THE SHORT FORM-36 HEALTH SURVEY QUESTIONNAIRE IN SPINE SURGERY

MICHAEL GREVITT, RABI KHAZIM, JOHN WEBB, ROBERT MULHOLLAND, JOHN SHEPPERD

From Queen's Medical Centre, Nottingham, England

The Short Form-36 (SF-36) health questionnaire has been put forward as a general measure of outcome in health care and has been evaluated in several recent studies in the UK. We report its use in three groups of patients after spinal operations and have compared it with the Oswestry and Low Back Pain disability scales.

There was a significant correlation between all variables of the SF-36 and the low-back scores. The mental-health items had the weakest correlation. Our study shows that the SF-36 questionnaire is valid and has internal consistency when applied to these patients.

J Bone Joint Surg [Br] 1997;79-B:48-52.
Received 8 August 1995; Accepted after revision 8 May 1996

The introduction of an internal market within the National Health Service has highlighted the need for valid and reliable measures for the assessment of the outcome of medical management. With limited resources and expanding demand, health purchasers are seeking means of comparing efficacy between specialties and individual procedures. There are many measures of outcome within orthopaedic surgery, but most are too specific for application to a general population.

Our aim was to compare a recently introduced health-survey questionnaire, the Short Form-36, with two commonly used rating systems for low back pain, the Oswestry Low Back Pain Questionnaire[1] and the Low Back Outcome Score.[2] We also describe its use in assessing outcome in two groups of patients after fusion of the lumbar spine.

The Short Form-36 (SF-36) Health Survey Questionnaire. The SF-36 is a shortened version of 149 health-related questions which have been validated on more than 22 000 patients as part of a study of medical outcome in the USA.[3]

The questionnaire measures three aspects of health: functional ability, well-being and overall health. These are quantified using eight multi-item variables (Table I); there is an additional unscaled item which relates to changes in the health of the patient over the previous year. For each variable, the item scores are coded, summated and transformed on to a scale from 0 (worst possible health) to 100 (best health). The score from each variable may be linked to form a graphical 'health profile' (Fig. 1) which varies between common medical conditions.[4] The questionnaire can be completed within ten minutes and normal levels have been collected for a UK population.[5]

Table I. The variables in the SF-36 health survey questionnaire

SF-36 variables	Number of items
A. Functional ability	
Physical functioning	10
Social functioning	2
Role limitations attributed to physical problems	4
Role limitations attributed to emotional problems	3
B. Well-being	
Mental health	5
Energy and fatigue	4
Pain	2
C. Overall health	
General health perception	5
Total	35

PATIENTS AND METHODS

Comparison of SF-36 with Oswestry and Low Back Outcome scales. We reviewed 120 patients as part of an audit of automated percutaneous lumbar discectomy (APLD). There were 63 men and 57 women with a mean age of 33 years (17 to 57). All had radicular symptoms due

M. Grevitt, FRCS Orth, Spinal Fellow
R. Khazim, FRCS C, Spinal Fellow
J. Webb, FRCS, Consultant Spinal Surgeon
R. Mulholland, FRCS, Special Professor of Orthopaedic Surgery
Centre for Spinal Studies and Surgery, Queen's Medical Centre, Nottingham NG7 2UH, UK.

J. Shepperd, FRCS, Consultant Orthopaedic Surgeon
Conquest Hospital, The Ridge, St Leonards on Sea, East Sussex TN37 7RD, UK.

Correspondence should be sent to Mr M. Grevitt at 46 Aldershot Road, Fleet, Hampshire GU13 9NT, UK.

0301-620X/97/11269 $2.00

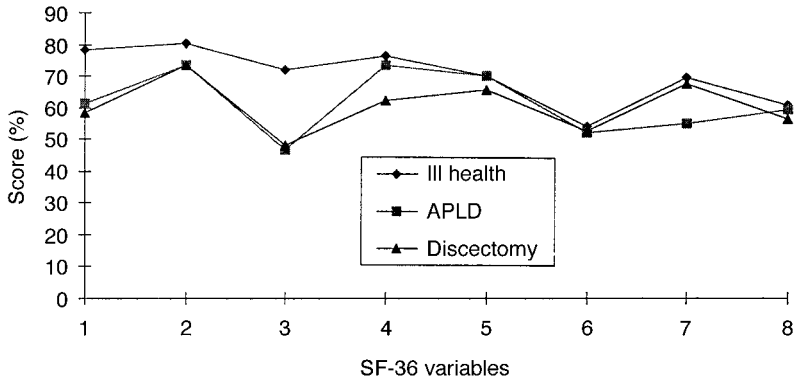

Fig. 1

SF-36 health profiles of patients treated by APLD or open discectomy compared with a population reporting chronic ill health. See Table I for the SF-36 variables.

to a prolapsed lumbar intervertebral disc. In each case conservative treatment for a minimum of three months was prescribed before further investigation was instigated. No patient was involved in a compensation claim at the time of the APLD.

For this study each patient was asked to complete the Oswestry Low Back Pain and the Low Back Outcome Score (LBOS) questionnaires. In addition, an SF-36 questionnaire was completed at the latest review. The results were collated and analysed. In patients with an LBOS we defined the outcome as follows: excellent, 65 to 75 points; good, 50 to 64; fair, 30 to 49; and poor, 0 to 29.[2] The maximum LBOS is 75, but to facilitate comparison with the Oswestry score it was modified to allow a score of between 0 and 100 points.

Statistical evaluation. We examined the correlation of each SF-36 variable with the Oswestry and LBOS questionnaires. Linear regression analysis was then performed and the significance of individual regression coefficients assessed. The scores for each SF-36 variable in the outcome categories excellent and good were pooled and compared with those in the fair and poor groups using Student's *t*-test and applying Bonferroni's correction with a significance level of $p \leq 0.006$.

Internal consistency was assessed by Cronbach's alpha and item-scale correlation.[6,7] The former is an interitem correlation statistic with a range of 0 to 1. A high value indicates that items on a scale are correlated and that the scale measures a single underlying quality in the questionnaire. The reliability of comparisons is considered satisfactory when alpha exceeds 0.7.[8] Item-scale correlations should be greater than 0.4 to indicate homogeneity of the scale.[9]

We performed all statistical calculations using a software package (Minitab, release 9 for Windows, State College, Pennsylvania).

Outcome assessment of minimal intervention fusion with instrumented posterolateral lumbar fusion. Eighty patients had posterolateral fusion using pedicular screws (Universal Spine System; Stratec Medical, Welwyn Garden City, UK). All had disabling chronic low back pain. We used provocative discography to identify the likely source of pain. Most had a two-level fusion. Some had leg pain and had additional nerve-root decompression. There were 40 men and 40 women with a mean age at operation of 47 years (21 to 73). They were reviewed at a mean follow-up of 18 months (12 to 48) by an independent assessor (RK). We assessed patients by interview, physical examination, and flexion-extension and oblique radiography. All completed an SF-36 questionnaire.

Twenty-six patients with chronic low back pain had a 'keyhole fusion'.[10] Fourteen had a degenerative spondylolisthesis with associated neurogenic claudication. There were 13 men and 13 women with a mean age of 59 years (37 to 88). Through a 2 cm incision, a bilateral fenestration approach was used to expose the disc which was evacuated and the endplates debrided. The lateral recesses were decompressed as necessary, a posterolateral gutter prepared and the transverse processes decorticated. A mixture of two parts morcellised autograft and one part hydroxyapatite granules was injected into the prepared bed. Most patients had a single-level fusion. They were mobilised the next day and wore a light corset for three months. This group was independently assessed (MG) at a mean follow-up of 18 months.

There were a similar number of smokers in each group (47% and 46%, respectively). The clinical details and the SF-36 scores for both groups were compared using Student's *t*-test. The level of significance was taken as $p = 0.05$.

RESULTS

Comparison of SF-36 with Oswestry and LBOS scales. Of the 120 patients, 18 were either lost to follow-up or failed to complete all items of the questionnaires.

There was significant correlation ($p < 0.001$) between the Oswestry and LBOS scores and the SF-36 questionnaire (Table II). The correlation coefficients were highest for pain and physical and social function; none was greater than 0.8. Physical function accounted for the largest variance (60%) in both the Oswestry and LBOS scores.

Mental health had the lowest correlation when compared with both back pain outcome measures, and also gave the lowest variance (< 20%).

Table II. Correlation ($p < 0.001$) and regression coefficients (SD) for the Oswestry and Low Back Outcome Scores compared with SF-36 variables

SF-36 variable	Correlation coefficient	Regression coefficient	
		Constant	Independent variable
Oswestry			
Physical function	-0.77	67.2 (3.6)	-0.62 (0.05)
Social function	-0.67	70.0 (4.8)	-0.55 (0.06)
Role limitation (physical)	-0.48	41.7 (3.1)	-0.26 (0.04)
Role limitation (emotional)	-0.50	51.9 (4.4)	-0.31 (0.05)
Mental health	-0.40	64.8 (8.3)	-0.51 (0.11)
Energy	-0.55	59.6 (4.0)	-0.58 (0.09)
Pain	-0.64	58.6 (4.0)	-0.51 (0.06)
Health perception	-0.54	59.5 (5.2)	-0.51 (0.08)
Low Back Outcome Score			
Physical function	0.78	27.4 (3.5)	0.64 (0.05)
Social function	0.66	26.1 (4.9)	0.54 (0.06)
Role limitation (physical)	0.64	49.6 (2.7)	0.35 (0.04)
Role limitation (emotional)	0.54	41.6 (4.3)	0.34 (0.05)
Mental health	0.43	28.8 (8.2)	0.54 (0.11)
Energy	0.57	34.8 (4.9)	0.60 (0.09)
Pain	0.75	31.5 (3.4)	0.61 (0.05)
Health perception	0.63	30.6 (4.7)	0.60 (0.07)

Table III. Comparison of mean SF-36 scores with the excellent/good and fair/poor LBOS groups. The difference between them was significant ($p < 0.0001$)

SF-36 variable	Excellent/good (n = 56) (SD)	Fair/poor (n = 46) (SD)
Physical function	81 (16)	35 (22)
Social function	91 (18)	52 (24)
Role (physical)	72 (38)	16 (31)
Role (emotional)	91 (24)	49 (36)
Mental health	78 (15)	59 (17)
Energy	63 (18)	38 (19)
Pain	76 (23)	33 (16)
Health perception	73 (19)	41 (22)

The regression equations derived from the analysis of the relationship between physical function and the Oswestry and LBOS scores were:

Oswestry = 67.2 – 0.62 physical function

LBOS = 27.4 + 0.64 physical function

The regression coefficients and the standard deviations for the remaining constants and independent SF-36 variables are listed in Table II. Comparison of the mean scores of each SF-36 variable for the excellent/good and fair/poor LBOS groups showed a significant difference between the two groups ($p < 0.0001$; Table III). Cronbach's alpha was significant for each variable of the SF-36; the lowest value was 0.75. The item-scale correlation was also high, with the lowest value being 0.6 (Table IV). Histograms of the mean SF-36 scores in the four LBOS outcome groups were produced for each health variable and Figure 2 illustrates the histogram profiles for physical function, pain and health perception. The profiles were similar and the same pattern was seen in the other SF-36 variables.

Comparison of fusion groups. The results of the inter-group comparison are summarised in Table V. The 'key-hole' fusion group was significantly older than the conventional fusion group but had a shorter duration of preoperative disability. The conventional fusion group had significantly greater postoperative Oswestry scores and also had reduced physical and social function and role limitation of the SF-36.

DISCUSSION

Nilsson et al[11] showed that significant reduction in the function and quality of life occurred in patients with radiological loosening of total hip replacements as measured by the Nottingham Health Profile (NHP) questionnaire. This was a small sample and the NHP has been criticised for its failure to quantify low levels of disability.[12] This deficiency is important in patients with back pain since many have minor residual disability even after 'successful' treatment.[13]

Garratt et al[4] found that such patients had lower scores in five of the eight SF-36 scales when compared with those with a suspected peptic ulcer, menorrhagia or varicose veins. The health profiles of these four chronic conditions differed significantly. Comparison with our results for APLD showed that the SF-36 scores for the fair/poor outcome groups were similar to those of the patients with

Table IV. Cronbach's alpha statistic and item-scale correlation coefficients for SF-36 variables

SF-36 variable	Cronbach's alpha	Item number									
		1	2	3	4	5	6	7	8	9	10
Physical function	0.90	0.70	0.79	0.79	0.70	0.69	0.64	0.84	0.85	0.75	0.60
Social function	0.82	0.91	0.96								
Role (physical)	0.97	0.91	0.79	0.88	0.78						
Role (emotional)	0.75	0.68	0.75	0.70							
Mental health	0.83	0.78	0.86	0.84	0.71	0.72					
Energy	0.90	0.88	0.92	0.84	0.90						
Pain	0.77	0.90	0.91								
Health	0.86	0.91	0.68	0.86	0.74	0.82					

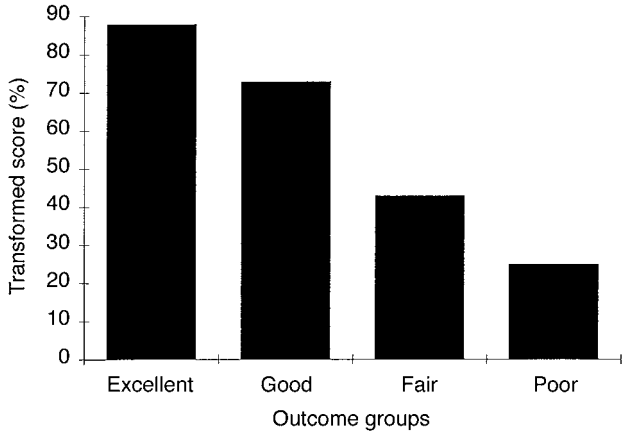

Fig. 2a

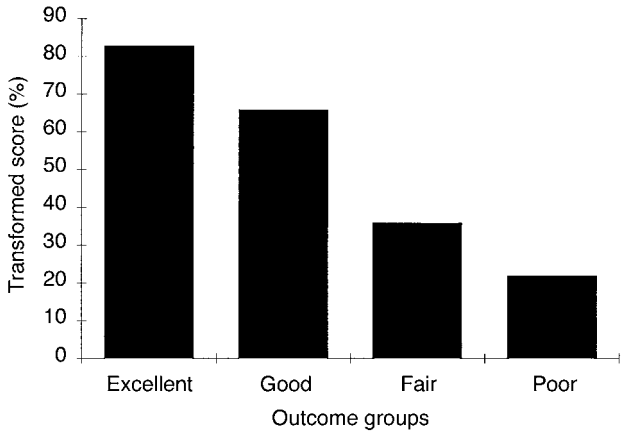

Fig. 2b

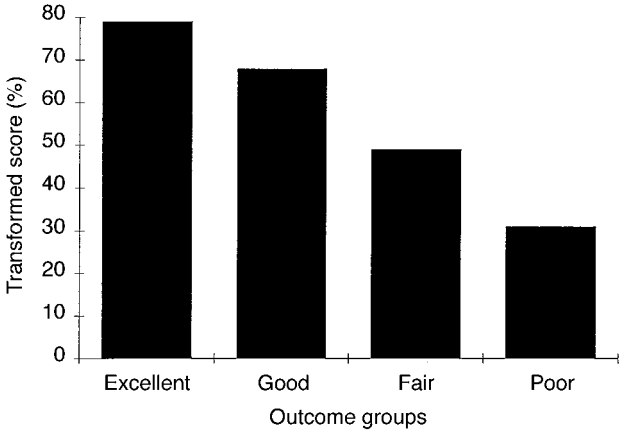

Fig. 2c

Fig. 2

Correlation of each LBOS group with mean SF-36 scores for physical function (a) pain (b) and health perception (c).

Table V. Comparison of clinical details and SF-36 scores in the fusion groups. Mean values (SD) are given

	Instrumented (n = 80)	**'Keyhole' (n = 26)**	**p value**
Age in years	47 (12)	59 (12)	0.0001
Preop disability in months	71 (39)	29 (21)	0.0001
Postop Oswestry score	51 (21)	38 (26)	0.002
Physical function	31 (29)	46 (29)	0.02
Social function	44 (32)	60 (39)	0.03
Role limitation (physical)	12 (27)	31 (41)	0.03
Role limitation (emotional)	35 (43)	67 (21)	0.002
Mental	58 (23)	67 (23)	0.07
Energy	34 (23)	44 (26)	0.06
Pain	35 (52)	45 (20)	0.2
Health perception	46 (28)	56 (2)	0.07

back pain who had been referred for specialist treatment by family practitioners. The excellent/good outcome group had scores which approximated to those of a normal population of the same age. Although the patients who had undergone APLD were a highly selected group, there was significant correlation between the SF-36 scores and back pain disability scales. The validity of the SF-36 for postoperative assessment is confirmed by the significant difference in the scores between the two main outcome groups.

Jenkinson et al[5] showed that the SF-36 has internal consistency when applied to both a general population and a group with chronic illness. This relationship was tested on patients with back pain in the community and on those referred for further treatment by their family practitioner.[4] Even although our patients represent a very selected population these findings were confirmed in our study. Similar histogram profiles were maintained for each SF-36 scale (Fig. 2).

The weakest correlation of both the LBOS and Oswestry scores was with the SF-36 mental-health variable. This may seem surprising given that psychological distress and illness behaviour are commonly associated with disorders of the lower back,[14] but may be because patients with back pain are not a homogeneous psychological population. Patients with the same level of physical disability may differ in their ability to cope or deal with depression.[15] We did not include measurement of the level of psychological distress in this study. Although five of the 35 SF-36 items assess mental

health the questions may have insufficient sensitivity to distinguish different psychological subgroups.

The assessment of a normal population showed an age-related decline in all SF-36 scores.[5] Although patients undergoing keyhole fusion were older than those who had a conventional fusion their physical and social impairment was less after the operation. This may reflect the longer period of preoperative disability and the more extensive surgery in the latter group. The pain scores were not significantly different, which suggests that physical disability and social dysfunction are modulated by factors other than pain. It may be that the greater physical demands and expectations of younger patients affect their subjective assessment of outcome after surgery but this preliminary report suggests the need for a randomised trial of the two types of operation using the SF-36 as an outcome measure.

Our study describes the analysis of patients with back pain using the SF-36 as an outcome measure. The questionnaire was not available at the start of the APLD trial and changes in the scores after surgery cannot be inferred from our results. We are now commencing studies to determine if the SF-36 scores and the overall health profile alter if the symptoms resolve naturally. We also include the SF-36 questionnaire in the prospective evaluation of all patients undergoing operation.

There are few reports of the use of SF-36 after other types of operation. Marsh, Smith and Do[16] used it in the late assessment of patients after treatment of fractures of the tibial plateau and found only moderate correlation with the Iowa knee scoring system. This comparison was limited, however, by the small number of patients and the lack of preoperative data. While pain and physical disability are common symptoms, other orthopaedic complaints such as deformity and disfigurement are not directly measured in the SF-36. Further studies are needed to compare the SF-36 questionnaire with other measures of orthopaedic outcome using prospective assessments to determine the sensitivity to change or 'responsiveness' of the SF-36 questionnaire in differing groups of patients.

We wish to thank Kate Tilling, Lecturer in Medical Statistics, St Thomas' Hospital, and Dr James Pearson, Senior Lecturer in Medical Statistics, Department of Public Health Medicine, University of Nottingham for their invaluable advice.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

REFERENCES

1. **Fairbank JC, Couper J, Davis JB, O'Brien JP.** The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271-3.
2. **Greenough CG, Fraser RD.** Assessment of outcome in patients with low-back pain. *Spine* 1992;17:36-41.
3. **Tarlov AR, Ware JE, Greenfield S, Nelson EC, Perrin E, Zubkoff M.** The medical outcomes study: an application of methods for monitoring the results of medical care. *JAMA* 1989;262:925-30.
4. **Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT.** The SF36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *BMJ* 1993;306:1440-4.
5. **Jenkinson C, Coulter A, Wright L.** Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. *BMJ* 1993;306:1437-40.
6. **Cronbach LJ.** Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297-34.
7. **Streiner DL, NormalGR.** *Health measurement scales: a practical guide to their development and use*. 2nd edition. Oxford, etc: Oxford University Press, 1995:65-6.
8. **Nunnally JC.** *Psychometric theory.* 2nd edition. New York: McGraw-Hill, 1978.
9. **Kline P.** *A handbook of test construction*. London: Methuen, 1986.
10. **Shepperd JAN, Peckham T, Knight G.** 'Keyhole' fusion. Procs European Spine Society, Nordwijk aan Zee, September, 1995.
11. **Nilsson LT, Franzén H, Carlsson AS, Onnerfalt R.** Early radiographic loosening impairs the function of a total hip replacement: the Nottingham Health Profile of 49 patients at five years. *J Bone Joint Surg [Br]* 1994;76-B:235-9.
12. **Kind P, Carr-Hill R.** The Nottingham Health Profile: a useful tool for epidemiologists? *Soc Sci Med* 1987;25:905-10.
13. **Grevitt MP, McLaren A, Shackleford IM, Mulholland RCM.** Automated percutaneous lumbar discectomy: an outcome study. *J Bone Joint Surg [Br]* 1995;77-B:626-9.
14. **Waddell G, Morris EW, Di Paola MP, Bircher M, Finlayson D.** A concept of illness tested as an improved basis for surgical decisions in low-back disorders. *Spine* 1986;11:712-9.
15. **Main et al.**The distress and risk assessment method. A simple patient classification to identify distress and evaluate the risk of poor outcome. *Spine* 1992;17:42-53.
16. **Marsh JL, Smith ST, Do TT.** External fixation and limited internal fixation for complex fractures of the tibial plateau. *J Bone Joint Surg [Am]* 1995;77-A:661-73.