Is the short form 36 (SF-36) suitable for routine health outcomes assessment in health care for older people? Evidence from preliminary work in community based health services in England

Sara Hill, Ursula Harries, Jennie Popay

Abstract

Study objective – To examine the short form 36 (SF-36) health status measure when used to assess older people's views of the outcome of community based health care.

Design – Completion of a structured questionnaire, before and after intervention alongside in-depth interviews with a subsample of the interviewees.

Setting – Community based continence and mental health services in two health authorities in the North West Health Region.

Patients – Forty seven older people newly referred to mental health services or continence services between December 1992 and April 1993 participated.

Main results – The SF-36 showed minimal change over time for both patient groups. The more detailed in-depth interviews showed that people experienced positive changes and derived value from contact with services in a number of important ways. For many reasons these benefits were not reflected in their SF-36 scores.

Conclusions – The SF-36 is not likely to be the "measure of choice" for this type of evaluation, particularly where it involves older patient groups who have high levels of comorbidity. The content of the SF-36 and its lack of detail for individual assessment of change means it masks rather than illuminates patients' views of outcome.

(*J Epidemiol Community Health* 1996;50:94-98)

Public Health
Research and
Resource Centre,
University of Salford,
Davenport House,
4th Floor,
Hulme Place,
The Crescent,
Salford M5 4QA
S Hill
U Harries
J Popay

Correspondence to:
S Hill.

Accepted for publication
July 1995

It is widely accepted that collecting patients' views on the outcome of their health care is a vital component in health care monitoring and assessment.¹ Identifying practical ways of capturing those views on a routine basis is problematic.² Most of the work in the field of patient assessed outcomes has concentrated on the development of standardised, self report health measures.³ In particular, generic health status measures have stimulated interest because they cover a number of health concepts such as psychological and emotional well being, physical health, and social functioning, which

should be relevant to patients across a whole range of services and interventions.⁴ One particular measure, the short form 36 (SF-36) is emerging as the current favourite, and is being used in a variety of settings.⁵ Despite its popularity, however, there seem to us to be two particular concerns that remain unanswered at present.

- There is a danger that the current drive towards an "evaluative culture" within the NHS, which we applaud, will result in measures which have been proved to be valid for use in one setting being transferred to other settings for which they may not be appropriate.⁶ That validity is a function of the context and the purpose, and not of an instrument in isolation is rarely discussed.⁷

- Although the SF-36 may be a valid health status measure in certain contexts, there is as yet little evidence from UK studies that it is responsive enough to detect changes in health status over time – obviously an important characteristic of an "outcome" measure.⁸

This paper sets out to shed light on these concerns by reporting findings from an investigation in which the SF-36 was used to assess changes experienced by older people referred to either community mental health services or continence services. There are two key points to make about the objectives of this study and the resulting design. Firstly, we have not attempted to assess the outcomes of individual interventions or care processes because the research was fundamentally a methodological investigation. Its main aim was to examine *different ways of obtaining people's views* about the *package of care* provided by the service and the *overall outcome* at a specified time point. Secondly, within this overarching aim our objective was to discover to what extent the SF-36 accurately summarised patients' views of change which had occurred after referral and treatment by comparing it at a group and individual level with detailed qualitative evaluations. An intensive method of this sort does not merit a large sample usually associated with statistically driven validation exercises, therefore our sample is small.

The SF-36 is a brief, generic self report questionnaire whose 36 questions are added together to form eight health dimensions (see

Table 1 Number of items in each SF-36 dimension

SF-36 dimensions	No of items	Response options
Functional status:		
Physical functioning	10	3 options
Social functioning	2	5 and 6 options
Role functioning – physical	4	2 options
Role functioning – emotional	3	2 options
Wellbeing:		
Mental health	5	6 options
Pain	2	5 and 6 options
Energy/vitality	4	6 options
Overall evaluation of health:		
General health perception	5	5 options

table 1). Investigations in the UK, and previous work in the US, indicate that it has good reliability and validity as a health status measure.^{9–11} These conclusions are based on its ability to distinguish between different “patient” and population groups, psychometric criteria for validity and internal consistency, and good response rates. With this weight of evidence behind it, the SF-36 is rapidly becoming the most widely used generic measure of health status and outcome in the UK.⁵

Our main concern is that the SF-36 is being used for purposes to which it is not suited because of its high profile. Much of the published information refers to its use in population health status assessment, although more information about its use in particular clinical conditions is becoming available.¹³ There is as yet little to guide people involved in local research, service audit, or other smaller scale investigations who wish to include the “patients’ perspective” in their evaluations. The SF-36 has received such a positive appraisal in the published reports that it stands out as “the best measure” and researchers and auditors may well feel compelled to use it when alternatives seem few and far between. Nevertheless, a measure cannot be adopted for all purposes without further investigation and debate. It is important to assess whether measures and measurement techniques which were originally designed for large group or general population monitoring are useful in populations such as the elderly or the young, and whether they are suitable for more more individualised assessments.^{14–16}

Subjects and methods

This paper reports the findings of a study of patients newly referred to continence services and mental health services in two separate health districts during a four month period. Patients’ names were passed to the research team after their first contact with the service provider where their suitability for the project was assessed according to agreed criteria. All patients had to be undergoing a new episode of treatment either with a continence advisor or a consultant psychiatrist/community psychiatric nurse. With regard to the mental health service only patients with a mini-mental state score of 20 or above were eligible for interview.¹⁷ Patients from both service groups were interviewed using a questionnaire which contained the SF-36, condition-specific questions, a psychological wellbeing measure and open-ended questions about expectations, other health problems, social factors, and recent life events.

Respondents were interviewed within a week of their referral and again after three months. Shortly after their second structured interview, a subsample of the interviewees participated in in-depth interviews. Overall we interviewed 18 patients referred to the community continence service in one district, and 29 patients referred to the community mental health service in another district. Nine patients from each district took part in both structured and semi-structured interviews. Most respondents were aged between 75 and 85 years and women outnumbered men in the sample by almost 3:1.

Data from the SF-36 were analysed using SPSS/PC. Frequency of individual SF-36 items and combined group scores were examined for both services. This indicated that dimension scores were not normally distributed, and for this reason and because of our small sample sizes, median and range provide a more useful group summary than means and standard deviations. A non-parametric test (Wilcoxon rank test) was used to test for statistically significant differences in dimension scores between first and second interview.¹⁸

The in-depth interviews were taped and transcribed and the transcripts were then indexed and analysed using a qualitative method.¹⁹ The in-depth interview data are not quantifiable and their detailed nature means that case studies and thematic overviews are the most satisfactory way of presenting the findings.

Results

Table 2 summarises SF-36 dimension data from the first and second interviews. The difference between the first and second column of figures in each service group is the amount of change recorded by the SF-36. In some instances, the median score was zero (indicating worst possible health on the SF-36) at both first and second interview. This was the case for the “role limitation: physical” dimension for both service groups, and for the “role limitation: emotional” item for the mental health group. In all instances, however, the

Table 2 Medians and ranges for SF-36 dimension scores (range indicated in brackets below median score)

SF-36 dimensions	Continence (18)		Mental health (29)	
	1st interview	2nd interview	1st interview	2nd interview
Physical functioning	0 (0–100)	5 (0–100)	40 (0–100)	50 (0–100)
Social functioning	11.11 (0–100)	11.11 (0–100)	44.44 (0–100)	33.33 (0–100)
Role limitation: physical	0 (0–100)	0 (0–100)	0 (0–100)	0 (0–100)
Role limitation: emotional	33.33 (0–100)	0 (0–100)	0 (0–100)	0 (0–100)
Mental health	52 (8–96)	48 (0–92)	36 (0–64)	40 (4–76)
Energy/vitality	15 (0–95)	15 (0–85)	25 (0–85)	35 (0–80)
Pain	44.44 (11.11–100)	44.44 (11.11–100)	33.33 (11.11–100)	66.67 (11.11–100)
General health perception	37 (0–100)	20 (0–100)	60 (0–87)	45 (0–87)

NB: 100 = best health and 0 = worst health. Medians were calculated as the score at which the cumulative percentage of respondents was 50% or more.

range of scores was extremely large. In many cases the range encompassed both minimum and maximum achievable dimension scores, indicating great variability within the groups.

Patients referred to the continence service tend to have lower median scores than the mental health service patients on a number of dimensions, particularly physical functioning, social functioning, and general health perception. The mental health service patients have lower mental health scores than the continence service patients. This indicates that the SF-36 did distinguish differences between the two patient groups.

OUTCOMES: THE QUANTITATIVE EVIDENCE

All the patients had received a package of treatment, care, and advice during the three months between their first and second interview. However, contrary to what one might hope, the aggregated SF-36 data in table 2 suggest any subsequent changes in health status were at best very small and occasionally negative (for example general health perception for both groups). Table 3 gives a more detailed breakdown of how individuals changed over time on each dimension score. It shows that almost as many people increased their dimension scores over the three month study period as decreased them. On a number of dimensions the tendency is for no change at all (role limitation dimensions in particular). Only the change on the pain dimension of the mental health group reached statistical significance at the 0.05 level.

The SF-36 therefore recorded no marked changes in either a positive or negative direction for these patients. Interpretation of these data alone may lead us to the conclusion that patients have experienced no significant health gain as a result of their referral.

OUTCOMES: THE QUALITATIVE EVIDENCE

When discussing their health in the in-depth interviews people referred to *how they felt* about their day to day lives and their *general mood and outlook* rather than their ability to perform tasks and activities. Their emphasis is different to that of the SF-36 which devotes a considerable number of its items to functional questions. The result is an often striking *difference* between evaluations of change made in the in-depth interviews and those indicated by the quantitative data outlined above.

The main themes which emerged from the in-depth interviews were positive feelings of support, confidence, changed outlook, and reduced sense of burden. All the patients interviewed in-depth reported some positive changes related to their referral.

COMPARING QUALITATIVE AND QUANTITATIVE EVIDENCE

In order to unpick the reasons for the lack of change on the SF-36 dimensions when the in-depth interviews suggest that people have experienced change, detailed comparisons of the qualitative and quantitative data were undertaken. We have selected two individuals who share many characteristics with the rest of the sample as illustrative case studies.

In the first case, an 86 year old woman and her carer talked about their delight at the outcome of referral to the continence advisor:

“You know the most pleasing thing we’ve had is her incontinence improving like that . . . you know you wouldn’t have been able to stand the situation we were in.”

This woman’s SF-36 dimension scores show minimal changes on any dimension but overall her health status was slightly worse by the time of her second interview. This effect is produced by a combination of factors: firstly, she has a number of other serious health problems which are worsening; secondly, the questions which the SF-36 poses (especially the physical functioning questions which begin by asking about vigorous activity and ask only one question about self-care) were largely meaningless to her because of her chronic ill health and were certainly not appropriate to gauge the impact of the pelvic floor exercises she was trained to do. Given her very poor general health and the severity of a number of other health problems, she felt delighted at an apparently small change in continence management which meant she could now “hang on” long enough to get to the toilet. For her and her husband the relief was considerable, and the support invaluable.

The second illustration concerns a 68 year old man in contact with the mental health service. In the in-depth interview he described in detail how, during the depths of his depression, he had been “just like a tramp” without the energy to get out of bed, to wash, or even to eat. By the time of his second structured interview he had begun to go out to his local social club, was feeling increasingly confident, and was, by his own definition, almost “back to normal”. Somewhat surprisingly, the improvements in physical and social wellbeing he described (which appear to relate directly to the dimensions of physical functioning, social functioning, and role limitation) are not reflected in his SF-36 dimension scores. His physical functioning and social functioning dimension scores worsened to a considerable degree (–15 and –55.5 points respectively). His role limitation (physical) score remained at zero. Negative changes on these dimensions cannot be fully explained even with substantial additional information, which can usually elim-

Table 3 Summary of increases and decrease in dimension scores between first and second interviews

SF-36 dimensions	Continence (18)			Mental health (29)		
	+	<>	–	+	<>	–
Physical functioning	7	9	2	14	6	9
Social functioning	4	10	4	12	6	11
Role limitation: physical	1	16	1	5	21	3
Role limitation: emotional	2	13	3	5	20	4
Mental health	10	0	8	15	2	12
Energy/vitality	7	3	8	13	4	12
Pain	5	7	6	14	10	5
General health perception	8	1	9	12	4	13

NB: + =higher scores on the second interview indicating improved health, <>=no change between first and second interview, – =a lower score on the second interview indicating worse health in that dimension.

inate confounding factors. One must assume that he did not understand the questions or that he found them hard to answer because they were not appropriate or meaningful to him. This kind of problem was not uncommon in our sample and raises queries about the nature of the aggregated data. On a more positive note, his mental health scores increased by 25% to 66/100, his energy score rose from 0 to 60, and his role limitation (emotional) score improved 100%. This corresponds with his in-depth interview to a much greater extent, nevertheless, the SF-36 profile as a whole raises more questions than we could answer and was not particularly helpful in establishing this man's perspective on changes in health status.

Discussion

Generic measures like the SF-36 are being used for a range of purposes, some of which seem to be inappropriate. It is vital that potential users of the SF-36 recognise that it may work well in certain settings but it cannot be uncritically transferred to others. Some queries have been raised about its suitability for use with older people but there seems to be no consensus.²⁰⁻²³ Evidence from this study suggests that the SF-36 is not helpful in assessing the outcome of some health care services for older people. The reasons for this appear to be linked to the nature of the services and their objectives, and the characteristics of the patient groups.

There are three characteristics of this context which stand out: the high comorbidity likely in these groups, the low number of patients coming through the service, and wide casemix, and the focus on inappropriate functional tasks for older people.

CASEMIX AND THROUGHPUT IN COMMUNITY BASED SERVICES

Routine assessment of outcomes in community settings is likely to differ from acute interventions in many ways. Firstly, for the two services we selected there was a relatively low throughput of patients, and secondly the variation between the patients – that is, the casemix, was great. Quantitative health status indicators lend themselves to producing summaries of change in relatively large and relatively homogeneous groups. When the SF-36 was used to summarise the health outcome for community continence and community mental health services the small numbers of patients and large variation between them resulted in an indication of “no change” in the group as a whole. However, this was in conflict with evidence obtained through in-depth interviews.

COMORBIDITY

Many patients interviewed in this age group have more than one health problem. This means that we saw many instances of a positive change in the health problem for which people were referred (continence or mental health) which was “masked” by changes in other health

problems. It is increasingly recognised that this “attribution problem” is one of the thorniest issues in outcomes assessment, and this is especially the case with older patients, and with services that do not result in dramatic and sudden changes in health status, but where the emphasis is on support and maintenance.

FUNCTIONAL EMPHASIS

The emphasis of the SF-36, and many other structured health status instruments on functional tasks, such as walking, climbing stairs, carrying out “work”, is inappropriate for some patient and intervention groups. There is a certain degree of “ageism”, or some would argue “healthism”, inherent in these kind of questions. The people we interviewed in-depth clearly expressed that what was important for them in the outcome of their referral to the particular service were feelings of support, increased confidence and improved outlook on life. Changes in what they were able to do in functional terms were often minor – what was important to them was how they felt about themselves and their future.

In circumstances such as those encountered in the incontinence service and the elderly mental health service we should perhaps spend more time assessing what the patients' desired outcomes are, and tailoring suitable outcomes measures to fit the monitoring needs. The SF-36 was not suited to the patient group or the type of routine service setting where we tested it and even in the presence of considerable amount of additional information it was impossible to draw any conclusions about the outcome of care from the patient's perspective. The SF-36 may be useful in some situations (for example, population health status assessment) but this does not mean that it can be used in an unlimited way. Applying it inappropriately is wasteful of resources, and it is also an unacceptable imposition on patients' time and goodwill.

Finally, we would stress that measures like the SF-36 have been developed to give patients a voice and may be the only opportunity that patients have to contribute to the evaluative process. We hope future research will consider ways of collecting patients' assessments of outcomes which truly reflect their views and priorities so that purchasers and providers of care have useful data from which to discuss areas of service development. This may mean taking a second look at the data we get from “off the shelf” health measures and consideration of new approaches to obtaining patients' views of outcome.

The authors would like to thank the King's Fund Centre for funding the research and all the health professionals and patients who took part. We would particularly like to thank Dr JS Bamrah for his assistance during the fieldwork and Sister Carol Hickman. The Public Health Research and Resource Centre is supported by Salford and Trafford Health Authority, Wigan and Bolton Health Authority and The University of Salford.

- 1 Long AF. Assessing health and social outcomes. In: Popay J, Williams G. *Researching the people's health*. London: Routledge, 1994:157–182.
- 2 Donovan JL, Frankel SJ, Eyles JD. Assessing the need for health status measures. *J Epidemiol Community Health* 1993;47:158–62.

- 3 Ware J. Measuring patients' views: the optimum outcome measure. *BMJ* 1993;306:1429-30.
- 4 Patrick DL, Mayo RA. Generic and disease specific measures in assessing health status and quality of life. *Med Care* 1989;27(suppl):s217-s232.
- 5 UK Clearing House on Health Outcomes. *The potential contribution of multidimensional profiles. Outcomes briefing no 2*. Leeds: Nuffield Institute for Health, 1993:16-22.
- 6 Hill S, Harries U. The outcomes process: some reflections from research with people in their 60's and 70's. *Critical Public Health* 1993;4(4):21-8.
- 7 Jenkinson C, Bardsley M, Lawrence K. Measurement in subjective health assessment: themes and prospects. In: Jenkinson C ed. *Measuring health and medical outcomes*. London: UCL Press, 1994:176-185.
- 8 Jenkinson C, Wright L, Coulter A. *Quality of life measurement in health care*. Oxford: Health Services Research Unit, 1992.
- 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 Jenkinson C, Coulter A, Wright L. Short-form 36 health survey questionnaire: normative data for adults of working age. *BMJ* 1993;306:1437-40.
- 11 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.
- 12 UK Clearing House on Health Outcomes. *Using the SF-36 in particular applications/settings. Outcomes Briefing No 4*. Leeds: Nuffield Institute for Health, 1994:14-18.
- 13 Wilkin D, Hallam L, Doggett M. *Measures of need and outcome for primary care*. Oxford: Oxford University Press, 1992.
- 14 McHorney CA, Ware JE, Rogers W, Raczeck A, Lu R. The validity and relative precision of MOS short and long form health status scales and Dartmouth Coop charts. *Med Care* 1992;30(suppl):ms253-ms265.
- 15 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36) 1. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83.
- 16 Folstein MF, Folstein SF, McHugh. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *J Psychiat Res* 1975;12:189-98.
- 17 Bryman A, Cramer D. *Quantitative data analysis for social scientists*. London: Routledge, 1990.
- 18 Social and Community Planning Research. *Individual in-depth interview: notes and papers from a five day course*. London: Qualitative Research Unit, SCPR, 1992.
- 19 Parker SG, Peet SM, Jagger C, Fahren M. Quality of life in older patients: the SF-36 in practice. *Age Ageing* 1993; 23(suppl 4):27(abstract).
- 20 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.
- 21 Singleton N, Turner A. Measuring patients' views of their health: SF-36 is suitable for elderly patients [letter]. *BMJ* 1993;307:126-7.
- 22 Johansson A. SF-36 may reinforce ageism [letter]. *BMJ* 1993;307:127.