

RESEARCH METHODS

Questionnaire design

Lesley Fallowfield

Measurement is a vital component of scientific research and much of the clinical research undertaken by the medical and allied professions utilises a questionnaire. Information gleaned from questionnaires may have important implications for the delivery of health care in a wide variety of areas. For example, the development of questionnaires with good psychometric properties has helped to establish health related quality of life parameters as valid outcome measures in clinical trials. Knowledge gained from questionnaire data about the toxicity and psychosocial burden imposed by different treatments may assist doctors and their patients in decision making about therapeutic options. Health care managers and economists responsible for budgets and policy decisions frequently integrate data derived from questionnaires and surveys when determining the allocation of resources or provision of services.

Questionnaires are sometimes the only means of gathering objective information, thus their construction demands considerable skill and thought as with any other form of measurement in medicine. Too often, people assume that anyone can cobble together a few items on a questionnaire. This may then be sent with little explanation to busy or anxious personnel or patients within the hospital, many of whom have neither the inclination nor the time to respond. Furthermore, if the questions seem irrelevant, insensitive, intrusive, uninteresting, or with a layout or format too confusing or complex to comprehend, response rates may be poor and decisions made on the basis of unsound results. Often the supposed 'softness' of questionnaire data is more a function of the lackadaisical fashion in which the instruments are constructed and administered rather than a true problem with the methodology.

Whether a survey or questionnaire is intended to provide descriptive data or to test hypotheses, the central issues for test designers are: does this questionnaire ask questions that are both valid and reliable? Has it been piloted on sufficient samples of the target population? Can the data be scored and suitably analysed? In this article I shall look at some of the principles of designing a questionnaire and discuss briefly some of the strengths and weaknesses of using questionnaires in paediatrics.

Designing a questionnaire

The principal considerations before embarking on any questionnaire development are:

- (1) Does a suitable questionnaire already exist?
- (2) Who will complete the questionnaire?
- (3) What response format will be used?
- (4) Are questions brief, relevant, and unambiguous?

(1) DOES A SUITABLE QUESTIONNAIRE ALREADY EXIST?

Surveying the literature in the quest for an off the shelf questionnaire can be daunting. The past decade has seen a proliferation of scales purporting to measure just about everything. Nevertheless, the investment of time at this stage is valuable. It is probably better to use a well validated, standardised measure that seems more or less up to the job than to add yet another naively constructed measure to the existing confusing pool. There is no compelling reason why *ad hoc* questionnaire items designed along standard lines examining the treatment condition or problem being evaluated should not be adapted from existing measures or used alongside one. However, this must be done with care, ensuring that the psychometric properties of the original instrument are preserved.¹ Apart from conducting Medline searches, other useful sources include books by McDowell and Newell² for general health measurement scales and Orvaschel and Walsh³ for scales suitable in epidemiological and clinical research in paediatrics. One of the problems with everyone constructing their own instrument is that it becomes rather difficult to pool data for meta-analysis or to compare outcomes across trials or disease states. Furthermore, most people have received little formal education about questionnaire design and assume that it is a simple task. Understanding more about the difficulties also allows one to appreciate better the strengths and weaknesses of established measures.

(2) WHO WILL COMPLETE THE QUESTIONNAIRE?

The question as to who should complete the questionnaire is particularly salient in paediatrics. Very young children may lack the appropriate constructs supposedly being tapped by various probes and some lack the ability to articulate their subjective distress. Consequently, many researchers rely on proxy

CRC Communication and Counselling Research Centre, Department of Oncology, UCL Medical School, 3rd Floor, Bland Sutton Institute, 48 Riding House Street, London W1P 7PL

Correspondence to:
Dr Fallowfield.

judgments made by patients, care givers, or interviewers. Unfortunately, when determining those aspects of well being that should be included in questionnaires, the perspectives of parents may differ markedly from that of health care professionals. For example, Cadman *et al* examined the relative values placed on different domains contributing to quality of life by professionals and parents of children with developmental impairments.⁴ There was poor correlation on each of the eight dimensions of prognosis, self care, mobility, communication, cognition, family relationship, mood, and behaviour. Issues such as the developmental appropriateness of different questions require considerable thought. However, there is evidence that the competence of children to rate their own physical and emotional well being is often underestimated.⁵

If questionnaires are designed well, children may give reliable responses. Pantell and Lewis have documented some of the potential technical and structural pitfalls when assessing the quality of life of children with self report measures.⁶ These include basic misinterpretations of language, difficulties with negatively worded questions, and position bias of questions or response options. Sometimes pictorial scales need to be used with younger children.⁷

Even if the questionnaire is designed to assess behaviour, attitude or emotions of adults, that is, parents or carers, a basic tenet is that the readability should not tax the skills of the average 12 year old. The literacy levels of the population are far lower than many realise and too many test designers pilot their questionnaires on professional colleagues or friends who may have literacy levels way beyond the target population.

(3) WHAT RESPONSE FORMAT WILL BE USED?

Categorical scales

Provided questions are clear and unambiguous, simple dichotomous, categorical judgments, for example those requiring yes/no or true/false responses are appealing. They are also easy to score. However, there are disadvantages, especially if the continuous nature of responses is ignored. Some people get frustrated if they are restricted to dichotomous variables and limiting responses can lead to information loss. Examples of categorical scales include the following:

- (i) Which of the following do you feel?

Tired	(tick one)
Dizzy	
Sad	
Lonely	
- (ii) Do your friends visit you?

Yes/no

- (iii) My friends have stopped visiting

True/false

Dichotomous variables obviously produce nominal data analysed with non-parametric statistical tests. Unfortunately, some questionnaires ask for dichotomous categorical responses when the symptom, attitude, or behaviour almost certainly lies on a con-

tinuum. Consider, for example, the question: 'Does your asthma stop you playing?' Categorical response of 'Yes' or 'No' would ignore just how much trouble the child was experiencing, and this could be assessed best with a continuous response format. Presumably one would wish to know how often the asthma prevented play.

Response formats which permit continuous judgments include adjectival, Likert, visual analogue, semantic differential, and Guttman scales:

Adjectival scales

Since my illness, friends (tick one)

Never visit
Rarely visit
Visit same as usual
Often visit
Visit more than ever

*Likert scales*⁸

Since my illness, friends do not like me: (tick one)

Strongly disagree	(tick one)
Disagree	
Neither agree nor disagree	
Agree	
Strongly agree	

Visual analogue scales

These scales are very popular for use in questionnaires. They usually consist of 10 cm lines, the ends of which represent extremes of the item being measured. People are asked to mark along the line where they think they lie between the two extremes. Although some respondents require some time to understand the concept, it has been used successfully with patients as young as 5 years.⁹

How severe has your pain been this week?

Pain as severe as it could be [] No pain at all

Semantic differential scales

My skin:

Worries me	[]	Doesn't bother me
Is embarrassing	[]	Is not embarrassing
Looks awful	[]	Looks great

A final, more complicated method for continuous judgments is Guttman scaling. Several items relating to the attitude, behaviour, or function in question are ranked tentatively in some hierarchy of best to worst, more or less, etc. After complex testing among subjects, certain items may be deleted or reordered. This response format is not recommended unless considerable resources are available for conducting reliability and validity checks.

Guttman scale

My child: (tick one)

Is confined to bed
Can walk with assistance across the room

Can walk unaided across the room
 Can climb the stairs
 Can walk less than a mile
 Can walk more than a mile

The various response modes have subtle advantages (see Streiner and Norman¹⁰ for an excellent discussion). Questionnaire developers should always conduct a small pilot study on the target population, using different formats. Having decided on one, it is best to stick with that format throughout. Changing response formats confuses some people and affects reliability of the data.

Another important issue to resolve is how many boxes or steps should be used. Opinions are divided, but between five and seven are probably optimal. Having an even number of categories demands that respondents commit themselves to one side or other of the continuum, but researchers may wish to permit neutral responses using an odd number. As far as children are concerned (and some adults), they do tend to show a position bias, choosing the first answer among response options. Swapping negative and positive positions of categories for different items forces the respondent to attend carefully (and the researcher doing the scoring), but this can sometimes confuse children. One other problem with many rating scales is that some people tend to exhibit what is called a halo effect, with their responses showing a strong positive bias. Therefore, it is important to encourage people filling in questionnaires with categorical response formats to use the extremes of the continuum. Discussions about the assumptions that rating scales produce interval data are beyond the scope of this article, but these issues are important considerations for appropriate statistical analysis.

(4) ARE QUESTIONS BRIEF, RELEVANT, AND UNAMBIGUOUS?

The diligent researcher must always make some compromises. Pragmatic considerations usually demand the sacrifice of comprehensive coverage of an area. Cooperations of patients might well decline if there are too many questionnaires or items included. It is axiomatic that quality of data is compromised as the length of time required to complete tests increases. This is even more likely to be true if multiple measures are made at frequent intervals by patients already burdened by disease. Thus, having chosen a list of items deemed necessary, some reduction will be required.

Validity and reliability

Establishing the reliability and validity of questionnaires is an extremely important area, frequently neglected by researchers. Comprehensive definitions of validity and reliability can be found in Fallowfield.¹¹ Briefly, reliability testing asks if a test measures accurately and consistently what it

is meant to be measuring. There are various forms of reliability, but at the very least test-retest reliability should be established. If a sample of people is given the same test on two occasions, then the correlations coefficient of scores can be obtained. Timing is all important in this as in health care situations, as changes in bodily state, disease progression, etc, could well produce an underestimate of reliability.

It is quite possible for a test to be reliable in that it measures *something* reliably, but this tells us little unless we know what that something is. Test validity is therefore more important than reliability.

A questionnaire should contain sufficient items to ensure face and content validity. At its simplest, this means: does the questionnaire appear on the face of it to be asking relevant questions about the topic of interest and are there sufficient domains or areas covered? This is primarily a subjective assessment determined by experts.¹² Criterion validity is necessary if the questionnaire is to be used as a predictive measure. Again, interested readers are advised to look at Streiner and Norman¹⁰ or Fallowfield¹¹ for a more detailed description of this. Often questionnaires are designed by researchers who are interested in finding out answers to specific questions about their practice or treatment and not especially interested in producing a reliable validated test, standardised for use in medicine. Nevertheless, consideration of many of the points already raised is important. Wording of items on questionnaires needs particular care.

Ambiguous questions

One of the most common failings seen in questionnaires that have not had psychometric properties sufficiently tested is ambiguity. Consider, for example, the statement 'I understood the physiotherapist's instructions' with a yes/no response. If the respondent says no, this could mean that he or she never saw a physiotherapist, or did not understand the instructions. Sometimes two questions are asked at the same time; for example 'My joints are stiff and tender'. What if joints are stiff, but not tender, or tender but not stiff? Some people will only tick yes if both are true and data are lost. Another common error of ambiguity concerns interpretation; for example: 'I have not played football recently'. Recently might mean within the past eight hours, a week, a month, or a year. The time frame is particularly important to state explicitly in questionnaires.

Negative wording

The validity coefficients decline for negatively worded items on questionnaires.¹³ It is really best to avoid words with a negative prefix, such as un-, in-, or im- and words such as not, rarely, or never, used in combination totally confuse. For example, 'I am rarely unwell' will

be less readily understood than 'I am usually well'.

Using jargon

The most obvious, innocuous words or phrases can serve to confuse lay populations and is another reason why piloting must be done on the target population not one's friends and colleagues.

Social desirability

It requires considerable skill to frame questions in such a way as to minimise implied value judgments or to prevent patients from giving socially desirable responses. Children may not wish to distress parents or care givers and parents may under-report some problems in their desire to be 'good' parents.

Conclusion

The well designed questionnaire can provide extremely valuable information and is often the only way of obtaining certain types of data. However, it is not a trivial process to design a test and spending sufficient time on the development is imperative if the data are to have any validity. Just as a poorly manufactured or

calibrated medical instrument would have unfortunate consequences for patients, data derived from faulty questionnaires can have a detrimental effect on patient care.

- 1 Guyatt GH, Bombardier C, Tugwell PX. Measuring disease specific quality of life in clinical trials. *Can Med Assoc J* 1986; **134**: 889-95.
- 2 McDowell I, Newell C. *Measuring health: a guide to rating scales and questionnaires*. Oxford: Oxford University Press, 1987.
- 3 Orvaschel H, Walsh G. *The assessment of adaptive functioning in children: a review of existing measures suitable for epidemiological and clinical services research*. Rockville MD: National Institute Mental Health, 1984.
- 4 Cadman D, Goldsmith C, Bashim P. Values, preferences and decisions in the care of children with developmental disabilities. *J Dev Behav Pediatr* 1984; **5**: 60-4.
- 5 Eiser C. Changes in understanding of illness as the child grows. *Arch Dis Child* 1985; **60**: 489-92.
- 6 Pantell RH, Lewis CC. Measuring the impact of medical care on children. *Journal of Chronic Disease* 1987; **40** (suppl 1): 995-1085.
- 7 Harter SL. The perceived competence scale for children. *Child Dev* 1982; **53**: 87-97.
- 8 Likert RA. A technique for the development of attitude scales. *Educational and Psychological Measurement* 1952; **12**: 313-5.
- 9 Scott J, Ansell BM, Huskisson EC. The measurement of pain in juvenile chronic polyarthritis. *Ann Rheum Dis* 1977; **36**: 186.
- 10 Streiner DL, Norman CR. *Health measurement scales: a practical guide to their development and use*. Oxford: Oxford University Press, 1989.
- 11 Fallowfield LJ. *The quality of life: the missing measurement in health care*. London: Souvenir Press, 1990.
- 12 Guilford JP. *Psychometric methods*. New York: McGraw-Hill, 1954.
- 13 Holden RR, Fekken GC, Jackson DN. Structured personality test item characteristics and validity. *Journal of Research in Personality* 1985; **19**: 386-94.