, **5**, 125–30.

Holt, V.L., Daling, J.R., Voigt, L.F., *et al.* (1989). Induced abortion and the risk of subsequent ectopic pregnancy. *American Journal of Public Health*, **79**, 1234–8.

Jobe, J.B. (2003). Cognitive psychology and self-reports: models and methods. *Quality of Life Research*, **12**, 219–27.

Jobe, J.B. and Mingay, D.J. (1989). Cognitive research improves questionnaires. *American Journal of Public Health*, **79**, 1053–5.

Kricker, A., Armstrong, B.K., Epglish, D.R., and Heenan, P.J. (1991). Pigmentary and cutaneous risk factors for non-melanocytic skin cancer: a case–control study. *International Journal of Cancer*, **48**, 650–662.

Krosnick, J.A. and Alwin, D.F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, **51**, 201–19.

Leigh, J.H. and Martin, C.R. (1987). 'Don't know' item nonresponse in a telephone survey: effects of question form and respondent characteristics. *Journal of Marketing Research*, **24**, 418–24.

McFarland, S.G. (1981). Effects of question order on survey responses. *Public Opinion Quarterly*, **45**, 208–15.

Means, B. and Loftus, E. (1991). When personal history repeats itself: decomposing memories for recurring events. *Applied Cognitive Psychology*, **5**, 297–318.

Mitchell, A.A., Cottier, L.B., and Shapiro, S. (1986). Effect of questionnaire design on recall of drug exposure in pregnancy. *American Journal of Epidemiology*, **123**, 670–6.

Nicholls, W.L. II, Baker, R.P., and Martin, J. (1997). The effect of new data collection technologies on survey data quality. In *Survey Measurement and Process Quality* (ed. L. Lyberg, P. Biemer, M. Collins, *et al.*), pp. 221–48. Wiley, New York.

Poe, G.S., Seeman, I., McLaughlin, J., Mehl, E., and Dietz, M. (1988). 'Don't know' boxes in factual questions in a mail questionnaire. Effects on level and quality of response. *Public Opinion Quarterly*, **52**, 212–22.

Schuman, H., Presser, S., and Ludwig, J. (1981). Content effects on survey questions about abortion. *Public Opinion Quarterly*, **45**, 216–23.

Schwarz, N. and Sudman, S. (ed.) (1996). *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. Jossey-Bass, San Francisco, CA.

Schwarz, N., Hippler, H.J., Deutsch, B., and Strack, F. (1985). Response scales: effects of category range on reported behaviour and comparative judgements. *Public Opinion Quarterly*, **49**, 388–95.

Singer, M. (1985). Mental processes of question answering. In *The Psychology of Questions* (ed. A.C. Graesser and J.B. Black), pp. 121–56. Erlbaum, Hillsdale, NJ.

Skinner, H.A. and Sheu, W.J. (1982). Reliability of alcohol use indices. The lifetime drinking history and the MAST. *Journal of Studies on Alcohol*, **43**, 1157–70.

Smith, T.W. (1987). That which we call welfare by any other name would smell sweeter. An analysis of the impact of question wording on response patterns. *Public Opinion Quarterly*, **51**, 75–83.

Sudman, S. and Andersen, R. M. (1977). Health survey research instruments. In *Advances in Health Survey Research Methods: Proceedings of a National Invitational Conference* (ed. L.G. Reeder), DHEW Publication No. (HRA) 77–3154, pp. 7–12. US Department of Health, Education and Welfare, Washington, DC.

Sudman, S. and Bradburn, N.M. (1983). *Asking Questions: A Practical Guide to Questionnaire Design*. Jossey-Bass, San Francisco, CA.

Tourangeau, R., Rips, L.J., and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge University Press.

Willis, G.B., Royston, P., and Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questionnaires. *Applied Cognitive Psychology*, **5**, 175–92.

Willis, G., DeMaio, T., and Harris-Kojetin, B. (1999). Is the bandwagon headed to the methodological promised land? Evaluating the validity of cognitive interviewing techniques. In *Cognition and Survey Research* (ed. M. Sirken, D.J. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau), pp. 133–54. Wiley, New York.