

The Short-Form 36 and older people: some problems encountered when using postal administration

Sara Mallinson

Abstract

Objective—To explore some of the problems encountered in the postal administration of the Short-Form 36 (SF-36). Questions that seem to present particular difficulties for the group are identified. In addition some of the written comments from the questionnaires and people's responses to questions on how difficult they found the SF-36 are discussed. **Design**—The study group were asked to complete a health questionnaire containing the SF-36 on three separate occasions (at zero, three, and six months). The first and final questionnaires were interviewer administered during a face to face interview. A shorter questionnaire containing only the SF-36 and another health status measure was sent by post to each patient in the interim.

Participants—People aged 65 years or above who were new referrals to community based occupational therapy or physiotherapy services in three areas in north west England.

Main results—Response and completion rates for the postal questionnaire were lower than expected, even though all the patients had already had a face to face interview and had therefore completed the SF-36 once. Only 34 of 56 respondents (60.7%) completed all the items on the SF-36.

Conclusions—All those planning to use the SF-36 (and similar measures) with older populations should be sensitive to the problems of postal administration. Non-return of questionnaires, high levels of missing data on those that are received, and ambiguities in response may mean that other measures, or perhaps alternative research methods, are more appropriate.

(J Epidemiol Community Health 1998;52:324-328)

they are generally cheap (compared with time intensive methods such as in depth interviewing), they typically have a quick turn around, and they are usually designed to produce quantitative data that are easy to summarise.⁵ The proliferation of health status measures reflects a particular interest in producing standardised data that can be subjected to statistical tests that give an indication of validity, reliability, and responsiveness. Amenability to this process confers credibility on the process of obtaining "subjective" assessments that appeases those familiar with more "objective" measures.⁶ It also presents the prospect of being able to generate comparable data sets across a whole range of different groups and that appeals to those with responsibility for decision making and health care management. It is this decision making imperative that creates the demand for rapid, clear, and easily digestible information to inform choices and therefore propels the search for the ideal multi-purpose health status measure.⁷

The Short-Form 36 Health Status Questionnaire (SF-36) is a generic self report health questionnaire whose 36 items are added together to form eight health dimensions.⁸ It can be administered by post or in an interview and investigations in the UK, and previous work in the US, indicate that it has good reliability and validity as a health status measure.⁹⁻¹³ With this weight of evidence it has become one of the most widely used health status measures.

Although postal administration of health surveys may be less expensive than using interviewer administration, those responsible for designing a research project need to be satisfied that response rates will be reasonable and that people who receive questionnaires will be able to comprehend them and provide answers. Some concern has been expressed about levels of non-response and missing data when the SF-36 has been used with older people. Brazier *et al* pointed out that levels of missing data in their general practice sample increased significantly with age so that the 65-74 year old respondents had almost twice as many missing items as people aged 55-64.⁹ Parker *et al* reported very low response rates (only 50%) when the measure was given to medical and surgical inpatients for self completion.¹⁴ In a study of older patients undergoing hip replacement surgery Dawson *et al* report that patients in their study who were sent the SF-36 by post were significantly less likely to return fully completed questionnaires and furthermore the

Involving people in evaluating the outcomes of particular treatments, packages of care, or services is acknowledged to be an important and yet challenging feature of the emergent evaluative culture within NHS and other public sector services.¹⁻⁴ Although there are numerous ways in which self assessment could be pursued, current research practice has concentrated on the development of social survey methodology for the assessment and quantification of the "subjective" aspects of health and illness. Survey methods are attractive because

Public Health
Research and
Resource Centre,
University of Salford,
Humphrey Booth
House, Hulme Place,
Salford M5 4NY

Correspondence to:
Sara Mallinson.

Accepted for publication
26 August 1997

SF-36 was more likely to have missing data than a condition specific measure (the Oxford Hip Score) that was sent at the same time.¹⁵ In a recent study of test-retest reliability in the US Andresen *et al* had a basic response rate of 60.1%, which was further reduced by missing items.¹⁶ Hayes *et al* compared different administration methods for the SF-36 with a sample of 195 older people recruited from a hospital outpatient department and general practice.¹⁷ They found that when the survey was self completed 42% (42 of 98) of people could not complete the form and of the 56 people who did complete it 34 (61%) missed one or more of the items. All those who could not self complete were offered an interview therefore the majority of their sample had an interviewer administered questionnaire (139 of 195). As might be expected the levels of missing data dropped significantly to just 12% (16 of 139)—although this still implies that people found some of the questions difficult to answer.

Interestingly, when patients were asked to evaluate the questionnaire 91% (177 of 195) of people said the questions in the SF-36 were all or mostly clear and easy to understand, and 88% (172) thought that all or most of the questions were relevant and applicable. From the response rates a less positive evaluation might have been predicted. Hayes *et al* concluded that the SF-36 was acceptable, quick to use, and potentially useful, although some items relating to “work” needed rephrasing. Their positive outlook may encourage others to use the measure despite the problems they recorded. Others are less positive and have suggested that its content may be both “healthist” and “ageist” and therefore inappropriate for use with older people, particularly those with health problems, because it is unable to reflect small changes in health that are often a very positive experience for the individual concerned.¹⁸⁻²⁰ Over the past two years the study reported in this paper has been exploring the issues of interpretation and meaning in the SF-36 and as part of that work the SF-36 was sent by post to a small sample of people aged 65 and above. The research design differs from that of the study done by Hayes *et al*¹⁷ but the data allow some comparison and further exploration of the problems of using the SF-36 with older people.

Methods

The research was conducted with the assistance of two teams of community physiotherapists and one team of community rehabilitation occupational therapists. They asked new referrals to their service who were aged 65 years or more if they would take part in the study. The only patient characteristic specified was that everyone should be able to self complete a questionnaire. The sample therefore includes people with a range of health problems who had a range of interventions. The purpose of the research was not to evaluate the effectiveness of each intervention but to mimic routine evaluation and explore patients' views on whether the package of care they received had made any impact on their health problem.

Patients were interviewed face to face on two occasions using a structured questionnaire: at zero months (a baseline assessment); and at six months (final assessment). In the interim (three months) a shortened version of the same questionnaire was sent by post to see how well people fared when the support (and pressure?) of an interviewer was not available. This paper will describe the problems encountered with the postal survey in terms of response rates and levels of missing data to inform others considering work of a similar nature.

Results

THE SAMPLE

Initially the sample comprised 71 people whose mean age was 76.8 years (range 65–89). During the course of the project nine of 71 (12.6%) died, four of 71 (5.6%) became too ill to continue with the research, and two of 71 (2.8%) asked to be withdrawn. As a result the total number of people who remained with the study for the full six months was 56 (78.9% of the original sample). This smaller group had a mean age of 77.1 years (range 65–89). Forty four of the sample were women and 12 were men. The remainder of this discussion will focus on this sample of 56 who completed both face to face interviews (zero and six months) and received the postal questionnaire (three months).

RESPONSE RATES

The questionnaire was mailed with a postage paid addressed envelope and was followed up after two weeks with a reminder letter offering to send another questionnaire, and a further letter after four weeks. Before receiving the postal questionnaire the group had already spent approximately 45 minutes being interviewed about their health and had completed the SF-36 health status questionnaire. They were therefore familiar with the researcher and the questions being asked, they had been informed about the purpose of the repeated administration of the questionnaire, and knew they would be contacted again after the postal survey. Because of these factors we predicted that the response rate would be good—higher than might be expected from a “cold-call” survey for example. In reality the response rate for the interim postal assessment was lower than expected. Firstly, the basic response rate in terms of returned questionnaires was 80% (45 of 56). Secondly, if we then take into account the standard of completion of the questionnaires that were returned the response rate drops further. The SF-36, which formed the main part of the postal questionnaire, was correctly and fully completed by only 34 of the 45 people who returned the questionnaire, thus lowering the overall response rate to a disappointing 60.7% (34 of 56).

Obviously this group encountered some significant problems when trying to answer the questionnaire without the aid of an interviewer. In the section below some of the items that were missed regularly are described.

Table 1 Problem questions from the SF-36: 4A-D (role limitation: physical) and 5A-C (role limitation: emotional)

4 During the past four weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?	yes or no
(a) Cut down on the amount of time you spent on work or other activities	
(b) Accomplished less than you would like.	
(c) Were limited in the kind of work or other activities	
(d) Had difficulty performing work or other activities (eg it took more effort)	
5 During the past four weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?	yes or no
(a) Cut down on the amount of time you spent on work or other activities	
(b) Accomplished less than you would like	
(c) Didn't do work as carefully as usual	

PROBLEM QUESTIONS ON THE SF-36

As with the study by Hayes *et al.*¹⁷ a large proportion of the difficulties centred on items that form the two role limitation dimensions (table 1). Five of the 45 people who returned the postal questionnaire missed out all of the items in question 4 and question 5 and a further three people missed two or more items from either question 4 or 5. While most people had obviously managed to complete these questions the frequency with which the questions are missed (eight of 45 or 18% of the respondents) suggests that there is a commonly experienced problem. Jenkinson has suggested that the wording of the initial rubric to the items in both question 4 and 5 may not make sense.²¹ It suggests that the questions will be phrased as problems but when you read them they are not. For example question 5C would make much more sense if it were phrased "I didn't do work as carefully as usual". Hayes *et al.* suggest that it is the reference to "work" that may cause confusion as people who are of retirement age may think the item does not apply to them.¹⁷

Another problematic set of items are those in question 10, which forms the main part of the General Health Perception dimension (table 2). Three people had omitted questions 10A, 10B, and 10C, and a further two people did not respond to 10B and 10D. Hayes *et al.* found question 10C was the most frequently missed item but in this study question 10B is missed most often—although all the questions in this dimension were missed by someone.

Questions that are used to create the mental health and energy/fatigue dimensions are formatted into a complex question matrix (usually numbered 9A-J). This means that 10 items are listed down one side of the page and the six response options—All of the time / Most of the time / A good bit of the time / Some of the time / A little of the time / None of the time—are listed across the top. Overall six people left out at least one of the items in the matrix, most often question 9H, which asks—Are you a happy person? Three people did not

Table 2 Problem questions from the SF-36: 10A-D (general health perceptions)

10 Please choose the answer that best describes how true or false each of the following statements is for you.	Definitely true / Mostly true / Not sure / Mostly false / Definitely false
(a) I seem to get ill more easily than other people	
(b) I am as healthy as anybody I know	
(c) I expect my health to get worse	
(d) My health is excellent	

KEY POINTS

- Postal administration of the SF-36 to an elderly patient group resulted in high levels of non-response or missing data.
- Missing data affected almost all of the health dimensions of the SF-36.
- Comments written on the questionnaires by respondents highlight problem areas that require future investigation.

answer this question. Of the six people with items missing three had missed out four or more of the 10 questions.

Interestingly items 3A-J (physical functioning dimension), which on the face of it seem to present people with quite straightforward statements of activity levels, also had their share of missing data. They cover a range of activities including vigorous activity, climbing stairs, walking 100 yards, kneeling, and bathing/dressing. People can choose one of three responses "yes, limited a lot, yes, limited a little, no, not limited". Two people did not respond to three or more of the questions and a further three people left out one item. It has been suggested that question 3A, which asks about "vigorous activities, such as running, lifting heavy objects, participating in strenuous sports" is particularly irrelevant to older people and is therefore prone to missing values.¹⁷ In this study this has not been the case. All respondents answered question 3A and missing data were spread across the other nine items. It is possible that people have difficulty relating to the activities specified in the items of the SF-36 items, or it may be the format of the questions (which are again presented as a matrix), which people find difficult.

Other questions were missed but not with the same frequency. For example question 8 "how much did pain interfere with your normal work (including work both outside the home and housework)?" was left out by two people. In addition one of the items that form the social functioning dimension (question 6) was left blank by a respondent, and question two on general health change was left out by a respondent.

QUESTIONNAIRE EVALUATION

At the end of the questionnaire a few simple questions were included to try and find out how difficult people thought the questionnaire was. Thirteen of 45 (28%) said they had no problems filling it but 21 of 45 (47%) said they had problems answering some questions and a further 11 of 45 (24%) said they found all of the questions difficult. Most of the respondents therefore had problems answering some or all of the questionnaire (32 of 45, 71%). This contrasts quite sharply with the overwhelmingly positive evaluations made by respondents in the Hayes study.

The questionnaire is designed for self completion but over two thirds of the respondents (29 of 45, 64%) said they had assistance completing the questionnaire from a relative or

friend. The need for such a high rate of assistance makes the whole process of postal administration seem uncertain.

WRITTEN COMMENTS ON THE QUESTIONNAIRES

Although not all of the respondents who left items missing wrote an explanation on the questionnaire, some people did add unsolicited comments or queries to the form. These are more than just “interesting reading”. They give us some clues to the type of difficulties people experience with the design and content of the SF-36. They are quite simplistic as we did not ask people to write on their questionnaires and did not provide extra space or encouragement for them to do so. Nevertheless, in taking account of them we can draw attention to issues that all potential users of the SF-36 need to be aware of. The comments of the respondents can be divided into three general areas.

Relevance and experience

One person did not answer any of the role limitation items in 4A-5C (see table 1) and wrote on the questionnaire “I am retired—don’t work”. Another person left out question 4D only and had written “don’t work”. The issue of “work” was also raised in relation to question 8 (how much did pain interfere with your normal work (including work both outside the home and housework?)) A respondent wrote by question 8 “I am elderly and not able to work.”

Even though the items in questions 4, 5, and 8 all try to use a wide definition of work and mention “other activities” or “housework” people still obviously read them as irrelevant to their situation. This supports evidence from the Hayes study.

Related to this idea of relevance are some of the written comments about the difference between being able to do something and actually doing it. People had difficulty deciding how to code their responses if the activity specified was one they had not tried recently. One woman reacted to the idea of housework included in several items by writing “home help does housework.” Another left out several parts of the physical functioning dimension with the explanation that she “hadn’t attempted”. Another made a similar comment of “cannot do or try” and left some items blank.

Misunderstanding

Misunderstanding of questions is hard to estimate from a postal survey and requires further research. However, one person did give an indication that something might be amiss with the role limitation: emotional dimension (table 1). She wrote by the side of the question 5A-C “only cannot do work because of hip—physical health”. She actually then went on to answer yes to all three items. This apparent divergence between the qualitative comment and the quantitative response has to be treated as an indicator of ambiguities in the data.

Face validity and formatting

Surprisingly for a questionnaire with such a long pedigree there seem to be quite basic

problems with the content of some items. These face validity issues make it difficult to interpret how people are answering some questions particularly in the physical functioning dimension where several items have “double barrelled” questions. That means they ask more than one question within a single item. Not everyone wrote comments but three people did point out some basic difficulties.

One woman left question 3J blank (3J: Does your health limit you in these activities? If so how much: bathing and dressing yourself) commenting that she could shower but not use the bath. Another had problems with 3J because she could dress herself but could not attempt to bathe. In response to question 3F (3F: Does your health limit you in these activities? If so how much: Bending, kneeling or stooping.) One person wrote that they could bend and stoop but not kneel. The latter two respondents both chose “limited a little” from the response options, presumably to reflect their dilemma of being able to do one thing but not another. However, once the score is coded and entered these nuances are lost. Asking people to make coding decisions of this kind should really be avoided and is a surprising flaw in a measure with such a long history of development and testing.

A potentially problematic confusion can be seen in the final questions of the SF-36. Two people had problems with 10 B (I am as healthy as anyone I know). One wrote in “for my age” but did not select a response. Another wrote “yes in my age group but I know a lot of young people” and chose “not sure”.

Several of the respondents had help in completing the questionnaire from a friend, partner or relative. Two of these helpers commented on the difficulty of some of the questions, particularly those that were in a matrix format (that is, the physical functioning dimension, and the mental health/energy and fatigue dimension). Although some time had been spent ensuring that the questions were set out as clearly as possible with large font, and boxed numbers for the person to circle people still found it hard to use. Several people had made a number of changes to their responses, which could be interpreted as a reflection of their difficulty using a matrix. One respondent who had completed the schedule alone wrote that the mental health/energy fatigue matrix made her “dizzy” and that she found it hard to find the right number. She also added that she didn’t like the numbers and would have preferred some “straight talking”.

Discussion

There are still uncertainties about the usefulness of the SF-36 with older populations, particularly when it is administered by post. The savings in time and staffing costs may have to be offset by disappointing response and completion rates. The comments written by respondents on the postal questionnaire help to highlight some of the challenges people faced in trying to interpret and answer it. Further work is required to explore these problems in more detail.

The author would like to thank the patients and therapists who took part in the study, Jean Siddall for her time and support, and the anonymous referees for their comments on an earlier draft of this paper.

Funding: this study is funded by The North West Regional Health Authority Research and Development Directorate as part of their Health Services Research Training Fellowship scheme.

Conflicts of interest: none.

- 1 Popay J, Williams G. Public health research and lay knowledge. *Soc Sci Med* 1996;42:759-68.
- 2 NHS Executive. *Patient partnership: building a collaborative strategy*. Leeds: NHS Executive, 1996.
- 3 Fitzpatrick R, White D. Public participation in the evaluation of health care. *Health Social Care in Community* 1997;5:3-8.
- 4 Bowling A. *Measuring health*. Milton Keynes: Open University Press, 1991.
- 5 Wilkin D, Hallam L, Doggett M. *Measures of need and outcome for primary care*. Oxford: Oxford University Press, 1992.
- 6 Donovan JL, Frankel SJ, Eyels JD. Assessing the need for health status measures. *J Epidemiol Community Health* 1993;47:158-62.
- 7 Long AF. Assessing health and social outcomes. In: Popay J, Williams G. *Researching the peoples' health*. London: Routledge, 1994:157-82.
- 8 Interstudy *SF-36 Health Status Questionnaire Users Manual*. Minneapolis: Quality Quest, 1989.
- 9 Brazier JE, Harper R, Jones NMB, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992;305:160-4.
- 10 Garrett AM, Ruta DA, Abdalla MI, et al. The SF-36 health survey questionnaire: an outcome measure suitable for use within the NHS? *BMJ* 1993;306:1440-4.
- 11 Jenkinson C, Coulter A, Wright L. Short-Form 36 Health Survey questionnaire: normative data for adults of working age. *BMJ* 1993;306:1437-40.
- 12 McHorney C A, Ware J E, Rogers W, et al. The validity and relative precision of MOS short and long form health status scales and Dartmouth Coop charts. *Med Care* 1992;30 (suppl):253-65.
- 13 Ware JE, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36) 1. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
- 14 Parker SG, Peet SM, Jagger C, et al. Quality of life in older patients: the SF-36 in practice. [Abstract]. *Age Ageing* 1993;23 (suppl 4):27.
- 15 Dawson J, Fitzpatrick R, Murray D, et al. Comparison of measure to assess outcomes in total hip replacement surgery. *Quality in Health Care* 1996;5:81-8.
- 16 Andresen EM, Bowley N, Rothenberg BM, et al. Test-retest performance of a mailed version of the medical outcomes study 36 item short form health survey among older adults. *Med Care* 1996;34:1165-70.
- 17 Hayes V, Morris J, Wolfe C, et al. The SF-36 health survey questionnaire: Is it suitable for use with older adults? *Age Ageing* 1995;24:120-5.
- 18 Hill S, Harries U, Popay J. Is the SF-36 suitable for routine health outcomes assessment in health care for older people? Evidence from preliminary work in community based health services. *J Epidemiol Community Health* 1996;50:94-8.
- 19 Lyons RA, Perry HW, Littlepage BNC. Evidence for the validity of the short-form questionnaire in an elderly population. *Age Ageing* 1994;23:182-4.
- 20 Johanssen A. SF-36 may reinforce ageism. [Letter]. *BMJ* 1993;307:127.
- 21 Jenkinson C. Evaluating the efficacy of medical treatment: possibilities and limitations. *Soc Sci Med* 1995;41:1395-401.