SPINE Volume 31, Number 6, pp 639–643
©2006, Lippincott Williams & Wilkins, Inc.

Quality of Life Assessment Using the Short Form-12 (SF-12) Questionnaire in Patients With Cervical Spondylotic Myelopathy

Comparison With SF-36

Anoushka Singh, PhD,* Kanna Gnanalingham, FRCS,† Adrian Casey, FRCS,* and Alan Crockard, FRCS*

Study Design. Clinical outcome study comparing the Short Form-36 (SF-36) and Short Form-12 (SF-12) assessment scales in patients with cervical spondylotic myelopathy (CSM).

Objectives. To compare the validity, reliability, and sensitivity to change of the SF-12 and SF-36 scales in CSM patients undergoing decompressive surgery.

Summary of Background Data. The SF-36 is a generic Health Related Quality of Life (HRQoL) questionnaire, consisting of 36 questions that can be reported as a Physical (PCS) and Mental Component Summary (MCS). Recently, an abbreviated version of SF-36, the SF-12, which uses only 12 questions drawn from the SF-36, has been described.

Methods. In this prospective study, patients with CSM undergoing decompressive surgery, self-completed the SF-36 questionnaire before surgery and at 6 months after surgery. SF-12 item responses were abstracted from the responses given to the SF-36 questionnaire. The validity, reliability, and sensitivity to change of the PCS and MCS components of SF-12 and SF-36 scales were compared.

Results. Overall, 105 patients underwent anterior (N = 58) or posterior (N = 47) decompressive surgery. After surgery, there were improvements in the PCS components of both the SF-36 (40 ± 2 to 54 ± 2) and SF-12 (34 ± 2 to 48 ± 3), as well as in the MCS component of the SF-36 (48 ± 2 to 63 ± 2) and SF-12 (43 ± 2 to 59 ± 2) ($P < 0.001$). The sensitivity to change and absolute sensitivity for both SF-12 and SF-36 were comparable, but the reliability of SF-36 was marginally greater. There were close and linear correlations between the SF-36 and SF-12 scores for both the PCS and MCS components, before and after surgery (R = 0.86 to 0.93; $P < 0.0001$).

Conclusions. Both the SF-12 and SF-36 scales are valid and sensitive to changes in physical and mental health status in CSM patients, undergoing decompressive surgery. Despite its abbreviated nature, the SF-12 appears to be an adequate substitute for SF-36, and its brevity should increase its attractiveness to both clinicians and patients.

Key words: cervical spondylotic myelopathy, Short Form-36, Short Form-12, comparison. **Spine 2006;31:639–643**

From the *Department of Surgical Neurology, National Hospital for Neurology and Neurosurgery, Queen Square, London, UK; and †Department of Neurosurgery, Hope Hospital, Stott Lane, Salford, UK.
Acknowledgment date: March 4, 2005. First revision date: March 25, 2005. Acceptance date: March 29, 2005.
The manuscript submitted does not contain information about medical device(s)/drug(s).
No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.
Address correspondence and reprint requests to Anoushka Singh, PhD, National Hospital for Neurology and Neurosurgery, Box No. 94, Queen Square, London, WC1N 3BG, UK. E-mail: anoushka.singh@uclh.org

Cervical spondylotic myelopathy (CSM) is a cause of significant morbidity, and there is ongoing debate as to the value and timing of surgical decompression in this condition.[1] In an attempt to obtain objective data about disease severity and the impact of surgical intervention, a variety of disease-specific and generic assessment scales have been used in CSM patients.[2–11]

Generic assessment scales are designed for use in a wide variety of diseases and also tend to be multidimensional in assessing Health Related Quality of Life (HRQoL). One of the most popular is the Short Form-36 (SF-36) questionnaire, which is a comprehensive measure of physical and mental HRQoL.[7] It is self-administered and consists of 36 questions that can be grouped in to 1 of 8 health domains and/or reported as a Physical (PCS) and Mental Component Summary (MCS).[7] The reliability and validity of SF-36 have been demonstrated in CSM patients.[8–11] More recently, an abbreviated version of SF-36, the SF-12, has been described, which is shorter and therefore quicker to complete.[12,13] We compared the validity, reliability, and sensitivity to change of the PCS and MCS components of SF-12 and SF-36 in CSM patients undergoing decompressive surgery.

Methods

A total of 105 patients, consecutively referred and accepted for decompressive surgery to the Neurosurgical Unit at the National Hospital for Neurology and Neurosurgery, were selected for study. The study had local ethical committee approval, and informed consent was obtained from each patient. CSM was diagnosed on the basis of appropriate history, clinical examination, and investigations, notably an MRI scan of the cervical spine. None of the patients had undergone previous neck surgery or had other coexisting pathology that might result in functional impairment.

The patients were under the care of six consultant neurosurgeons. The same assessor, an independent nurse practitioner, experienced in the use of assessment scales in CSM, evaluated each patient before surgery and at 6 months postsurgery.[11] All patients were given the SF-36 questionnaire to complete, and the nurse practitioner who had no input to the surgical management of the patients was available to answer any queries and to ensure that all parts of the questionnaire were completed.

Analysis of the SF-36 and SF-12 Questionnaires. The data from the 36 questions of the SF-36 questionnaire can be

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

grouped into 1 of 8 separate health domains: physical functioning (PF), physical role (PR), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), emotional role (ER), and mental health (MH).[7] Each health domain is scored on a 0 to 100 metric, the higher score reflecting better health.[7,14] These 8 health domains may be further reduced to a PCS and MCS.[7,14]

SF-12 uses only 12 questions (Appendix available for online viewing through ArticlePlus) drawn from each of the 8 domains of the SF-36.[12,13] The SF-12 item responses can thus be abstracted from the responses given to the SF-36 questionnaire, and this approach gives comparable results to those obtained when SF-12 and SF-36 are administered separately.[12] SF-12 is also reported as PCS and MCS components.[12,13]

For both the SF-36 and SF-12, the data from completed questionnaires were converted to a 0 to 100 metric. The data were not converted to "norm-based scoring" with respect to a general population, since this requires normative data, which is not available for a corresponding U.K. population.[7,12–14]

A number of properties of the PCS and MCS components of SF-36 and SF-12 were considered, including the:

1. Mean change in PCS and MCS scores: [mean (postop) − (preop)]
2. Sensitivity to change (*i.e.*, responsiveness to change), measured as the standardized response mean: [mean (postop) − (preop)/standard deviation].[15] Higher values indicate greater sensitivity of the scale to detect changes following surgical intervention.
3. Absolute sensitivity, measured as the coefficient of variation: [interquartile range/median].[15] Higher values indicate greater sensitivity of the scale to distinguish between different levels of disease severity in the patient population.
4. Reliability (*i.e.*, internal consistency), measured by Cronbach's alpha.[16,17] This is the normalized measure of correlations between multiple components of one scale, and a score of 1 indicates perfect correlation and high reliability between different components of the same scale.
5. Validity was assessed in terms of correlations between the PCS, MCS, and the respective changes between the SF-36 and SF-12 scores, before and after surgery.[18]

For "absolute sensitivity" and "reliability," the preoperative and postoperative scores were pooled together.

Statistical Analysis. Parametric analysis was used, after confirming the approximate normal distribution of the data, using the Kolmogorov-Smirnov test. PCS and MCS component scores of the SF-36 and SF-12, before and after surgery were compared using repeated-measures analysis of variance (ANOVA), with the different SF scales entered as the between-subjects factor. Correlations between the PCS, MCS, and respective changes between the SF-36 and SF-12 scores, before and after surgery were analyzed using the Pearson test. Scatter plots were constructed, demonstrating the correlation coefficient (R), least squares regression line, and 95% confidence interval lines. All tests were performed on the SPSS statistical package (Statistical Programs for the Social Sciences, Chicago, IL).

■ Results

Patient Demographics

From a total of 105 patients, 68 (65%) were male and the mean age was 58 ± 2 years (range, 33–89 years). The following surgical procedures were undertaken: 58 patients with anterior decompression (13, anterior cervical discectomies without fusion; 27, Smith Robinson; 15, Clowards; 3, anterior cervical discectomies with bone graft and plating) and 47 patients with posterior decompression (9, cervical laminectomies; 34, open door laminoplasties with mini-plates; 4, cervical laminectomies with lateral mass plating).

Preoperative and Postoperative Changes in SF-36 and SF-12

The PCS components of both SF-36 and SF-12 demonstrated higher postoperative scores following surgical intervention ($P < 0.001$; $F = 97$; repeated-measures ANOVA; Table 1). For the PCS component, the preoperative and postoperative scores of SF-12 were lower in comparison with the corresponding SF-36 scores ($P < 0.05$, $F = 4$; repeated-measures ANOVA; Table 1).

The MCS components of both SF-36 and SF-12 demonstrated higher postoperative scores following surgical intervention ($P < 0.001$; $F = 124$; repeated-measures ANOVA; Table 1). For the MCS component, there was no difference between the corresponding preoperative

Table 1. Comparison of the PCS and MCS Components of the SF-36 and SF-12, Preoperatively and Postoperatively in Patients With CSM (N = 105)

	Preoperative SF-36 PCS	Postoperative SF-36 PCS	Preoperative SF-12 PCS	Postoperative SF-12 PCS
Physical component summary (PCS)				
Mean ± SEM	39.7 ± 2	53.7 ± 2*	33.7 ± 2†	48.2 ± 3*†
Interquartile range	28–47	35–72	20–43	28–70
	Preoperative SF-36 MCS	Postoperative SF-36 MCS	Preoperative SF-12 MCS	Postoperative SF-12 MCS
Mental component summary (MCS)				
Mean ± SEM	47.8 ± 2	62.7 ± 2*	42.7 ± 2	59.3 ± 2*
Interquartile range	33–58	46–79	26–54	37–79

Values represent transformed scores ranging from 0 to 100, the higher score reflecting better health.
*$P < 0.001$, postoperative *vs.* corresponding preoperative score (repeated-measures ANOVA).
†$P < 0.05$ SF-12 *vs.* corresponding SF-36 component (repeated-measures ANOVA).

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

and postoperative scores of SF-12 and SF-36 scores ($P = 0.6$, $F = 0.4$; repeated-measures ANOVA; Table 1).

Comparison of the Properties of SF-36 and SF-12

The "mean change," "sensitivity to change," and "absolute sensitivity" were similar between the SF-36 and SF-12 scales, for both PCS and MCS components (Table 2). For both PCS and MCS components, the "reliability" (Cronbach's alpha) was marginally lower for SF-12 in comparison to SF-36 (Table 2).

Correlations Between SF-36 and SF-12 Scores

There were good correlations between the PCS and MCS components of the SF-36 and SF-12 scales ($R = 0.93$ to 0.96; $P < 0.0001$; Pearson test; Figure 1). The association was linear and applied to both the preoperative and postoperative scores (Figure 1), as well as the changes in scores (*i.e.*, [postop] − [preop]) ($R = 0.92$–0.93; $P < 0.0001$; Pearson test; Figure 2).

Table 2. Comparison of the Properties of the PCS and MCS Components of the SF-36 and SF-12 Scales (N = 105)

Properties	SF-36	SF-12
Mean change ± SEM Mean (postoperative) − (preoperative scores)	14.0 ± 2 (PCS) 14.9 ± 2 (MCS)	14.5 ± 2 (PCS) 16.6 ± 2 (MCS)
Sensitivity to change Mean (postoperative) − (preoperative scores)/SD	0.73 (PCS) 0.80 (MCS)	0.64 (PCS) 0.75 (MCS)
Absolute sensitivity Interquartile range/median	0.85 (PCS) 0.67 (MCS)	1.02 (PCS) 0.76 (MCS)
Reliability (Cronbach's alpha)	0.93 (PCS); 0.89 (MCS)	0.77 (PCS); 0.77 (MCS)

Mean change and sensitivity to change were calculated as the differences between the postoperative and preoperative scores. For absolute sensitivity and reliability, the preoperative and postoperative scores were pooled together.

■ Discussion

There is growing interest in generic HRQoL assessment scales to quantitate the impact of diseases and their treatments. SF-36 provides a multidimensional measure of physical and mental HRQoL in a variety of disease states, including CSM.[7–10] However, this 36-item questionnaire can be time-consuming for the patient to complete and for the physician to analyze. In the present study, we validated the recently described SF-12, an abbreviated version of the SF-36, in CSM patients undergoing decompressive surgery.

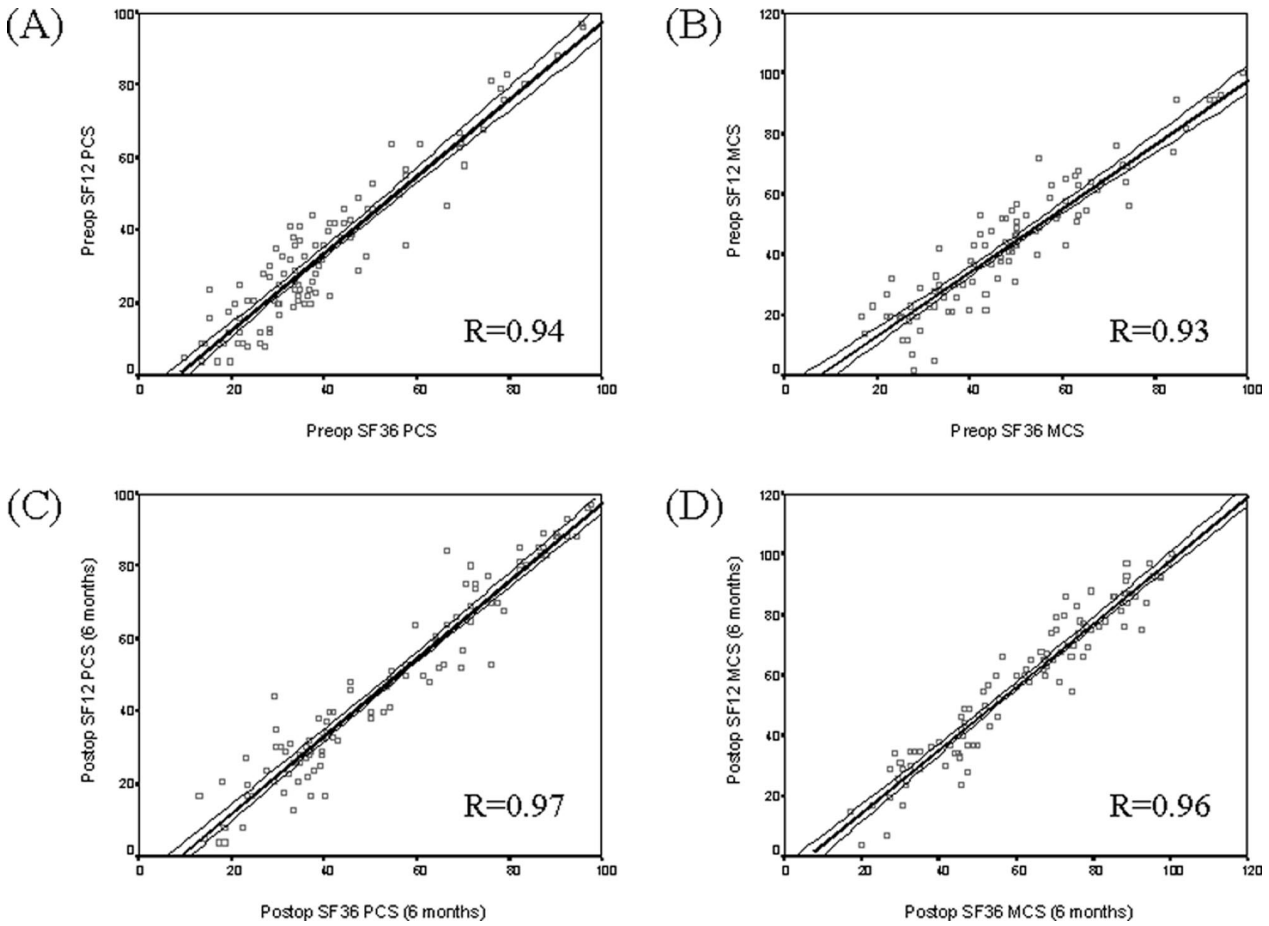

Figure 1. Scatter plots showing the correlation between the PCS and MCS components of the SF36 and SF12, pre- (**A** and **C**) and postoperatively (**B** and **D**) in patients with CSM (n = 105). The values represent transformed scores ranging from 0 to 100, the higher score reflecting better health. Pearson test with correlation coefficient (R) and a least squares regression line, with 95% confidence interval lines are shown for each scatter plot ($P < 0.0001$).

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

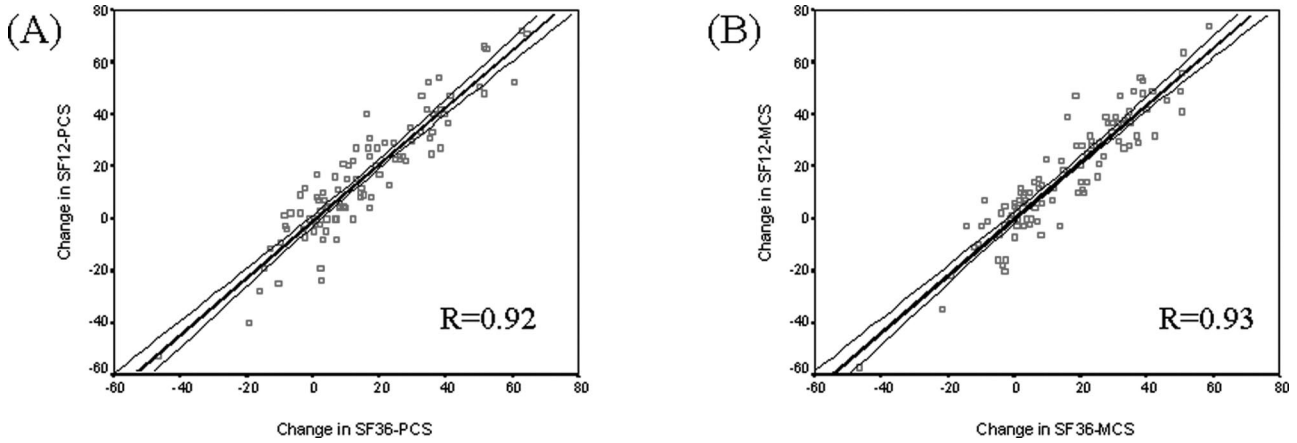

Figure 2. Scatter plots showing the correlation between the changes in the PCS and MCS components of the SF36 and SF12, pre- (**A**) and postoperatively (**B**) in patients with CSM (n = 105). The values represent differences between the postoperative and preoperative transformed scores ranging from 0 to 100, the higher score reflecting better health. Pearson test with correlation coefficient (R) and a least squares regression line, with 95% confidence interval lines are shown for each scatter plot ($P < 0.0001$).

At 6 months following surgery, there were similar improvements in the PCS and MCS scores for both SF-36 and SF-12. Moreover, the "sensitivity to change" (*i.e.*, sensitivity of the scale to detect change after surgery) and "absolute sensitivity" (*i.e.*, sensitivity of the scale to distinguish between different levels of disease severity) were comparable for the PCS and MCS components of SF-36 and SF-12. However, reliability as assessed by the internal consistency of different items of each scale was greater for SF-36 than SF-12. This most likely reflects the threefold reduction in the number of questions answered in SF-12. Nonetheless, the value of 0.77 obtained for SF-12 is still adequate, using the Nunnally criterion of acceptability for values over 0.7.[16,17] As both SF-36 and SF-12 questionnaires are patient rated, interrater reliability would not be applicable for these assessment scales. Furthermore, the validity of SF-12 was confirmed by the near-linear correlations between the PCS and MCS scores of SF-12 and SF-36, before and after surgery. Indeed, such close correlations were also apparent for the changes in the PCS and MCS scores, before and after surgery.

Evidently, despite capturing more limited information, the SF-12 appears comparable to SF-36 in recording the changes in physical and mental HRQoL over time. Thus, the reduction of scale length from 36 to 12 items did not greatly affect the validity, sensitivity to change, or the absolute sensitivity of the SF-12 scale. The small loss of reliability associated with the fewer items in the SF-12 scale should be regarded as an acceptable trade-off for the practical benefits gained. The brevity of SF-12 should increase its attractiveness to clinicians and patients alike in both routine clinical practice as well as in a research setting. In practical terms, the single page version of the SF-12 questionnaire takes less than 3 to 4 minutes to complete, in comparison to the 10 to 15 minutes needed to cover the 3 pages of the SF-36 scale. The shorter SF-12 scale may also reduce the problems associated with noncompletion of some items in longer questionnaires. Even so, there are some advantages in using the longer SF-36 questionnaire, since it can provide a more detailed breakdown of HRQoL, in terms of the 8 health domains. SF-12, being briefer, cannot be reliably reported in terms of the 8 domains, but only as PCS and MCS summaries.[12,13]

A related issue is whether SF-12 provides any advantage over the more traditional disease-specific instruments used in assessing CSM. It is important to note that generic assessment scales such as the SF-12 and SF-36 provide a more comprehensive measure of HRQoL, specifically assessing physical, social, and mental health in disease states. This is perhaps more important from a patient's perspective. In contrast, disease-specific assessment scales, such as Nurick, Ranawat, Japanese Orthopedic Association score, Walking test, and Myelopathy Disability Index, tend to focus more on the physical changes in CSM, including the neurologic symptoms and signs.[2–6] As a result, disease-specific assessment scales may be more sensitive to the physical deficits noted in CSM patients but neglect the global impact on other aspects of patient's health.[9–11] In this respect, generic scales such as the SF-12 complement the information provided by disease-specific assessment scales. Furthermore, while the SF-12 can be used to monitor the changes in HRQoL in a group of patients, its role as an adjunct to clinical decision-making in individual patients needs further study.

■ Key Points

- The Short Form-36 (SF-36) is a generic Health Related Quality of Life (HRQoL) questionnaire, consisting of 36 questions.
- Recently, an abbreviated version of SF-36, the SF-12, which uses only 12 questions drawn from the SF-36, has been described.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

- Outcome following decompressive surgery in patients with cervical spondylotic myelopathy (CSM) was measured using the SF-12 and SF-36 scales.
- Both the SF-12 and SF-36 scales are valid and sensitive to changes in physical and mental health status in CSM patients.
- Despite its abbreviated nature, the SF-12 appears to be an adequate substitute for SF-36, and its brevity should increase its attractiveness to both the clinicians and patients.

References

1. Rowland LP. Surgical treatment of cervical spondylotic myelopathy: time for a controlled trial. *Neurology* 1992;42:5–13.
2. Yonenobu K, Hosono N, Iwasaki M, et al. Neurologic complications of surgery for cervical compression myelopathy. *Spine* 1991;16:1277–82.
3. Nurick S. The pathogenesis of the spinal cord disorder associated with cervical spondylosis. *Brain* 1972;95:87–100.
4. Ranawat CS, O'Leary P, Pellicci P, et al. Cervical spine fusion in rheumatoid arthritis. *J Bone Joint Surg Am* 1979;65:1003–10.
5. Casey ATH, Bland JM, Crockard HA. Development of a functional scoring system for rheumatoid arthritis patients with cervical myelopathy. *Ann Rheum Dis* 1996;55:901–6.
6. Singh A, Crockard HA. Quantitative assessment of cervical spondylotic myelopathy by a simple walking test. *Lancet* 1999;354:370–3.
7. Ware-JE J, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;30: 473–83.
8. King JT, McGinnis KA, Roberts MS. Quality of life assessment with the medical outcomes study in short form-36 among patients with cervical spondylotic myelopathy. *Neurosurgery* 2003;52:113–20.
9. King JT, Roberts MS. Validity and reliability of the Short Form-36 in cervical spondylotic myelopathy. *J Neurosurg Spine* 2002;97:180–5.
10. Latimer M, Haden N, Seeley HM, et al. Measurement of outcome in patients with cervical spondylotic myelopathy treated surgically. *Br J Neurosurg* 2002;16:545–9.
11. Singh A, Crockard HA. Comparison of seven different scales used quality severity of cervical spondylotic myelopathy and post-operative improvement. *J Outcome Measurement* 2001;5:798–818.
12. Ware JE, Kosinski M, Keller SD. SF-12: *How to Score the SF-12 Physical and Mental Health Summary Scales*, 2nd ed. Boston: Health Institute, New England Medical Centre, 1995.
13. Ware JE, Kosinski M, Keller SD. A 12-item Short Form Health Survey: construction of scales and preliminary test of reliability and validity. *Med Care* 1996;34:220–33.
14. Kalantar-Zadeh K, Kopple JD, Block G, et al. Association among SF-36 quality of life measures and nutrition, hospitalisation and mortality in haemodialysis. *J Am Soc Nephrol* 2001;12:2797–806.
15. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis* 1987;40:171–8.
16. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull* 1955;52:281–302.
17. Nunnally JC, Bernstein IH. *Psychometric Theory*, 2nd ed. New York: McGraw-Hill, 1994.
18. Payton OD. *Research: The Validation of Clinical Practice*, 3rd ed. Philadelphia: Davis, 1994:67–83.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.