Questionnaires

Asking questions

If we want to know things about subjects, sometimes the easiest or only way is to ask them. Many healthcare studies make use of a questionnaire to elicit some or all of the data. Because questionnaires are familiar, written in words, and the best ones are designed to be simple and straightforward to complete, researchers sometimes fall into the trap of thinking that they must be easy to design. This is not so. Designing a questionnaire which is easy to complete, obtains the required information, and is easy to analyse is a difficult and time-consuming process, requiring just as much work as any other part of the research process. Jotting down a few questions in half an hour and passing them on to the typist is a recipe for disaster.

The way in which a question is asked may influence the reply. We should avoid questions which are leading, ambiguous, or in language which the respondent will not understand. Sometimes the bias in a question is obvious. Compare these:

(a) Do you think people should be free to provide the best medical care possible for themselves and their families, free of interference from a State bureaucracy?

(b) Should the wealthy be able to buy a place at the head of the queue for medical care, pushing aside those with greater need, or should medical care be shared solely on the basis of need for it?

Version (a) expects the answer 'yes', version (b) expects the answer 'no'. These are leading questions, directing the respondent to a particular answer. Another technique for asking a leading question is to start with a piece of apparently factual information:

'Most people think that medical statisticians are grossly underpaid. Do you agree?'.

Sometimes questions may lead by implying that an answer is foolish. An example would be:

'Do you have an unreasonable fear of heights?'

where to answer 'yes' is to admit to being unreasonable. A colleagues asked 120 women who had just had a cervical smear

'Do you understand the importance of having a cervical smear test?'

with the possible answers 'yes', 'no', and 'partly'. Not surprisingly, 118 respondents said 'yes'. Leading questions should be avoided.

Ambiguity is another problem in questioning. For example, Hedges (1978) reports several examples of the effects of varying the wording of questions. He asked two groups of about 800 subjects one of the following:

(a) Do you feel you take enough care of your health, or not?

(b) Do you feel you take enough care of your health, or do you think you could take more care of your health?

In reply to question (a), 82% said that they took enough care, whereas only 68% said this in reply to question (b). The second question is ambiguous, as it is quite possible to feel that you take enough care of your health while not doing everything possible.

Even more dramatic was the difference between this pair:

(a) Do you think a person of your age can do anything to prevent ill-health in the future or not?

(b) Do you think a person of your age can do anything to prevent ill-health in the future, or is it largely a matter of chance?

Not only was there a difference in the percentage who replied that they could do something, but this answer was related to age for version (a) but not for version (b).

| | Age (years) | | | |
	16-34	35-54	55+	Total
Can do something (a)	75%	64%	56%	65%
Can do something (b)	45%	49%	50%	49%

Here version (b) is ambiguous, as it is quite possible to think that health is largely a matter of chance but that there is still something one can do about it. Only if it is totally a matter of chance is there nothing one can do. Reasonably enough, older respondents were less likely to answer 'yes' to the unambiguous question (a), as decisions about health related behaviour in the past cannot be changed. For (b), however, the view that health is largely a matter of chance may be unrelated to age.

Ambiguity may occur in the possible replies to a question. The following comes from a questionnaire about health checks in general practice:

When was your check-up?   Less than one month
(Tick one answer only)   1 to 6 months ago
                                 6 to 12 months ago

Respondents who had a check 6 months ago would find this difficult to complete. A better version would be

When was your check-up?   Less than one month ago
(Tick one answer only)   1 to 6 months ago
                                 More than 6 months ago but less than one year ago

We may have two questions confused among the possible answers:

Would your prefer your smear to be taken by:
                 A female doctor
                 A male doctor
                 A nurse
                 I don't mind

Here the preference for a female and the preference for a doctor are mixed together and the respondent who wants a female to take the smear cannot answer.

Sometimes the respondents may interpret the question in a different way from the questioner. Children and their parents were asked:Do you (Does your child) usually cough first thing in the morning? Schoolchildren 3.7%   Parents 2.4%.Do you (Does your child) usually cough at other times in the day or at night?Schoolchildren 24.8%   Parents 4.5%.The different percentages giving positive answers to the second question showed that the children and their parents were not reporting the same thing. However, these reported symptoms all showed relationships to the child's smoking and other potentially causal variables, and also to one another, so they are measuring something.Another possibility is that respondents may not understand the question at all, especially when it includes medical terms. In a study of

cigarette smoking by children, we found that 85% of a sample agreed that smoking caused cancer, but that 41% agreed that smoking was not harmful (Bewley *et al.*, 1974). There are at least two possible explanations for being asked to agree with this: the negative statement 'smoking is not harmful' may have confused the children, or they may not see cancer as harmful. We have evidence for both of these possibilities. In a repeat study in Kent we asked a further sample of children whether they agreed that smoking caused cancer and that 'smoking is bad for your health' (Bewley and Bland 1976). In this study 90% agreed that smoking causes cancer and 91% agreed that smoking is bad for your health.

In another study, we asked children what was meant by the term 'lung cancer' (Bland *et al.*, 1975). Only 13% seemed to us to understand and 32% clearly did not, often saying 'I don't know'. They nearly all knew that lung cancer was caused by smoking, however.

Here is another example where respondents may not understand the question. The consultants Deloitte & Touche were commissioned to evaluate audio-visual services at St. George's Hospital Medical School. They sent round a questionnaire asking this:

> How often have you used this service?
>> Frequently   Often
>> Rarely   Never

Is 'frequently' more or less than 'often'? Deloitte & Touche think more.

We should always use simple words rather than complex ones, and we should always pilot questions very carefully to see that medical or technical terms are understood by our respondents.

Interviews and self-administered questionnaires

Questionnaires can be administered by an interviewer or completed by the subjects themselves, a self-administered questionnaire. Each approach has its advantages.

Self-administered questionnaires can be used either through the mail or for subjects who have come to the place of research, e.g. visiting a clinic. Compared to interviewer-administered questionnaires they are cheap and private, as the respondent does not have to tell anyone the replies directly. They can also be anonymous. They are suitable when the purpose of the study is fairly straightforward and can be explained in a few paragraphs of text. The questionnaire should be fairly short, particularly for mail questionnaires, and the questions must be very clear and unambiguous. Conditional questions of the form

> 'If "yes" go to question 7, if "no" go to question 23'

should be avoided if possible, as they make following the questionnaire difficult for the respondent.

Self-administered questionnaires should be avoided if there is a large amount of information to get, and if the study is difficult to explain. They should be avoided if there is likely to be a problem of literacy among the respondents, particularly where there are immigrants who may not have good command of the questionnaire language. Our experience in the UK has been that we obtained a very poor response from people of Asian origin, even when using own-language questionnaires. Such issues must be explored in pilot studies. Mail questionnaires are not suitable if it is important that the views of only one person are obtained, e.g. the views of a child rather than the parents, or of a patient rather than those of a carer. We can't be sure who completes a postal questionnaire. It may happen, for example, that a subject may pass the questionnaire to their spouse for completion. Scott (1961) reports a mail survey where 10% of questionnaires had been passed on to someone else to complete.

Interviews and self-administered questionnaires may produce different answers. For example, two random samples of GPs were asked about provision of counselling services in their practice (Sibbald *et al.*, 1994). One sample were approached by post and then by telephone if they did not reply after two reminders, and the other were contacted directly by telephone. These were the results:

Provided counselling:

	themselves	by health visitor
Postal sample	19%	14%
Telephone sample	36%	30%

The interviewer was able to probe.

Mail questionnaires usually have a lower response rate than interviewer questionnaires or questionnaires outside the home (e.g. in schools or clinics). If the questionnaire is not anonymous, we can send follow-up letters, preferably with another copy of the questionnaire. Moser and Kalton (1971) recommend two follow-up letters. They suggest as a rough guide that one gets the same percentage response rate each time, thus if 70% reply the first time then sending a reminder to the remaining 30% will generate 70% of 30% = 21% further response, and a second reminder will generate a further 70% of 9% = 5.4% further response. Clearly this is only an approximation, as if it ere true sending out repeated questionnaires indefinitely would quite rapidly approach a 100% response, which is unlikely. The bloody-minded are always with us!

Interviews are preferable if the issues are complex, if the questionnaire is long, or if a high response rate is essential. When things are complex it may be very helpful if the interviewer can probe ambiguous and incomplete answers, with supplementary questions such as 'How do you mean?' and 'In what way?'. One problem is interviewer bias, where the interviewer might by change words in the question or add explanations which indicate an answer. Interviewers must be trained. See Moser and Kalton (1971) for a discussion of interviewer training and interviewing techniques. A discussion of this in the setting of populations with low levels of literacy is given by Smith and Morrow (1991).

Confidentiality

In medical research confidentiality should be a fundamental part of the study design. We must tell our research subjects that we will respect the privacy of the data with which they provide us, and really mean it. In particular, we must assure our subjects that their replies will not influence any treatment which they may receive.

One way to guarantee confidentiality is anonymity, where we do not collect any identifying information at all. This has considerable disadvantages, however. It prevents us using interviewers. In postal surveys, it prevents us from following up non-responders, as we won't know who they are. It also prevents us from linking the questionnaire to other records about the subject. Another problem is that we sometimes want to use our questionnaire to select a sub-sample for further study, for which we must identify respondents.

The linking of anonymous questionnaires can sometimes be done by asking respondents to invent their own serial numbers. This can be done by asking them to quote some combination of letters and numbers which they will remember but which will not enable the investigator to work out who they are. Birth dates are a good basis for this. Clearly such methods need very careful piloting, as the serial number must be one which the respondent will be able to recreate when next asked to complete a questionnaire.

The invented serial number method cannot be done if we want to select a sub-sample, as we must actually identify the subjects. We can use identifying information other than the name, however. For example, Chadwick *et al.* (1989) wanted to select a group of school children who were habitual abusers of volatile substances ('glue sniffers') from a questionnaire given

to all children in several school years. Abusers and a control sample of non-users would then undergo a battery of neuropsychological test to look for any deficits associated with volatile substance abuse. The questionnaire was self-administered in the classroom, and asked questions about cigarette smoking, volatile substance abuse, alcohol consumption, and health. Because of the possibly sensitive nature of the data we did not want to ask the children to put their names on the questionnaire. We asked the children to give us their dates of birth and the name or number of their school class or tutor group. We then used school registers to identify those who we wished to study. This may have fooled some of the children some of the time. It did not appear to guarantee truthfulness of replies. In this study we use a mass spectrometer to analyse the exhaled breath of the subjects in the sub-sample for traces of abused substances. We detected 1,1,1-trichloroethane or toluene in the breath of seven index children, who had reported volatile substance abuse on the self-completion questionnaire, and toluene in one control, who had denied volatile substance abuse.

Questionnaire design

Questionnaires should be clearly set out and legible, and any branches in the questionnaire should be very clearly indicated. We find that horizontal rules between questions, or boxes round them, are a useful way of clarifying the structure of a questionnaire:

1) Do you usually cough first thing in the morning?	
(please tick one box)	YES ☐
	NO ☐

2) Do you usually cough during the day or at night?	
(please tick one box)	YES ☐
	NO ☐

3) Do you get short of breath when hurrying on flat ground or walking up a slight hill?	
(please tick one box)	YES ☐
	NO ☐

Questionnaires which are to be completed by the respondent should be attractive documents. We should try to make the respondents' task as pleasant and interesting as possible. They are providing us with information which is for our benefit rather than theirs, usually for no reward. We should therefore be polite to our respondents, inserting words like 'please' where appropriate. Only tax and immigration authorities can afford to be rude!

Questionnaires get lost. Coloured paper is useful, as it makes it much easier for respondents to locate the questionnaire among the pile of bills. When several questionnaires are used in a study, it is a good idea to make each a different colour.

Types of question

Most questions ask about facts, such as age, sex, etc., or opinions, such as whether smoking should be allowed in public places. Several styles of question can be used. Questions are open or closed. **Open questions** allow the respondents to answer in whatever way they wish, e.g.:

What one improvement would you make to this course?

_____

_____

_____

This style of question is useful when we want to get some ideas, as in this example where we want to get ideas for improving the course. Such questions can be used in pilot studies at an early stage in an investigation, where they can help us to design more structured questions in suitable language for use in the main study. They are not much use in large studies where the data are to be used in statistical analysis. **Closed questions** present the respondent with a choice of predetermined responses. Most questionnaires are of this type.

The simplest questions are of the multiple choice type, where the respondent has to choose one of two or more possible answers:

Please read these statements carefully and tick the one box which best describes you. (Please tick one box only)
I have never smoked a cigarette ☐
I have only tried smoking once ☐
I have smoked sometimes, but I don't smoke as much as one cigarette a week ☐
I usually smoke between one and six cigarettes a week ☐
I usually smoke more than six cigarettes a week ☐

It is important in wording such questions that the categories are mutually exclusive and include all the possibilities. In the layout, the answers should have sufficient vertical space between them that the respondent cannot mistake which box applies to which answer.

Another popular style of question is the **check-list**, where respondents can choose more than one answer:

Has your child ever had any of the following diseases:	YES	NO
asthma	☐	☐
bronchitis	☐	☐
croup	☐	☐
hay fever	☐	☐
pneumonia	☐	☐
tonsillitis	☐	☐
whooping cough	☐	☐

When a check-list question is laid out like this, many respondents will tick only the 'YES' boxes for the relevant diseases, and leave the 'NO' boxes blank. The investigator must then decide whether to treat these as missing information or as genuine 'NO' answers. We can avoid the problem by presenting the question without the 'NO' boxes:

Has your child ever had any of the following diseases: (Tick all the boxes which apply)	
asthma	☐
bronchitis	☐
croup	☐
hay fever	☐
pneumonia	☐
tonsillitis	☐
whooping cough	☐

Note that the boxes should not be so far from the responses that the respondent can become confused over which box is which.

Occasionally we use questionnaires to ask for numerical information, such as age, height, weight, family size, etc. For example

How old are you? _____ years

or

How old are you? ☐☐☐ years

Remember to allow sufficient space for the answer. If you use boxes, give sufficient boxes for the largest number. If your population could include someone who is 100 years old, give three boxes. The addition of the unit, years, is useful and is essential if the answer may have more than one unit, as for height or weight:

Would you tell us your weight, please?
☐☐ stones ☐☐ pounds
OR ☐☐☐ pounds
OR ☐☐☐ kilogrammes

If possible, we should collect such data as accurately as possible. It is possible to ask questions about numerical variables where we group the possible answers:

How old are you? (Please tick one box)	
less than 18 years	
18 to 44 years	☐
45 to 64 years	☐
65 to 74 years	☐
75 years or more	☐

Such grouping should be avoided unless there is a very good reason for it, e.g. for income. Asking age in groups restricts the analysis which can later be done, and may make it difficult to compare your study with others. Certainly, if we have asked the question so as to elicit a number, we should not group the data before we enter them into the computer. We should be able to use all the information offered by the respondent. Should we wish to group the variable later, the computer can do that for us.

The question designs discussed so far are mainly concerned with factual information. To ask about opinions we mostly use different styles. Of course, we can simply ask

Are you in good health?   YES ☐   NO ☐
   (please tick one box)

but this is a fairly crude instrument. It is better to use a **rating scale**, where a graded series of options is offered:

Which word best describes your health? (Please tick one box only)	
excellent	☐
good	☐
fair	☐
poor	☐

A useful method of asking about opinions is the **Likert scale**, where the respondent is asked how much they agree with a statement of opinion. Often several such statements are asked together.

		Strongly agree	Agree	Don't know	Disagree	Strongly disagree
1.	A pupil who plays truant or skives from school should be punished.	☐	☐	☐	☐	☐
2.	Cigarettes should be harder to get.	☐	☐	☐	☐	☐
3.	Others make fun of you if you don't smoke.	☐	☐	☐	☐	☐
4.	Sometimes my brother or sister gives me a cigarette.	☐	☐	☐	☐	☐
5.	My parents do not mind whom I go around with.	☐	☐	☐	☐	☐
6.	Smoking is a dirty habit.	☐	☐	☐	☐	☐
7.	Smoking is only bad for you if you smoke a lot.	☐	☐	☐	☐	☐

There are a few general principles which should be applied to such attitude statements:

- ❖ They should be single sentences including only one idea.
- ❖ They should be short, fewer than 20 words.
- ❖ They should avoid absolute terms like 'all', 'none', 'always', and 'never'.
- ❖ They should avoid statements which are either true or false.

The fourth item violates two of these principles: it includes two ideas, whether or not the respondent has a sibling and whether or not this sibling gives cigarettes, and it is factual. It would have been better asked in a different way.

Sometimes we want to know how respondents would choose between a set of items where all might be rated positively (or all negatively) if asked separately. We can ask respondents to rank the items in order of importance. For example:

The following terms all might be used to describe a GP. Please put them in order of how important they would be to you when choosing a new GP. Put numbers 1 to 5 in the boxes, from 1, most important, to 5, least important.
Keen on preventive medicine
Good with children
Up to date with medical research
Patient
Friendly

Such questions should be used sparingly, as they are very difficult to analyse. With only five items there are 120 different possible orderings. The rank given to each item should be entered into the computer, each item forming a separate variable. The mean rank for each question can be used to order the items to give a descriptive summary.

We often ask questions to which there is a graded response, e.g.

How would you describe your health?
    1. excellent

    2. good

    3. fair

    4. poor

We would use the numbers 1, 2, 3, and 4 as our data. It is a short step to thinking of these numbers as a scale of health. For example, we used a nine-point scale in a trial where patients with psychological problems were randomized to treatment by a clinical psychologist or by their GP (Robson *et al.*, 1984). Subjects were asked to rate the severity of the problem from 0 to 8, with verbal labels being attached to alternate numbers:

  0 no problem
  1
  2 only very slight (and/or occasional)
  3
  4 fairly severe (and/or quite frequent)
  5
  6 quite severe (and/or most the time)
  7
  8 very severe (and/or all the time)

GPs, subjects, and another member of the subject's household were asked to score the problem at the start of treatment and at four subsequent times. The improvement in the problem was then measured by the change in score.

We do not need to include labels for points on the scale. We can simply ask for a number. For example, we might ask:

Can you give the pain a number between one and ten, where 1 means no pain at all and 10 means the worst pain you can imagine?
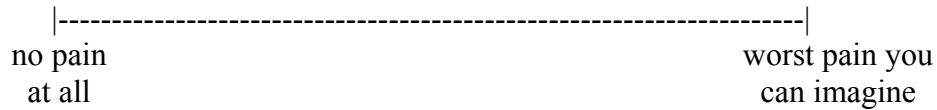
    Pain (1 to 10): _____

or we can present it like this:

    Can you give the pain a number between one and ten, where 1 means 'no pain at all' and 10 means the worst pain you can imagine?
    Circle the number which best describes the pain:
    no pain at all 1 2 3 4 5 6 7 8 9 10 worst pain you can imagine

It is a natural step from such scales to a measurement on a continuous scale, which can be done using a visual analogue scale. A **visual analogue scale** (**VAS**) consists of a straight line ruled on the questionnaire, marked at either end with words which describe the extremes which the end of the line represents. For example, a line used to measure pain might be marked 'no pain at all' and 'worst pain you can imagine'.

```
    |------------------------------------------------------------------|
  no pain                                                    worst pain you
    at all                                                      can imagine
```

For ease of measurement and interpretation, most investigators use a 10cm line. It is very important that all coders use the same units, e.g. millimetres! Sometimes scales are marked at 1 cm intervals, and some investigators record the scale only to the nearest cm or 0.5 cm. It is better to measure as accurately as possible.

Coding

Most questionnaires are analysed using a computer program. Whichever program you use, you will have to code the data before you can put it into the computer. Some statistical programs will accept alphabetic characters as input, others only numeric. On the whole, it is a good idea to stick to numerical codes for statistical purposes.

We shall consider how numerical codes are assigned to some typical questions, adapted from a patient survey of health checks in general practice (Ochera *et al*, 1994). The general principles are that coding should be clear, unambiguous, simple, and help us to avoid keying errors.

First we look at a simple multiple choice question with only two possible answers:

Sex: Male   Female

We need numeric codes for 'male' and 'female'. The usual choice of codes is male=1, female=2. We should ***not*** use male=0, female=1. Some programs do not distinguish between zero and blank. It is thus very easy to type a zero by mistake. Zero codes should be avoided if possible.

It is a good idea to record the code on the form itself, lest it be forgotten:

Sex: Male $_1$ Female $_2$

What about the comedian who answers this rather crudely designed question with 'Yes, please'? (Every youth who scrawls this thinks that it is highly original!) We do not know the sex of the respondent, so we need a **missing data code**. Some programs use a numeric missing data code, some use a special symbol, such as '.' or '*'. If a numeric code is used, it is conventional to use a string of '9's. Blank and zero should not be used, to avoid input errors.

The next example is a multiple choice question with several possible answers, but only one choice is allowed:

How many follow-up appointments have you had at the surgery (including attendances at groups)?	
one appointment	1
2 to 5 appointments	2
6 to 10 appointments	3
more than 10 appointments	4

Here we can code the answers 1, 2, 3, 4, with a missing data code 9 for those who do not answer or tick more than one. This question should only be answered by patients who had had a check-up, and who were invited back to the surgery for a further visit after the check-up. These will not be all patients, so we need another code for 'not applicable'. This could be 5. When the missing data code is 9, the not applicable code is often 8. Some programs (e.g. SPSS) allow you to define more than one missing data code for a variable, so that 'not applicables' can be excluded from analysis easily if this is required.

The next question is a check-list. It has several possible answers, and several possible choices are allowed:

At your check-up, did the nurse or doctor give you advice about any of these things?
Smoking
How much alcohol to drink
Exercise
What food to eat
Your weight
Your blood pressure

We cannot code these as 1, 2, 3, 4, 5, 6. How would we code someone who ticked all the items? In fact this is not one question, but six. It could equally be written:

	Yes	No
At your check-up, did the nurse or doctor give you advice about smoking?		
At your check-up, did the nurse or doctor give you advice about how much alcohol to drink?		
At your check-up, did the nurse or doctor give you advice about exercise?		
At your check-up, did the nurse or doctor give you advice about what food to eat?		
At your check-up, did the nurse or doctor give you advice about your weight?		
At your check-up, did the nurse or doctor give you advice about your blood pressure?		

We therefore code each item separately as yes=1, no=2. The question produces six separate variables. In this case there will also be a code for not applicable, because not all respondents are asked the question. The question might be presented like this:

	Yes	No
At your check-up, did the nurse or doctor give you advice about smoking?	1	2
At your check-up, did the nurse or doctor give you advice about how much alcohol to drink?	1	2
At your check-up, did the nurse or doctor give you advice about exercise?	1	2
At your check-up, did the nurse or doctor give you advice about what food to eat?	1	2
At your check-up, did the nurse or doctor give you advice about your weight?		2
At your check-up, did the nurse or doctor give you advice about your blood pressure?	1	2

The next question is open:

If you have any other views on the check-up, please write them in the space opposite:	| | | |

Questions like this are asked because we do not have a list of options. If we want to code it, we first carry out a **content analysis** of either all or a sample of questionnaires. We read the answers and note down the ideas or topics which respondents mention. We then code this as for the advice question, with a separate variable and separate code 1 or 2 for each topic. We can then analyse them like any other 'yes/no' question. Questions like this are very useful in pilot studies or in small in-depth surveys, but in large studies they are seldom of much value. It takes too long to code them and coding is too subjective.

Validity of questions How well do questions measure what we want them to measure? For factual questions we can test by checking other sources. E.g., to check the validity of

'Has your child ever had asthma?'

we can compare parents' answers to medical records. Sometimes there is no other direct source of information. E.g. we might ask a child

'Have you ever smoked a cigarette?'

This is factual but the only available source of information is the subject. We must rely on reliability or repeatability, i.e. to what extent do the same people give us the same answers, and whether we get consistent relationships with other variables.

Sometimes validity is difficult because the question is ill-defined. For example, the MRC Chronic Bronchitis Questionnaire contains:

'Do you usually cough first thing in the morning?'Exactly what is meant by 'usually', 'cough', and 'first thing in the morning'? These terms are not well defined. This question is often asked to children, who do not have chronic bronchitis. How can we assess the validity? We can measure reliability, we can compare results of questioning different observers to get their subjective opinions, e.g. children and their parents, and we can test for differences in related objective measurements between those giving yes and no answers, e.g. measured lung function.When there is no factual component, as in attitude statements, we rely on construct validity. This means that we look for internal consistency between related questions and for expected relationships with other variables.**Questionnaire scales**In medicine we often want to measure ill-defined and abstract things, like disability, depression, anxiety and health. The obvious way to decide how depressed someone is to ask them. However we cannot just ask 'how depressed are you out of 10?', as people would not have a common scale. Instead, we ask a series of questions relating to different aspects of depression and then combine them to give a depression score.

For example, this is the depression scale of the GHQ:HAVE YOU RECENTLY

been thinking of yourself as a worthless person?	Not at all	No more than usual	Rather more than usual	Much more than usual
felt that life is entirely hopeless?	Not at all	No more than usual	Rather more than usual	Much more than usual
felt that life isn't worth living?	Not at all	No more than usual	Rather more than usual	Much more than usual
thought of the possibility that you might make away with yourself?	Definitely have	I don't think so	Has crossed my mind	Definitely not
found at times you couldn't do anything because your nerves were too bad?	Not at all	No more than usual	Rather more than usual	Much more than usual
found yourself wishing you were dead and away from it all?	Not at all	No more than usual	Rather more than usual	Much more than usual
found that the idea of taking your own life kept coming into your mind?	Definitely have	I don't think so	Has crossed my mind	Definitely not

Scoring for the depression scale of the GHQ:HAVE YOU RECENTLY

been thinking of yourself as a worthless person?	Not at all	No more than usual	Rather more than usual	Much more than usual
	0	1	2	3
felt that life is entirely hopeless?	Not at all	No more than usual	Rather more than usual	Much more than usual
	0	1	2	3

14

felt that life isn't worth living?	Not at all 0	No more than usual 1	Rather more than usual 2	Much more than usual 3
thought of the possibility that you might make away with yourself?	Definitely have 3	I don't think so 2	Has crossed my mind 1	Definitely not 0
found at times you couldn't do anything because your nerves were too bad?	Not at all 0	No more than usual 1	Rather more than usual 2	Much more than usual 3
found yourself wishing you were dead and away from it all?	Not at all 0	No more than usual 1	Rather more than usual 2	Much more than usual 3
found that the idea of taking your own life kept coming into your mind?	Definitely have 3	I don't think so 2	Has crossed my mind 1	Definitely not 0

Questions are scored 0, 1, 2, 3 for the choices from left to right for items 1, 2, 3, 5, and 6, and 3, 2, 1, 0 for items 4 and 7. The sum of these is the score on the depression scale.

The questions are clearly related to one another and together should make a scale. Anyone who truthfully gets a high score on this is depressed. The full questionnaire has four such scales.

Questions are formed into a scale as follows:A set of questions which are expected to be related to the concepts of interest is devised, based on experience.The questions are answered by test subjects.The scales are checked for internal consistency.Dubious questions are excluded and the scale tested again.Validation of the scale is by tests of reliability and by its relationship to other measures of related quantities. For example the depression scale can be given to patients with diagnosed clinical depression, patients with other diagnoses and people with no psychiatric diagnosis, to see how well it distinguishes between them.This is another depression scale, the depression scale of the CCEI (Crown Crisp Experiential Index):

Can you think as quickly as you used to?	Yes	No	
Do you feel that life is too much effort?	At times	Often	Never
Do you regret much of your past behaviour?	Yes	No	
Do you wake unusually early in the morning?	Yes	No	
Do you experience long periods of sadness?	Never	Sometimes	Often

Do you have to make a special effort to face up to a crisis or difficulty?	Very much so	Sometimes	Not more than anyone else
Do you find yourself needing to cry?	Frequently	Sometimes	Never
Have you lost your ability to feel sympathy for other people?	No	Yes	

This is the coding for the CCEI:

Can you think as quickly as you used to?	Yes 2	No 0	
Do you feel that life is too much effort?	At times 1	Often 2	Never 0
Do you regret much of your past behaviour?	Yes 2	No 0	
Do you wake unusually early in the morning?	Yes 2	No 0	
Do you experience long periods of sadness?	Never 0	Sometimes 1	Often 2
Do you have to make a special effort to face up to a crisis or difficulty?	Very much so 2	Sometimes 1	Not more than anyone else 0
Do you find yourself needing to cry?	Frequently 2	Sometimes 1	Never 0
Have you lost your ability to feel sympathy for other people?	No 0	Yes 2	

In practice, these questions are interspersed between questions related to five other psychiatric scales.

Presenting scales Both the GHQ and CCEI share some features in their presentation. Some answers go from left is low to right is high, and some the opposite way. This reduces the tendency to tick the first box all the way down. The answers are varied in wording. This is to avoid monotony and to encourage respondents to read and think about the items. In the CCEI, the order of high scoring answers is varied, so that sometimes the highest or lowest is in the middle of three options, not at the end. This further encourages respondents to read and think about the question. In the full questionnaires, the sub-scales are mixed up, so that it is less obvious to the respondent what the questions are trying to elicit.There are many types of scale in regular use. This is one of several possible formats. Scales are difficult to design and validate, and so whenever possible we use one which has been developed previously, such as the GHQ.

This also makes it easier to plan and to interpret the results of studies, as the properties of the scale are already known. However, we should always check that the language is appropriate to the population being studied, particularly when using questionnaires developed in other countries. Language may change with place. For example, US questionnaires might refer to the 'doctor's office', whereas in the UK we call this the 'doctor's surgery'. The doctor's office is where he writes deathless prose or plays solitaire, To be frivolous, in the UK 'blow me!' is a mild expression of surprise, in the USA it is a request which can get you struck off.

Language may change over time, and we should check that questions still mean what they used to mean. For example, the EPI (Eysenck Personality Inventory) used to include the question

Do you like gay parties?

This became

Do you like lively parties?

following a change in usage of the word 'gay' to mean 'homosexual'.

Sensitive questions

People in the UK are remarkably willing to tell their most intimate secrets to complete strangers with clip-boards. They will tell an interviewer things they would never dream of telling their spouses, and may tell the interviewer how good it is to be able to talk about these topics to someone. Of course, we can rarely be sure what is truthful and what concealment or exaggeration. Sometimes respondents are reluctant to tell an interviewer the truth.

It is surprising what topics people find sensitive. We might think that sex, drugs and criminal activity would be the difficult ones. In fact, many respondents seem to find talking to a researcher about their sexual behaviour quite easy and unthreatening, perhaps because they can't talk about it to anyone else. What bothers some people in the UK is questions about their income and other financial arrangements. One parent 'phoned us in a great rage because we had asked his son whether the family home was rented or owned, ignoring the detailed questions we had asked about smoking, drinking and solvent abuse. Income is surprisingly sensitive. Perhaps respondents think that we will pass this information directly to the Inland Revenue. All we can say is that we suffer as much at their hands as everyone else.

Even a simple question such as 'For whom would you would vote in an election?' can be very sensitive. Opinion pollsters International Communications and Market Research conducted a poll in which half the subjects were questioned by interviewers about their voting preference and half were given a secret ballot, which they sealed in an envelope before handing back to the interviewer (McKie 1992). By each method 33% chose 'Labour', but 28% chose 'Conservative' at interview and 7% would not say, whereas 35% chose 'Conservative' by secret ballot and only 1% would not say. Hence the secret method produced a Conservative majority, as at the then recent 1992 UK general election, and the open interview a Labour majority. As the polls had got it wrong in 1992, it seems likely that reluctance to tell an interviewer of the intention to vote Conservative may have been an important factor.

The sensitivity of items may differ from culture to culture. The investigators should already have some insight into what their respondents will deem sensitive. This issue should also be explored in pilot studies.

When a question is sensitive, one possibility is a secret ballot, as was done in the opinion poll study described above. This may give a much better estimate of the population proportion, but the main difficulty is that we cannot link the answer to other data. For example, if we ask about voting intention by secret ballot and ask about social class at interview, we cannot look at the relationship between voting intention and social class. This leads us to add the social class question to the secret ballot and we end up with a self-administered questionnaire.

Does sensitivity matter? If our main purpose is to estimate the population value, such as the proportion of people in the population who have used an illegal drug in the past year, then refusals to answer and misleading answers from drug-users will cause the estimate to be

wrong. If we are mainly concerned with comparing the sensitive item between different groups of people, things may not be quite so bad. However, it is quite possible that the sensitivity of the question will vary between groups we wish to compare. An obvious example would be the comparison of contraception between different religious groups. This might then produce quite spurious relationships, where what is related is not the actual thing we wish to study but the willingness to talk about it. This issue is one which must be considered very carefully in the design and piloting of questionnaire studies.

Inevitably, we often want to ask sensitive questions in medical research. Despite the problems mentioned above, we can do this. First, we must gain the trust of our respondents. Second, we must convince them that their replies will remain confidential. Third, we must ask the question in a non-threatening way.

We can make our sensitive question less threatening by putting it in a group of similar but unthreatening questions. Rather than ask:

Have you ever had gonorrhoea?   YES   NO
(please tick one box)

we can place the item in a check list:

Have you ever had any of the following diseases?	
(please tick all that apply)	
measles	German measles (rubella)
chicken pox	gonorrhoea
tuberculosis	syphilis
scarlet fever	rheumatic fever

If we want to ask several questions about a sensitive subject, we can include them in the middle of a questionnaire which asks about other subjects. In their study of volatile substance abuse, Chadwick *et al.* (1989) wanted to identify children who had abused. The questionnaire began with general questions about age, sex, and social circumstances. These were followed by three groups of similar questions: first a group about cigarette smoking, then one about volatile substance abuse, and finally a group about alcohol consumption. The questionnaire finished with some questions about general health. This seemed to work quite well, and the children decided that it was alcohol that we were really after.

Sometimes we need to reassure the respondent that the behaviour which we are asking about is not going to shock us. For example, rather than asking a sensitive question like this:

Do you masturbate? YES NO
(please tick one box)

If YES, how often do you do this?
(please tick one box)

    most days

    most weeks

    most months

---

we can start with a reassuring comment:

Most people masturbate. How frequently do you?
(please tick one box)

    most days

    most weeks

    most months

    sometimes but not most months

    never

A similar approach can be used when we want a numerical answer which respondents might be reluctant to supply. We suggest to respondents that an answer much more extreme than theirs would not surprise us. For example, if we want to ask a population of alcoholics how much they drink, we can expect an underestimate if we ask:

How many bottles of spirits do you drink in a typical week? _____ bottles

We can reassure the respondent like this:

How many bottles of spirits do you drink in a typical week?
    less than one bottle

    one or two bottles

    between three and five bottles

    between six and nine bottles

    between ten and fifteen bottles

    between sixteen and nineteen bottles

    between twenty and twenty-four bottles

    between twenty-five and twenty-nine bottles

    between twenty-five and twenty-nine bottles

    thirty bottles or more

The idea is that the respondent who drinks fifteen bottles a week will feel happier to tell us

this if we suggest that twice this would not startle us. Such techniques must be tested thoroughly in pilot studies before the final questionnaire is designed.

J. M. Bland

References

Bewley, B.R. and Bland, J.M. (1976) Academic performance and social factors related to cigarette smoking by schoolchildren. *British Journal of Preventive and Social Medicine* **31**, 18-24.

Bland, J.M., Bewley, B.R., Banks, M.H., and Pollard, V.M. (1975) Schoolchildren's beliefs about smoking and disease. *Health Education Journal* **34**, 71-8.

Chadwick, O., Anderson, R., Bland, M., Ramsey, J. (1989) Neuropsychological consequences of volatile substance abuse: a population based study of secondary school pupils. *British Medical Journal* **298**, 1679-84.

Mckie, D. (1992) Pollsters turn to secret ballot. *The Guardian* London, 24 August, 20.

Moser, C.A. and Kalton, G. (1971) *Survey Methods in Social Investigation, 2nd. ed.* Heinemann, London.

Ochera, J., Hilton, S., Bland, J.M., Jones, D.R., Dowell, A.C. (1994) Patients' experiences of health checks in general practice: a sample survey. *Family Practice* **11**, 26-34.

Robson, M.H., France, R., and Bland, M. (1984) Clinical psychologist in primary care: controlled clinical and economic evaluation. *British Medical Journal* **288**, 1805-8.

Sibbald, B., Addington Hall, J., Brenneman, D., Freeling, P. (1994) Telephone versus postal surveys of general practitioners. *British Journal of General Practice* **44**, 297-300.

Scott, C. (1961) Research on Mail Surveys. *Journal of the Royal Statistical Society, A* **124**, 143-205.