Problems in using health survey questionnaires in older patients with physical disabilities. The reliability and validity of the SF-36 and the effect of cognitive impairment

D. Gwyn Seymour BSc MB ChB MD FRCP,[1] Anne E. Ball BSc MPH,[2]
Elizabeth M. Russell MB ChB MD FFCM FRCP,[3] William R. Primrose MB ChB MRCGP FRCP,[4]
Andrew M. Garratt BA MSc PhD[5] and John R. Crawford BSc MSc PhD[6]
[1]Professor, Medicine for the Elderly, University of Aberdeen, Foresterhill Health Centre, Westburn Road, Aberdeen AB25 2AY, UK
[2]Research Assistant, Department of Public Health, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen AB25 2AY, UK
[3]Professor, Department of Public Health, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen AB25 2AY, UK
[4]Consultant Physician, Department of Medicine for the Elderly, Woodend Hospital, Eday Road, Aberdeen AB15 6XS, UK
[5]Research Fellow, Institute of Health Sciences, Old Road, Oxford OX3 7LF, UK
[6]Reader in Psychology, Department of Psychology, King's College, University of Aberdeen, Aberdeen AB24 2UB, UK

Correspondence:
Professor D. Gwyn Seymour
Medicine for the Elderly
Foresterhill Health Centre
Westburn Road
Aberdeen AB25 2AY
UK

Keywords: aged, elderly, reliability, SF-36 questionnaire, validity

Accepted for publication:
19 January 2001

Abstract

Reliability and validity of the SF-36 Health Survey Questionnaire was assessed in older rehabilitation patients, comparing cognitively impaired with cognitively normal subjects. The SF-36 was administered by face-to-face interview to 314 patients (58–93 years) in the day hospital and rehabilitation wards of a department of medicine for the elderly. Reliability was measured using Cronbach's alpha (for internal consistency) on the main sample and intraclass correlation coefficients on a test–retest sample; correlations with functional independence measure (FIM) were examined to assess validity. In 203 cognitively normal patients (Mini-Mental State Examination ≥24), Cronbach's alpha scores on the eight dimensions of the SF-36 ranged from 0.545 (social function) to 0.933 (bodily pain). The range for the 111 cognitively impaired patients was 0.413–0.861. Cronbach's alpha values were significantly higher (i.e. reliability was better) in the cognitively normal group for bodily pain ($P = 0.003$), mental health ($P = 0.03$) and role emotional ($P = 0.04$). In test–retest studies on a further 67 patients, an intraclass correlation coefficient of 0.7 was attained for five out of eight dimensions in cognitively normal patients, and four out of eight dimensions in the cognitively impaired. Only the physical function dimension in the cognitively normal group attained the criterion level ($r > 0.4$) for construct validity when correlated with the FIM. In this group of older physically disabled patients, levels of reliability and validity previously reported for the SF-36 in younger subjects were not attained, even on face-to-face testing. Patients with coexistent cognitive impairment performed worse than those who were cognitively normal.

Introduction

The Short Form 36 Health Survey Questionnaire (SF-36) has been widely recommended for people in all age-groups (Ware *et al*. 1993; Garratt *et al*. 1994; McHorney *et al*. 1994), but some recent reports have questioned its performance in older patients, particularly those with a physical disability or cognitive impairment (Garratt *et al*. 1994; Hill *et al*. 1996; Gladman 1998; O'Mahony *et al*. 1998; Parker *et al*. 1998), or those classified as 'frail' (Stadnyk *et al*. 1998). The present study has therefore formally assessed the reliability and validity of the SF-36 in a consecutive series of patients referred for physical rehabilitation to a department of medicine for the elderly, dividing patients into two subgroups based on the presence or absence of cognitive impairment.

As well as a transition question, the SF-36 comprises eight multi-item dimensions, which are physical function (10 items), role physical (role limitations due to physical problems, four items), bodily pain (two items), general health (five items), vitality (four items), social functioning (two items), role emotional (role limitation due to emotional problems, three items) and mental health (five items) (Ware *et al*. 1993; McHorney *et al*. 1994; Garratt *et al*. 1994; Stadnyk *et al*. 1998). Each of the dimensions is scored from 0 to 100, with higher scores indicating better health. The SF-36 was primarily conceived as a postal survey for self-completion, but this leads to poor response rates and high levels of missing data in many older patients (Hayes *et al*. 1995; Brazier *et al*. 1996; Hobson & Meara 1997; Mallinson 1998; Parker *et al*. 1998; O'Mahony *et al*. 1998), especially those with physical disability and cognitive impairment (Parker *et al*. 1998). The present investigation has therefore used face-to-face interviews to avoid the problems associated with self-completion.

Methods

Subjects in the main patient sample ($n = 314$)

These were patients referred for physical rehabilitation to the Department of Medicine for the Elderly, Woodend Hospital, Aberdeen. Eighty-two per cent were assessed in the day hospital, the remainder being recently admitted in-patients. Patients with an acute disability requiring immediate admission (such as an acute vertebral crush fracture) were excluded, as were dysphasic patients and residents of nursing homes. Of 338 patients who met the inclusion criteria, 314 agreed to take part. The SF-36 (UK Standard Acute Version, Ware *et al*. 1993) was administered during a face-to-face interview by a research nurse or a research assistant. At the same time, cognitive function was tested by the Mini-Mental State Examination (Tombaugh & Mcintyre 1992), with a score of 23 or less on the MMSE being taken as evidence of cognitive impairment (McDowell & Newell 1996). Physical function was assessed by a research occupational therapist using the motor subscale of the functional independence measure (FIM) (UDS Data Management Service 1990). The latter produces a score of between 13 and 91, with higher scores representing better function. The test instruments were administered in the same order for all patients.

Subjects in the test–retest group ($n = 67$)

An additional group of 91 day hospital patients who were not in the main study group had the SF-36 administered on two occasions, 1 week apart. Of these 91 patients, 67 said that their health had not changed between the two assessments, and they were used to estimate the test–retest reliability of the SF-36.

Statistical analysis

Our prior hypothesis was that the group of patients with cognitive impairment would differ from the group who were cognitively normal in respect of their reliability and validity on the eight separate dimensions of the SF-36. In line with the approach of Perneger (1998), we have not applied Bonferroni corrections to the individual *P*-values.

Measurement of reliability, main patient sample As the eight dimensions of the SF-36 are each made up of more than one item, it is possible to calculate whether answers given to items within each dimension are internally consistent with one another. This method of testing reliability does not require the questionnaire to be administered twice. In the main patient sample, the internal consistency of the eight dimensions of the SF-36 was tested by Cronbach's alpha (Cronbach 1951; Bland & Altman 1997) and statistical comparisons of alpha scores from cognitively normal and cognitively impaired groups were carried out using the methods of Feldt (1969, 1980) and Feldt *et al.* (1987).

Measurement of reliability, test–retest sample Here the reliability was tested using the intraclass correlation coefficient (ICC) (Deyo *et al*. 1991).

Criteria of reliability For measures of reliability, such as Cronbach's alpha or the ICC, it is usually suggested that levels of 0.7 and above are acceptable when groups of patients are being compared, but that levels between 0.90 and 0.95 are required for monitoring individuals (Bland & Altman 1997; Nunnally 1978).

Measurement of validity The SF-36 looks at various aspects of health status from the patient's point of view, but there is no 'gold standard' against which its eight dimensions can be validated. The commonest technique for establishing validity in health status instruments of this type is to assess 'convergent validity', which is a subtype of 'construct validity' (McHorney & Tarlov 1995). To do this, the dimensions are correlated in turn with another, 'external', variable that should be broadly related to them. Correlation coefficients between 0.4 and 0.6 are evidence that the same construct is being embraced by the instrument and the external standard (Streiner & Norman 1995). In the present group of physical rehabilitation patients, we used the FIM Motor Scale (UDS Data Management Service 1990) as the external variable. Our two prior hypotheses were, first, that the more 'physical' dimensions of the SF-36 (such as physical function and role physical) would show the highest correlations with the FIM motor score, and secondly, that correlations between SF-36 scores and the external variable would be higher in the cognitively normal group than in the cognitively impaired group.

Results

The mean age of the 314 patients interviewed was 79.7 years (range 58–95 years), 212 (67.5%) were female, and 111 (35.4%) were classified as cognitively impaired (MMSE score 23 or less), of whom 38 had scores of 17 or lower. The mean baseline FIM motor score was 71.3 (range 21–91, median 74). Table 1 shows mean values of the eight dimensions of the SF-36 in the cognitively normal and cognitively impaired groups. For two dimensions (role physical and social function) the cognitively impaired group had statistically significantly higher mean values than the cognitively normal group. However, the FIM motor score as measured by the occupational therapist was actually worse in the cognitively impaired group (66.2 vs. 73.5, $t = 4.8$, $P < 0.0001$), perhaps casting doubts on the ability of the cognitively impaired patients to estimate how much their activities were limited by their physical status.

Table 1 Mean scores on the eight dimensions of the SF-36

SF-36 dimension	*Cognitively normal patients* Mean score (95% CI)*	*Cognitively impaired patients† Mean score (95% CI)*	*P-value for t-test comparison of the two cognitive groups*
Physical function	24.4 (21.5, 27.4)	26.1 (22.2, 30.0)	0.81
Role physical	17.3 (13.6, 20.9)	27.5 (21.3, 33.7)	0.01
Bodily pain	54.5 (50.0, 58.9)	59.0 (52.9, 65.1)	0.32
General health	51.8 (48.4, 55.1)	54.2 (50.0, 58.6)	0.44
Vitality	40.3 (37.1, 43.6)	44.1 (39.4, 48.8)	0.22
Social function	53.6 (49.3, 57.8)	63.0 (57.4, 68.5)	0.02
Role emotional	77.9 (72.7, 83.2)	80.7 (74.1, 87.3)	0.55
Mental health	72.4 (69.4, 75.3)	71.8 (67.8, 75.8)	0.75

* Mini-Mental State ≥24, $n = 203$, mean age 78.6, range 58–93. † Mini-Mental State ≤23, $n = 111$, mean age 81.6, range 65–95.

Table 2 Internal consistency of the SF-36 questionnaire. Cronbach's alpha in cognitively normal and cognitively impaired groups

SF-36	*Number of items in dimension*	*Cognitively normal patients** α *(95% CI)*	*Cognitively impaired patients†* α *(95% CI)*	*P-value‡*
Physical function	10	0.860 (0.829–0.887)	0.812 (0.754–0.861)	0.11
Role physical	4	0.679 (0.599–0.746)	0.737 (0.646–0.810)	0.30
Bodily pain	2	0.933 (0.911–0.949)	0.861 (0.797–0.905)	0.003
General health	5	0.739 (0.677–0.792)	0.633 (0.511–0.732)	0.08
Vitality	4	0.712 (0.640–0.772)	0.716 (0.616–0.795)	0.95
Social function	2	0.545 (0.398–0.656)	0.413 (0.141–0.599)	0.29
Role emotional	3	0.907 (0.882–0.927)	0.857 (0.803–0.898)	0.04
Mental health	5	0.801 (0.753–0.842)	0.692 (0.587–0.777)	0.03

* Mini-Mental State $\geq$24, $n = 197–201$ (occasional missing values), mean age 78.6, range 58–93. † Mini-Mental State $\leq$23, $n = 104–109$ (occasional missing values), mean age 81.6, range 65–95. ‡ For comparison of the two cognitive groups using the W statistic of Feldt (1969, 1980) and Feldt *et al.* (1987).

Reliability (internal consistency of main patient sample, $n = 314$)

Internal consistency for the two cognitive groups, as measured by Cronbach's alpha, is shown in Table 2. In cognitively normal patients the 0.7 (group comparison) criterion for Cronbach's alpha was attained for six out of eight dimensions, and the 0.9 (individual comparison) criterion for two out of eight dimensions. In cognitively impaired patients, five out of eight dimensions reached the 0.7 criterion, but none reached the 0.9 criterion.

Our prior hypothesis was that impaired cognitive function would be associated with poorer internal consistency. The final column of Table 2 therefore uses the statistical techniques of Feldt (1969, 1980) and Feldt *et al.* (1987) to show that, compared with the cognitively normal group, cognitively impaired patients had significantly lower Cronbach's alpha values for the SF-36 dimensions of bodily pain ($P = 0.003$), role emotional ($P = 0.04$) and mental health ($P = 0.03$) with borderline values for general health ($P = 0.08$) and physical function ($P = 0.11$). These differences are seen more readily in Fig. 1.

Reliability (intraclass correlation coefficient, test–retest sample, $n = 67$)

The intraclass correlation coefficient is an alternative to Cronbach's alpha when testing reliability, but the questionnaire needs to be administered twice in a group of subjects whose status has remained stable. In the test–retest group, among the 48 stable patients with normal cognitive status, five of the eight SF-36 dimensions (physical function, bodily pain, general health, vitality, mental health) yielded ICCs greater than the 0.7 criterion level. Among the 19 stable patients who were cognitively impaired, four dimensions (physical function, bodily pain, general health, vitality) had ICCs above 0.7. Formal testing of the differences between ICC values in the cognitively normal and cognitively impaired groups did not show any significant differences, but sample size was small.

Validity (main patient sample, $n = 314$)

Table 3 shows that, as hypothesized, the strongest correlations on Pearson's r statistic occurred between the motor FIM (the 'external' variable) and the physical function scale, but even here the previously stated criterion for construct validity (an r-value between 0.4 and 0.6) was reached only in the cognitively normal group. Correlations between the motor FIM and the remaining seven dimensions of the SF-36 were much more modest, with none being above 0.22. Many of these r-values were statistically significantly different from zero, but, as can be seen from Table 3, the proportion of variance explained (r^2) was low. When the r-values of the cognitively normal and the cognitively impaired group were compared (Table 3, final column), a statistically significant

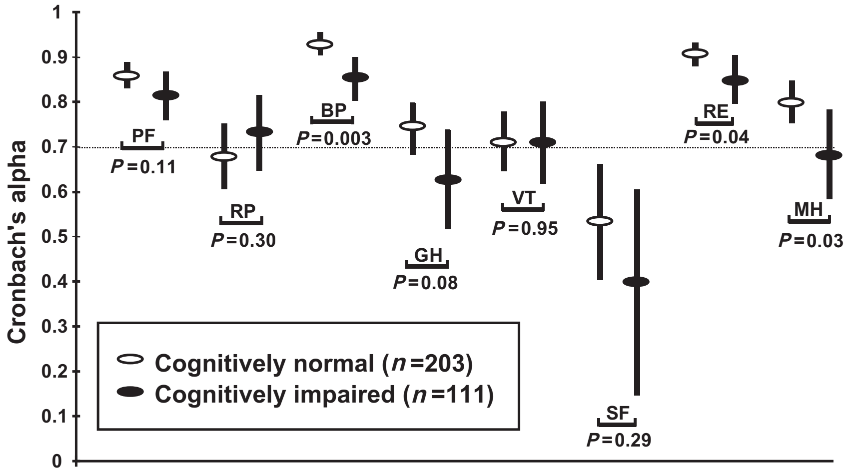

Figure 1 The reliability (internal consistency) of scores in cognitively normal and cognitively impaired patient groups is shown for each of the eight dimensions of the SF-36, using Cronbach's alpha. Cronbach's alpha levels of 0.7 are usually advised for comparisons of groups but levels of 0.9–0.95 are needed for comparisons involving individuals (Nunnally 1978; McHorney & Tarlov 1995; Bland & Altman 1997). The 0.7 criterion is shown as a dotted line in the figure. 95% confidence intervals for Cronbach's alpha and statistical comparisons between the two cognitive groups have been calculated using the methods of Feldt (1969, 1980) and Feldt *et al.* (1987). Abbreviations for the eight dimensions of the SF-36 are as follows: PF = physical functioning; RP = role limitation due to physical problems; BP = bodily pain; GH = general health perception; VT = vitality; SF = social functioning; RE = role limitation due to emotional problems; MH = mental health.

Table 3 Examination of construct validity. Correlations (Pearson's *r*) between the functional independence measure (FIM) motor scale and each of the eight dimensions of the SF-36

SF-36	*Correlation with FIM in cognitively normal patients** *r*	*Correlation with FIM in cognitively impaired patients†* *r*	*P-value for two-tailed comparison of the two cognitive groups‡*
Physical function	0.528	0.331	0.04
Role physical	0.132	0.197	0.58
Bodily pain	0.156	0.184	0.81
General health	0.133	0.088	0.70
Vitality	0.214	0.180	0.77
Social function	0.092	0.032	0.61
Role emotion	0.120	–0.029	0.21
Mental health	0.192	0.220	0.81

* Mini-Mental State ≥24, $n = 203$, mean age 78.6, range 58–93. † Mini-Mental State ≤23, $n = 111$, mean age 81.6, range 65–95. ‡ Zar (1996).

difference was found for the physical function dimension ($P = 0.04$), indicating better construct validity for the SF-36 in the cognitively normal group. As well as these parametric analyses, non-parametric correlations using Spearman's Rho were carried out, but the results were very similar to those using Pearson's r.

Discussion

One of the attractions of a generic health survey questionnaire such as the SF-36 is that, in theory at least, a wide range of people with a variety of medical conditions can be assessed with the same instrument. However, if certain sections of the population are unable to complete the questionnaire satisfactorily, they become 'disenfranchised' because they are excluded from studies and surveys. While there have been reports about the adverse effects of age, cognitive impairment and physical status on rates of self-completion of the SF-36 (Hayes *et al*. 1995; Brazier *et al*. 1996; Hobson & Meara 1997; Gladman 1998; Parker *et al*. 1998; Mallinson 1998), the present study appears to be the first formal assessment of the effect of cognitive impairment on the reliability and validity of the SF-36 in a group of older people.

In view of the widely acknowledged difficulties of postal administration of the SF-36 in older patients (Gladman 1998; Parker *et al*. 1998; Hayes *et al*. 1995; Brazier *et al*. 1996; Hobson & Meara 1997; Mallinson 1998), face-to-face administration was used in the present study. Even so, we found the reliability (internal consistency measured by Cronbach's alpha) of the questionnaire to be worse in patients with cognitive impairment. To make statistical comparisons, we used methods for comparing Cronbach's alpha values that have been available for many years as a result of the elegant work of Feldt (1969, 1980) and Feldt *et al.* (1987) and which deserve to be better known.

How do the levels of reliability we found in the present study compare with general recommendations about 'acceptable' levels for health status instruments? The 0.9 criterion for Cronbach's alpha, widely recommended for studies involving individual comparisons (McHorney & Tarlov 1995; Bland & Altman 1997), was never attained in our cognitively impaired patients, and even in the cognitively normal group it was reached only for the bodily pain and role emotional dimensions. More encouragingly, the 0.7 group comparison (Bland & Altman 1997; Nunnally 1978) criterion was attained for six of the eight dimensions in the cognitively normal group and five of the eight dimensions in the cognitively impaired group. However, these levels of internal consistency are worse than those previously reported in studies of the SF-36 in which middle-aged patients predominated. Thus, in the USA a large postal community survey (McHorney *et al*. 1994) found that the 0.7 criterion was reached in all but one of 192 subgroup analyses, while hospital outpatients attained alpha values between 0.74 and 0.90 (McHorney & Tarlov 1995). Within the UK there have been three widely quoted postal studies of general populations (Brazier *et al*. 1992; Jenkinson *et al*. 1993; Garratt *et al*. 1994), all of which reported alpha values of 0.80 or over, except for the social function dimension where the lowest level was 0.73.

Information from SF-36 postal surveys in very elderly patients is limited, but in one survey of subjects aged over 75 years in the USA (McHorney *et al*. 1994) the lowest Cronbach's alpha value was 0.77 (for general health), and among UK women in the same age range, alpha values were 0.8 and above, except for social function (0.56) and general health (0.66) (Brazier *et al*. 1996). However, this high level of reliability was found in successfully completed postal questionnaires. Many of the elderly women in the UK study appeared to encounter difficulty in completing the questionnaire, with missing data rates reaching 32% for some dimensions (Hayes *et al*. 1995). The non-completion rate of the SF-36 was even higher in a recent American postal community survey of subjects aged over 65 years (Andresen *et al*. 1996). In that study, while Cronbach's alpha values were 0.70 or above for seven out of eight dimensions, only 253 out of 422 people returned the forms, and many individual questions went unanswered. One hundred and eighty-six of the respondents returned a second form 4 weeks later, and in this subgroup the test–retest performance of the SF-36 was very good (ICCs 0.65–0.87). However, this final sample comprised only 44% of those originally sent postal questionnaires, and it contained a disproportionate number of younger and fitter subjects.

A Canadian study has used face-to-face administration of the SF-36 questionnaire in 146 patients aged 65 and over who were designated as 'frail' (Stadnyk *et al.* 1998). The study differed from our own in that patients with an MMSE of 17 or less were excluded, and the effect of cognitive impairment on reliability and validity of the SF-36 was not studied. The Canadian estimates of Cronbach's alpha values were within the 95% confidence intervals of our cognitively normal group (see Table 2) for three of the SF-36 dimensions (general health, role emotional, mental health), were higher for four (physical function, role physical, vitality, social function) and were lower for one (bodily pain). The authors concluded that the SF-36 could not be regarded as 'the optimum outcome measure' for frail elderly patients, and that further research using frail elderly people was required (Stadnyk *et al*. 1998). A recent study from the USA administered SF-36 questionnaires to nursing home residents (Andresen *et al*. 1999). The inclusion criteria were similar to those used in the Canadian study (Stadnyk *et al*. 1998), but this meant that only one in five of all nursing home residents were able to take part. In this selected group reliability characteristics were judged to be 'fairly good', with intraclass correlation coefficients exceeding 0.70 on four of the eight dimensions of the SF-36.

As well as testing the reliability of the SF-36 in our patient group, we looked at construct (convergent) validity. Before the start of the study it was argued that the physical function and the role physical dimensions of the SF-36 should correlate best with the FIM motor score, but we found that only in the case of the physical function dimension, and then only in cognitively normal patients, was the recommended range of r (0.4–0.6) achieved. However, if a test has poor reliability, this puts an upper limit on the levels of validity that can be achieved (Kline 1993), and so some of the low correlations seen in Table 3 may have been due to the poor reliability already demonstrated in Table 2. The nursing home study by Andresen *et al*. (1999) reported correlations of 0.37 and 0.43, respectively, when the physical function and role physical scales were correlated with a measure of activities of daily living, but, as has been mentioned above, patients with poor cognitive function were excluded.

Conclusion

We conclude that in older physically disabled patients, particularly in those with cognitive impairment, levels of reliability and validity of the SF-36 previously reported for younger or fitter people are not attainable, even when questionnaires are administered face to face. There is an even bigger question mark over postal or self-completed questionnaires in such patients. Our study demonstrates that, even for an instrument as well-designed as the SF-36, the reliability and validity of a health survey questionnaire needs to be specifically tested in the type of patient group in which it is to be used.

Acknowledgements

This research was funded by the Chief Scientist Office of the Scottish Office Home and Health Department, but the opinions in this article are those of the authors. We also thank the staff and patients at Woodend Hospital for their assistance and Lorraine Davidson and Claire Caie for their help with data collection. For statistical advice we thank Jill Mollison. For the use of their questionnaires we acknowledge the following: the Medical Outcomes Study Trust, Boston, USA, for the use of the SF-36 health survey questionnaire, and the Center for Functional Assessment Research, State University of New York at Buffalo, for the FIM.

References

Andresen E.M., Bowley N., Rothenberg B.M., Panzer R. & Katz P. (1996) Test-retest performance of a mailed version of the Medical Outcomes Study 36-item Short-Form health survey among older adults. *Medical Care* **34**, 1165–1170.

Andresen E.M., Gravitt G.W., Aydelotte M.E. & Podgorski C.A. (1999) Limitations of the SF-36 in a sample of nursing home residents. *Age and Ageing* **28**, 562–566.

Bland J.M. & Altman D.G. (1997) Cronbach's alpha. *British Medical Journal* **314**, 572.

Brazier J.E., Harper R., Jones N.M.B., O'Cathain A., Thomas K.J., Usherwood T. & Westlake L. (1992) Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *British Medical Journal* **305**, 160–164.

Brazier J.E., Walters S.J., Nicholl J.P. & Kohler B. (1996) Using the SF-36 and Euroqol on an elderly population. *Quality of Life Research* **5**, 195–204.

Cronbach L.J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* **16**, 297–333.

Deyo R.A., Diehr P. & Patrick D.L. (1991) Reproducibility and responsiveness of health status measures. *Controlled Clinical Trials* **12**, 142S–158S.

Feldt L.S. (1969) A test of the hypothesis that Cronbach's Alpha or Kuder-Richardson Coefficient Twenty is the same for two tests. *Psychometrika* **34**, 363–373.

Feldt L.S. (1980) A test of the hypothesis that Cronbach's Alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika* **45**, 99–105.

Feldt L.S., Woodruff D.J. & Salih F.A. (1987) Statistical inference for Coefficient Alpha. *Applied Psychological Measurement* **11**, 93–103.

Garratt A.M., Ruta D.A., Abdalla M.I., Buckingham J.K. & Russell I.T. (1994). The SF-36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *British Medical Journal* **306**, 1440–1444.

Gladman J.R.F. (1998) Assessing health status with the SF-36. *Age and Ageing* **27**, 3.

Hill S., Harries U. & Popay J. (1996) 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. *Journal of Epidemiology and Community Health* **50**, 94–98.

Hayes V., Morris J., Wolfe C. & Morgan M. (1995) The SF-36 health survey questionnaire: is it suitable for use with older adults? *Age and Ageing* **24**, 1120–1125.

Hobson J.P. & Meara R.J. (1997) Is the SF-36 health survey questionnaire suitable as a self-report measure of the health status of older adults with Parkinson's disease? *Quality of Life Research* **6**, 213–216.

Jenkinson C., Coulter A. & Wright L. (1993) Short Form 36 (SF-36) health survey questionnaire. Normative data for adults of working age. *British Medical Journal* **306**, 1437–1440.

Kline P. (1993) *The Handbook of Psychological Testing*. Routledge, London.

McDowell I. & Newell C. (1996) *Measuring Health: a Guide to Rating Scales and Questionnaires*, 2nd edn. Oxford University Press, New York.

McHorney C.A., Ware J.E., Rachel J.F. & Sherbourne C.D. (1994) The MOS 36-item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care* **32**, 40–66.

McHorney C.A. & Tarlov A.R. (1995) Individual patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research* **4**, 293–307.

Mallinson S. (1998) The Short-Form 36 and older people: some problems encountered when using postal administration. *Journal of Epidemiology and Community Health* **52**, 324–328.

Nunnally J.C. (1978) *Psychometric Theory*, 2nd edn. McGraw-Hill, New York.

O'Mahony P.G., Rodgers H., Thompson R.G., Dobson R. & James O.F.W. (1998) Is the SF-36 suitable for assessing health status of older stroke patients. *Age and Ageing* **27**, 19–22.

Parker S.G., Peet S.M., Jagger C., Farhan M. & Castleden C.M. (1998) Measuring health status in older patients. The SF-36 in practice. *Age and Ageing* **27**, 13–18.

Perneger T.V. (1998) What's wrong with Bonferroni adjustments? *British Medical Journal* **316**, 1236–1238.

Stadnyk K., Calder J. & Rockwood K. (1998) Testing the measurement properties of the Short Form-36 Health Survey in a frail elderly population. *Journal of Clinical Epidemiology* **51**, 827–835.

Streiner D.L. & Norman G.R. (1995) *Health Measurement Scales. A Practical Guide to Their Development and Use*, 2nd edn. Oxford University Press, Oxford.

Tombaugh T.N. & Mcintyre N.J. (1992) The Mini-Mental State Examination: a comprehensive review. *JAGS* **40**, 922–935.

UDS Data Management Service (1990) *Guide for Use of the Uniform Data Set for Medical Rehabilitation Including the Functional Independence Measure, Version 3.0.* Uniform Data System for Medical Rehabilitation, The Buffalo General Hospital, Buffalo, New York.

Ware J.E., Snow K.K., Kosinski M. & Gandeck B. (1993) *SF-36 Health Survey: Manual and Interpretation Guide.* The Health Institute, New England Medical Center, Boston, Massachusetts.

Zar J.H. (1996) *Biostatistical Analysis*, 3rd edn. Prentice Hall, Upper Saddle River NJ.

