

A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies?

Crispin Jenkinson, Richard Layte, Damian Jenkinson, Kate Lawrence, Sophie Petersen, Colin Paice and John Stradling

Abstract

Background The SF-36 is a generic health status measure which has gained popularity as a measure of outcome in a wide variety of patient groups and social surveys. However, there is a need for even shorter measures, which reduce respondent burden. The developers of the SF-36 have consequently suggested that a 12-item sub-set of the items may accurately reproduce the two summary component scores which can be derived from the SF-36 [the Physical Component Summary Score (PCS) and Mental Health Component Summary Score (MCS)]. In this paper, we adopt scoring algorithms for the UK SF-36 and SF-12 summary scores to evaluate the picture of change gained in various treatment groups.

Methods The SF-36 was administered in three treatment groups (ACE inhibitors for congestive heart failure, continuous positive airways therapy for sleep apnoea, and open vs laparoscopic surgery for inguinal hernia).

Results PCS and MCS scores calculated from the SF-36 or a sub-set of 12 items (the 'SF-12') were virtually identical, and indicated the same magnitude of ill-health and degree of change over time.

Conclusion The results suggest that where two summary scores of health status are adequate then the SF-12 may be the instrument of choice.

Keywords: SF-36, SF-12, summary scales, sensitivity to change

Introduction

The value of patient-based measures of outcomes has become established in medical research and evaluation.¹ Consequently, the search for short, easily completed health status measures has been a growth area. Much of this research has concentrated on the development of disease-specific measures,² although the search for a short, easily completed, cheap and psychometrically sound generic measure has been an undertaking with relatively few successes. At present, the most popular generic measure of outcomes is probably the 36-item short form health survey (SF-36).³ This is a questionnaire developed in the United

States,^{4–7} which has been adopted as one of the standard health outcomes measures both in America and elsewhere.^{8–11} The questionnaire contains 36 questions covering eight dimensions of health status (see Table 1). The wording of the questionnaire has been anglicized for use in the United Kingdom, and substantial research has been undertaken to determine the validity and reliability of the measure across various populations and patient groups in this country.^{12–19}

Recently, the developers of the SF-36 have suggested a method of gaining two summary scores from the instrument.²⁰ Using data from the Medical Outcomes Study,²¹ a Physical Component Summary Scale Score (PCS) and Mental Component Summary Scale Score (MCS) have been generated, and their validity and reliability evaluated. The intention of the developers when constructing the PCS and MCS was to reduce the SF-36 from an eight-scale profile to two summary measures of health status without substantial loss of information. Furthermore, they have suggested that the summary scores may make trial and other longitudinal data sets more easily interpretable, as they reduce the number of statistical comparisons required from eight to two.²² Within the UK context we undertook an analysis of the Oxford Healthy Lifestyle Survey²³ dataset and have developed scoring algorithms for the summary scores in the UK setting.²⁴

¹Health Services Research Unit, University of Oxford, Institute of Health Sciences, Headington, Oxford OX3 7LF.

²Nuffield College, Oxford OX1 1NF.

³Stroke Unit, The Royal Bournemouth and Christchurch Hospitals NHS Trust, Bournemouth BH7 7DW.

⁴Chest Unit, Churchill Hospital, Oxford OX3 7LJ.

Crispin Jenkinson, Deputy Director¹

Richard Layte, Research Officer²

Damian Jenkinson, Consultant Physician³

Kate Lawrence, MRC Training Fellow¹

Sophie Petersen, Research Officer¹

Colin Paice, Computer Officer¹

John Stradling, Consultant Physician⁴

Address correspondence to Dr C. Jenkinson.

Table 1 Dimensions of the SF-36, and number of items in each (there is one further unscaled item in the SF-36 asking respondents about health change over the past year)

| Scales evaluating functional status | Scales evaluating well-being | Scale evaluating general health perception |
|--|------------------------------|--|
| Physical functioning (10 items) | Mental health (5 items) | (5 items) |
| Social functioning (2 items) | Energy/vitality (4 items) | |
| Role limitations due to physical problems (4 items) | Pain (2 items) | |
| Role limitations due to emotional problems (3 items) | | |

The developers have consequently suggested that a 12-item sub-set of the original 36 items of the SF-36 can be used to construct a shorter form health survey which can produce the PCS and MCS without substantial loss of information.^{25,26} They have suggested that the SF-12 is able to produce the two summary scales originally developed from the SF-36 with considerable accuracy and yet with far less respondent burden. Consequently, the SF-12 may be an instrument of choice where a short generic measure providing summary information on physical and mental health status is required. Similarly, we have replicated the work of the developers in a UK context, and produced summary scores from the SF-12 which closely mirror those produced from the SF-36.²⁷ However, to date, little evidence has been available concerning the sensitivity of these measures to change. This is an important aspect of any health status measure which is used to assess outcomes of treatment.²⁸⁻³⁰ The purpose of this paper, therefore, is to compare the sensitivity to change of the PCS and MCS calculated from the SF-36 with the PCS and MCS calculated from the 12-item subset (the 'SF-12') from three separate data sets in which the SF-36 was included. We will adopt the convention suggested by the developers, namely, that the summary scores when calculated from the SF-36 will be referred to as the PCS-36 and MCS-36, and when calculated from the SF-12 they will be referred to as the PCS-12 and MCS-12.²⁶

Methods and materials

The data presented here are based upon three longitudinal datasets. Full details of recruitment, sample characteristics, power calculations and randomization procedures are provided in full elsewhere, and are available on request from the first author. We briefly outline below the study designs and entry criteria.

Congestive heart failure ACE inhibitor study

The SF-36 was given to elderly patients before and after they were placed on angiotensin converting enzyme inhibitor therapy (ACE inhibitors) for congestive heart failure (CHF). The entry criteria to the study were that patients had to be over 60 years of age and were included if they met the criteria which

were used to establish a diagnosis of CHF in the Framingham study.³¹ All patients were assured that inclusion in the study would not alter their treatment in any way and would not subject them to extra examinations or medical tests. Patients received a test dose of 6.25 mg of captopril under close medical supervision in the Radcliffe Infirmary NHS Trust day hospital in Oxford. No significant hypotensive events were encountered and it was possible to establish patients on a maintenance dose of between 10 and 20 mg daily of enalapril in one or two divided doses by the time of follow-up at approximately one month after the initial interview. Patients received additional conventional treatment in the form of diuretics and vasodilators at the discretion of their primary physician. They were followed up approximately four weeks after being placed on therapy.

Sleep apnoea CPAP study

The SF-36 was given to patients diagnosed with sleep apnoea before and after they were placed on continuous positive airways pressure therapy (CPAP).³² Patients were recruited from the sleep clinic at the Churchill Hospital, Oxford. A diagnosis of sleep apnoea was made if a patient presented with an Epworth Sleepiness Scale (ESS)³³ score outside of the normal range, and if the Desaturation index was more than ten per hour, which approximates to the number of apnoeas per hour and is well above what is considered normal.³² The ESS is a standard measure of hypersomnolence used to evaluate the impact of sleep apnoea on daytime function.³² Patients were followed up approximately four weeks after being placed on therapy.

Randomized trial of open vs laparoscopic surgery for inguinal hernia

Between December 1992 and August 1993 consecutive patients presenting with inguinal hernia were screened for entry into this study, which took place at the John Radcliffe Hospital, Oxford. Patients were eligible for the study if they had a primary, unilateral inguinal hernia on examination and met the local criteria for day surgery [American Society of Anaesthesia (ASA) grade 1 or 2, age less than 70 years]. Patients with previous major abdominal surgery or requiring overnight admission were excluded. Sealed randomization envelopes

**THE U.K. SHORT FORM 12 HEALTH SURVEY QUESTIONNAIRE
(UK SF-12)**

The following questions ask for your views about your health, how you feel and how well you are able to do your usual activities.

If you are unsure about how to answer any questions please give the best answer you can and make any of your own comments if you like. Do not spend too much time in answering as your immediate response is likely to be the most accurate.

1. In general, would you say your health is (please tick one box)

| | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Excellent | Very good | Good | Fair | Poor |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

2. HEALTH AND DAILY ACTIVITIES

The following questions are about activities you might do during a typical day. Does your health limit you in these activities? If so, how much? (please tick one box on each line)

| | | | | |
|----|--|--------------------------|-----------------------------|------------------------------|
| | | Yes, limited a lot | Yes, limited a little | No, not limited at all |
| a) | Moderate activities, such as moving a table, pushing a vacuum, bowling or playing golf | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| b) | Climbing several flights of stairs | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

3. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health? (Please answer Yes or No to each question)

| | | | |
|----|--|--------------------------|--------------------------|
| | | Yes | No |
| a) | Accomplished less than you would like | <input type="checkbox"/> | <input type="checkbox"/> |
| b) | Were limited in the kind of work or other activities | <input type="checkbox"/> | <input type="checkbox"/> |

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)? (Please answer Yes or No to each question)

| | | | |
|----|--|--------------------------|--------------------------|
| | | Yes | No |
| a) | Accomplished less than you would like | <input type="checkbox"/> | <input type="checkbox"/> |
| b) | Didn't do work or other activities as carefully as usual | <input type="checkbox"/> | <input type="checkbox"/> |

5. During the past 4 weeks how much did pain interfere with your normal work (including work both outside the home and housework)? (Please tick one box)

| | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Not at all | A little bit | Moderately | Quite a bit | Extremely |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

YOUR FEELINGS

6. These questions are about how you feel and how things have been with you during the past month. For each question, please indicate the one answer that comes closest to the way you have been feeling. (Please tick one box on each line)

| | | | | | | | |
|----|--|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | How much time during the last month: | All of the time | Most of the time | A good bit of the time | Some of the time | A little of the time | None of the time |
| a) | Have you felt calm and peaceful? | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| b) | Did you have a lot of energy? | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| c) | Have you felt downhearted and low? | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| d) | Has your health limited your social activities (like visiting friends or close relatives)? | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

UK SF-12 is a trade mark of the Medical Outcomes Trust

Figure 1 The UK SF-12.

Downloaded from http://pubs.aphis.gov/journals.org/ by guest on October 16, 2014

from tables of random numbers were kept in the outpatient clinic. Patients who agreed to take part in the study were allocated in the clinic to open or laparoscopic surgery using unrestricted randomization in a 1:1 ratio. Patients received either standard open hernia repair or laparoscopic repair. All surgery was performed, using a general anaesthetic, by four surgeons who were experienced laparoscopists of senior registrar or consultant grade. Patients were followed up at various intervals after surgery, although we report here only on the first follow-up, at 10 days. Results of longer-term follow-up from the trial have been presented elsewhere.^{34,35} The purpose here is to compare the SF-36 and SF-12 summary scores on a clinical trial dataset.

Construction of SF-36 summary measures

The methods for deriving the summary scale scores have been published in full elsewhere.^{20,22,24} Briefly, the procedure was as follows. SF-36 scores were calculated for the eight dimensions from data from a large-scale population survey, the Oxford Healthy Lifestyle Survey (OHLS).²³ These data were factor analysed in accordance with the recommendations of the MCS-PCS developers, using principal components analysis and orthogonal rotation.²⁰ Each of the eight individual scales of the SF-36 was then standardized using a z-score transformation using means and standard deviations calculated for all respondents in the OHLS. Each z-score was calculated by subtracting the OHLS mean from each individual respondent's scale score and dividing the difference by the corresponding scale's standard deviation from the OHLS. After the z-scores had been calculated for each scale the aggregate scores for the physical and mental component scale scores were computed. In the case of the PCS, this involved multiplying each SF-36 scale z-score by its respective factor score coefficient. Similarly, in the case of the MCS, this involved multiplying each SF-36 scale z-score by its respective factor score coefficient. Finally, these scores were standardized to a T-score where the mean was set to 50 and the SD to 10. It is these factor loadings and calculation algorithms that are applied to datasets reported in this paper.

Construction of SF-12 summary measures

The methods for deriving the summary scale scores have been published in full elsewhere.²⁵⁻²⁷ Briefly, the procedure was as follows. Once again, the procedure is based upon the normative OHLS dataset. The summary scores for the SF-12 were created using procedures recommended by the developers.³⁶ The 12 items recommended for inclusion in the SF-12 were selected from the SF-36. These are shown in Fig. 1. The first step in creating the PCS-12 and MCS-12 was to reverse score four of the 12 chosen items so that on all items a higher score indicates better health. The second step was to create indicator variables (1/0) for all but one response choice category for each item. A one is assigned to the response choice if endorsed and a zero is assigned if not endorsed. Indicator variables are not created for

the highest health state for each item. Therefore out of 48 total response categories among the 12 items only 36 indicator variables are created. It should be noted that this is one more than in the US SF-12 summary scores owing to the fact that the standard UK SF-36 is based upon the original questionnaire made available from InterStudy in the United States. Minor changes have altered the response categories in the US version, now made available from the Medical Outcomes Trust, but, in the United Kingdom, a network of users agreed to standardize on the original questionnaire.³⁷ The indicator variables were then weighted. This was undertaken by regressing the indicator variables against the PCS-36 to gain PCS-12 indicator variable weights, and the indicator variables against the MCS-36 to gain MCS-12 indicator variable weights. Computation of PCS-12 is achieved by multiplying each indicator variable by its respective physical regression weight, and summing the 36 products. Similarly, MCS-12 is computed by multiplying each indicator variable by its respective mental regression weight and summing the 36 products. The summary scales are then transformed to norm-based scoring, where the mean is set to 50 and the SD to 10. This is accomplished by adding the constant, generated in the regression, for the PCS to the sum of the 36 products gained from multiplying the indicator variables by the PCS weights, and, similarly, by adding the constant, generated in the regression, for the MCS to the sum of the 36 products gained from multiplying the indicator variables by the MCS weights. It is these regression loadings and calculation algorithms that are applied to datasets reported in this paper.

Results

The results reported here are based solely on cases where there are both baseline data and follow-up data sufficient to calculate the summary scores. The median age of patients with CHF was 82 (range 60–92 years; $n=61$). Eighteen patients were male and 43 female. The median age of patients with sleep apnoea was 50 (range 28–68 years; $n=63$). All patients recruited into this study were male. The median age of patients randomized to open repair was 47 (range 20–69 years; $n=69$) and for those randomized to laparoscopic repair 47 (range 20–77 years; $n=66$). Three patients randomized to open repair were female, and one randomized to laparoscopic repair was female.

Descriptive statistics from the eight dimensions of the SF-36 for all treatment groups are reported in Table 2. Effect sizes have been calculated to determine the magnitude of change across dimensions in these patient groups (see Table 3). They have been calculated as recommended by Kazis *et al.*,³⁸ by dividing the mean change by the baseline standard deviation. Hakzis *et al.*³⁸ suggested that as a benchmark for assessing the relative magnitude of a change, an effect size of 0.20 is regarded as small, 0.50 as moderate and 0.80 as large.³⁹ Consequently, results suggest that ACE inhibitor therapy for CHF in the elderly has limited impact on all eight dimensions,

Table 2 SF-36 domain scores (mean, with standard deviation given in parentheses) for the four treatment groups: congestive heart failure (CHF), sleep apnoea patients receiving CPAP, and patients before and after operations for inguinal hernia [open and laparoscopic (lap.)]

| SF-36 dimensions | CHF (n = 61) | | Apnoea (n = 63) | | Hernia, open (n = 69) | | Hernia, Lap. (n = 66) | |
|---------------------------|------------------|------------------|------------------|------------------|-----------------------|------------------|-----------------------|------------------|
| | Time 1 | Time 2 | Time 1 | Time 2 | Time 1 | Time 2 | Time 1 | Time 2 |
| Physical functioning | 34.43 (26.38) | 29.75 (24.55) | 74.60 (23.39) | 79.68 (22.09) | 86.32 (14.67) | 61.45 (27.04) | 80.51 (20.09) | 65.99 (22.38) |
| Social functioning | 68.49 (31.96) | 66.85 (34.75) | 70.02 (27.12) | 89.24 (15.58) | 81.32 (15.42) | 51.21 (26.21) | 76.09 (20.56) | 59.06 (23.51) |
| Role – physical | 30.74 (39.63) | 43.03 (43.81) | 58.73 (37.34) | 77.38 (34.41) | 81.88 (31.76) | 13.59 (25.34) | 65.91 (38.64) | 29.55 (36.93) |
| Role – emotional | 54.64 (48.70) | 67.21 (43.67) | 59.79 (42.82) | 82.54 (31.60) | 89.37 (25.24) | 75.85 (37.00) | 83.39 (25.59) | 68.69 (40.88) |
| Mental health | 72.98 (19.74) | 75.93 (21.29) | 68.25 (16.30) | 76.83 (15.12) | 79.94 (17.91) | 80.59 (17.57) | 77.82 (15.28) | 80.27 (15.07) |
| Energy/vitality | 50.33 (22.43) | 45.33 (25.80) | 39.05 (20.40) | 60.40 (20.25) | 68.26 (19.77) | 54.39 (20.56) | 65.08 (17.53) | 59.67 (17.52) |
| Pain | 71.77 (28.15) | 71.58 (28.73) | 73.37 (24.24) | 80.60 (23.18) | 73.59 (21.57) | 44.93 (21.00) | 66.09 (23.64) | 59.78 (19.69) |
| General health perception | 60.84 (20.35) | 58.38 (21.34) | 61.27 (22.33) | 63.19 (26.61) | 80.04 (15.29) | 75.67 (18.17) | 74.79 (18.01) | 70.89 (16.55) |

whereas treatment for sleep apnoea has a large impact on 'social functioning' and 'energy/vitality', and a moderate impact on a number of other dimensions ('role limitations due to physical problems', 'role limitations due to emotional problems' and 'mental health'). At ten days the impact of surgery for inguinal hernia has adversely affected a number of areas of functioning and well-being dramatically, as one might expect. However, the level of adverse outcome at this stage is lower, for the most part, for the laparoscopic arm.

Table 3 Effect size calculations for the four treatment groups: congestive heart failure (CHF), sleep apnoea patients receiving CPAP, and patients before and after operations for inguinal hernia [open and laparoscopic (lap.)]

| SF-36 dimensions | CHF | Apnoea | Hernia, open | Hernia, lap. |
|---------------------------|-------|--------|--------------|--------------|
| Physical functioning | -0.18 | 0.22 | -1.70 | -0.72 |
| Social functioning | -0.05 | 0.71 | -1.95 | -0.83 |
| Role – physical | 0.31 | 0.50 | -2.15 | -0.94 |
| Role – emotional | 0.26 | 0.53 | -0.54 | -0.81 |
| Mental health | 0.15 | 0.53 | 0.04 | 0.16 |
| Energy/vitality | 0.22 | 1.05 | -0.70 | -0.31 |
| Pain | -0.01 | 0.30 | -1.33 | -0.27 |
| General health perception | -0.12 | 0.09 | -0.29 | -0.22 |

Negative effect sizes indicate a decline in reported health.

The summary scale scores were then calculated using the SF-36 and SF-12. These results are reported in Table 4. What is striking about these results is the high level of concordance between PCS-36 and PCS-12 scores, and MCS-36 and MCS-12 scores. Means, standard deviations and 95% confidence intervals (CIs) are calculated for both times across all treatment groups. Mean scores are almost identical and 95 per cent CIs always substantially overlap. Mean differences between the PCS and MCS summary scores were computed using the SF-12 and SF-36; paired *t*-tests did reveal some significant differences between the scores gained from the two measures (see Table 5). However, mean differences were never greater than two and a half points in any illness group on either the physical or mental component scores, and such a difference would never be either subjectively or clinically meaningful.⁷ At the group level, the extent and direction of change indicated by the effect size statistic are virtually identical for summary scores calculated from the 36- or 12-item instrument. Finally, although the summary scores tend to obscure which exact factors are improving or worsening, the overall story remains similar to that gained from the SF-36 eight dimension scores: notably, that the CHF group have not reported much change, the sleep apnoea group have improved both on the physical summary score and, to a greater extent, on the mental health summary score (the developers of the SF-36 and summary scores claim that 'social functioning' and 'energy/vitality' are largely aspects of emotional/mental health), and that soon after surgery for inguinal hernia there is an overall decline in functioning and well-being for both open and, to a lesser extent, laparoscopic surgery patients across dimensions.

Table 4 Comparison of summary scores calculated from the SF-12 and SF-36: mean (SD, given in parentheses) 95% CI and effect sizes

| | Time 1 Mean 95% CI | Time 2 Mean 95% CI | Effect size |
|---|----------------------------|----------------------------|-------------|
| <i>CHF (n = 61)</i> | | | |
| PCS-12 | 32.04 (11.76) 29.0–35.1 | 31.91 (14.19) 28.3–35.5 | -0.01 |
| PCS-36 | 31.47 (12.38) 28.3–34.6 | 31.91 (13.70) 28.4–35.4 | 0.04 |
| MCS-12 | 38.70 (12.31) 35.5–41.9 | 39.94 (14.01) 36.4–43.5 | 0.10 |
| MCS-36 | 39.34 (13.30) 35.9–42.7 | 40.81 (13.89) 37.3–44.4 | 0.11 |
| <i>Sleep apnoea (n = 63)</i> | | | |
| PCS-12 | 39.21 (12.41) 36.1–42.3 | 46.75 (11.62) 43.8–49.7 | 0.61 |
| PCS-36 | 39.36 (12.31) 36.3–42.5 | 46.56 (12.35) 43.4–49.7 | 0.58 |
| MCS-12 | 39.97 (11.55) 37.1–42.9 | 48.83 (10.48) 46.2–51.5 | 0.77 |
| MCS-36 | 39.69 (11.51) 36.8–42.6 | 49.01 (10.30) 46.4–51.6 | 0.81 |
| <i>Hernia (open arm) (n = 69)</i> | | | |
| PCS-12 | 47.61 (9.22) 45.4–49.8 | 27.88 (9.97) 25.5–30.3 | -2.14 |
| PCS-36 | 48.72 (9.20) 46.5–50.9 | 30.36 (9.75) 28.0–32.7 | -2.00 |
| MCS-12 | 49.88 (9.05) 47.7–52.1 | 39.30 (9.31) 37.1–41.5 | -1.17 |
| MCS-36 | 51.43 (9.13) 49.2–53.6 | 41.27 (9.72) 38.9–43.6 | -1.13 |
| <i>Hernia (laparoscopic arm) (n = 66)</i> | | | |
| PCS-12 | 42.69 (11.44) 39.9–45.5 | 33.25 (10.78) 30.6–35.9 | -0.83 |
| PCS-36 | 44.06 (11.76) 41.3–46.9 | 34.91 (10.59) 32.4–37.5 | -0.77 |
| MCS-12 | 47.61 (9.22) 45.4–49.8 | 41.55 (10.16) 39.1–44.0 | -0.66 |
| MCS-36 | 49.02 (9.65) 46.6–51.4 | 42.57 (10.38) 40.1–45.1 | -0.67 |

Negative effect sizes indicate a decline in reported health.

Discussion

This paper has presented evidence for the suitability of the SF-12 as a shorter form measure for the assessment of the summary scales in longitudinal studies. The SF-12 health survey represents a step forward in the attempt to find a short yet meaningful measure of health status. The PCS-12 and MCS-12 show similar levels of precision to the summary scores derived from the longer 36-item set. Furthermore, at the level of group analysis, the level of change detected by the summary scores of the SF-12 was, to all intents and purposes, the same as that gained from the longer form instrument. Consequently, it is possible to gain physical and mental health summary scale scores which are within at most a few points difference from

Table 5 Mean differences (with SD given in parentheses) between PCS and MCS scores calculated using the SF-36 or SF-12

| | Time 1 Mean 95% CI | Time 2 Mean 95% CI |
|---|---------------------------------|----------------------------------|
| <i>CHF (n = 61)</i> | | |
| PCS | 0.57 (3.18) -0.2–1.4 | 0 (3.26) -0.8–0.8 |
| MCS | -0.64 (2.83) -1.3–0.0 | -1.47 (2.81)*** -2.2 to -0.75 |
| <i>Sleep apnoea (n = 63)</i> | | |
| PCS | -0.15 (2.97) -0.6–1.0 | 0.19 (3.34) -0.6–1.0 |
| MCS | 0.28 (2.43) -0.4–1.0 | -0.18 (2.79) -0.8–0.4 |
| <i>Hernia (open arm) (n = 69)</i> | | |
| PCS | -1.10 (2.56)** -1.7 to -0.5 | -2.47 (2.91)*** -3.2 to -1.8 |
| MCS | -1.56 (1.93)*** -2.0 to -1.1 | 1.97 (2.53)*** -2.6 to -1.4 |
| <i>Hernia (laparoscopic arm) (n = 66)</i> | | |
| PCS | -1.37 (3.23)*** -2.2 to -0.6 | -1.67 (3.87)*** -2.6 to -0.7 |
| MCS | -1.41 (2.15)*** -1.9 to -0.9 | -1.02 (3.76)* -1.9–0.0 |

* $p < 0.05$, ** $p < 0.01$; *** $p < 0.001$.

those gained from the SF-36 in a questionnaire which can be produced on one or two pages of paper and is, for the most part, sufficient for self-completion in only a few minutes. Such a questionnaire may be ideal in, for example, large-scale surveys and 'mega-trials' where there is limited time and/or resources, or in studies in which respondent burden is already high and the inclusion of the 36-item questionnaire would be considered excessive.

It should be borne in mind, however, that the SF-12 is not without its limitations. First, the summary scores gained from the SF-12 are not identical to those gained from the SF-36. However, the level of difference is small and the developers themselves have stated that differences of two to three points are not meaningful either subjectively or clinically.⁷ Second, the questionnaire contains a number of areas of health tapped with only a single item. Consequently, as the developers themselves note, the SF-36 will provide a more reliable profile of scores across the eight domains than could be gained using the SF-12,²⁶ hence producing the eight dimension summary from the SF-12 is not recommended. After all, it was lack of precision in the multi-dimensional 20-item short form (SF-20),⁴⁰ caused in part through single-item domains, which brought about the development of the SF-36. Furthermore, the scoring algorithms for the SF-12 are more complicated than those used to gain the eight dimension scores from the SF-36. In the United States, the developers have licensed certain companies to produce software for the calculation of the

PCS-12 and MCS-12. At present, no such licensing system has been adopted in the United Kingdom. Users should, therefore, assure themselves that the scoring algorithms and regression weights they are using are from appropriate population surveys. For this purpose, we are happy to provide information on the regression weights and constants gained from the OHLS: contact the first named author, or alternatively the information is available on the web at <http://hsru.dphpc.ox.ac.uk/sf12.htm>. Consequently, where a brief generic health survey is required which produces summary indices of physical and mental health, the SF-12 is likely to be of particular value.

References

- 1 Spilker B. Introduction. In Spilker B, ed. *Quality of life and pharmacoeconomics in clinical trials*, 2nd edn. Philadelphia, PA: Lippincott-Raven, 1996.
- 2 Bowling A. *Measuring disease*. Buckingham: Open University Press, 1995.
- 3 Jenkinson C, Layte R, Wright L, Coulter A. *Manual and interpretation guide for the UK SF-36*. Oxford: Health Services Research Unit, 1996.
- 4 Ware J, Sherbourne C. The MOS 36-item short-form health survey 1: Conceptual framework and item selection. *Med Care* 1992; **30**: 473–483.
- 5 McHorney CA, Ware JE, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993; **31**: 247–263.
- 6 McHorney CA, Ware JE, Lu JFR. The MOS 36-item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions and reliability across diverse patient groups. *Med Care* 1994; **32**: 40–66.
- 7 Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36 health survey: manual and interpretation guide*. Boston, MA: Health Institute, New England Medical Center, 1993.
- 8 Ware JE, Keller SD, Gandek B, Brazier J, Sullivan M. Evaluating translations of health status questionnaires. Methods from the IQOLA project. *Int J Technol Assessment Hlth Care* 1995; **11**: 525–551.
- 9 Aaronson NK, Acquadro C, Alonso J, et al. International Quality of Life Assessment (IQOLA) project. *Qual Life Res* 1992; **1**: 349–351.
- 10 Sullivan M, Karlsson J, Ware JE. The Swedish SF-36 health survey – 1. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. *Social Sci Med* 1995; **41**: 1349–1358.
- 11 Bullinger M. German translation and psychometric testing of the SF-36 health survey: preliminary results from the IQOLA project. *Social Sci Med* 1995; **41**: 1359–1366.
- 12 Brazier J, Harper R, Jones, N, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *Br Med J* 1992; **305**: 160–164.
- 13 Garratt A, Ruta D, Abdalla M, Buckingham J, Russell I. The SF 36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *Br Med J* 1993; **306**: 1440–1444.
- 14 Ruta D, Abdalla M, Garratt A, Coutt A, Russell I. SF 36 health survey questionnaire I: Reliability in two patient based studies. *Qual Hlth Care* 1994; **3**: 180–185.
- 15 Garratt A, Ruta D, Abdalla M, Russell I. SF 36 health survey questionnaire II: Responsiveness to changes in health status in four common clinical conditions. *Qual Hlth Care* 1994; **3**: 186–193.
- 16 Jenkinson C, Coulter A, Wright L. Short Form 36 (SF 36) health survey questionnaire: normative data for adults of working age. *Br Med J* 1993; **306**: 1437–1440.
- 17 Jenkinson C, Layte R, Coulter A, Wright L. Evidence for the sensitivity of the SF-36 health status measure to inequalities in health: results from the Oxford healthy lifestyles survey. *J Epidemiol Commun Hlth* 1996; **50**: 377–380.
- 18 Lyons RA, Lo SV, Littlepage BNC. Comparative health status of patients with 11 common illnesses in Wales. *J Epidemiol Commun Hlth* 1994; **48**: 388–390.
- 19 Lyons RA, Perry HM, Littlepage BNC. Evidence for the validity of the short-form 36 questionnaire (SF-36) in an elderly population. *Age Ageing* 1994; **23**: 102–104.
- 20 Ware JE, Kosinski M, Keller SD. *SF-36 physical and mental health summary scales: a user manual*. Boston, MA: Health Institute, New England Medical Center, 1994.
- 21 Tarlov AR, Ware JE, Greenfield S, et al. The Medical Outcomes Study: an application of methods for monitoring the results of medical care. *JAMA* 1989; **262**: 925–930.
- 22 Ware JE, Kosinski M, Bayliss MS, et al. Comparison of methods for scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care* 1995; **33**: AS264–AS279.
- 23 Wright L, Harwood D, Coulter A. *Health and lifestyles in the Oxford region*. Oxford: Health Services Research Unit, 1992.
- 24 Jenkinson C, Layte R, Lawrence K. Development and testing of the SF-36 summary scale scores in the United Kingdom: results from a large scale survey and a clinical trial. *Med Care* (in press).
- 25 Ware JE, Kosinski M, Keller SD. SF-12: an even shorter health survey. *Med Outcomes Trust Bull* 1996; **4**(1): 2.
- 26 Ware JE, Kosinski M, Keller SD. A 12-item short-form health survey. Construction of scales and preliminary tests of reliability and validity. *Med Care* 1995; **34**: 220–233.
- 27 Jenkinson C, Layte R. Development and testing of the UK SF-12. *J Hlth Serv Res Policy* 1997; **2**: 14–18.

- 28 Deyo R, Inui T. Towards clinical applications of health status measures: sensitivity of scales to clinically important changes. *Hlth Serv Res* 1984; **19**: 275–289.
- 29 Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. The importance of sensitivity to change as a criterion for selection of health status measures. *Qual Hlth Care* 1992; **1**: 89–93.
- 30 Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*, 2nd edn. Oxford: Oxford University Press, 1995.
- 31 McKee PA, Castelli WP, McNamara PM, Kannel WB. The natural history of congestive heart failure: the Framingham study. *N Engl J Med* 1971; **285**: 1441–1446.
- 32 Stradling J. *Handbook of sleep-related breathing disorders*. Oxford: Oxford University Press, 1993.
- 33 Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 1991; **14**: 540–545.
- 34 Lawrence K, Doll H, McWhinnie D. Relationship between health status and postoperative return to work. *J Publ Hlth Med* 1996; **18**: 49–53.
- 35 Jenkinson C, Lawrence K, McWhinnie D, Gordon J. Sensitivity to change of health status measures in a randomized controlled trial: comparison of the COOP charts and the SF-36. *Qual Life Res* 1995; **4**: 47–52.
- 36 Ware JE, Kosinski M, Keller SD. *SF-12: how to score the SF-12 physical and mental health summary scales*, 2nd edn. Boston, MA: Health Institute, New England Medical Center, 1995.
- 37 Brazier J. The SF-36 health survey questionnaire – a tool for economists. *Hlth Econ* 1993; **2**: 213–215.
- 38 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989; **27**: S178–S189.
- 39 Cohen J. *Statistical power for the behavioural sciences*. New York: Academic Press, 1977.
- 40 Bindman AB, Keane D, Lurie N. Measuring health changes among severely ill patients: the floor phenomenon. *Med Care* 1990; **28**: 1142–1152.

Accepted on 3 November 1996