Assessment of the SF-36 version 2 in the United Kingdom

Crispin Jenkinson, Sarah Stewart-Brown, Sophie Petersen, Colin Paice

Abstract

***Objectives*—To introduce the UK SF36 Version II (SF36-II), and to (a) gain population norms for the UK SF36-II in a large community sample as well as to explore the instrument's internal consistency reliability and construct validity, and (b) to derive the Physical Component Summary (PCS) and Mental Component Summary (MCS) algorithms for the UK SF36-II.**

***Design*—Postal survey using a questionnaire booklet, containing the SF-36-II and questions on demographics and long term illness.**

***Setting*—The sample was drawn from General Practitioner Records held by the Health Authorities for Berkshire, Buckinghamshire, Northamptonshire, and Oxfordshire.**

***Sample*—The questionnaire was sent to 13 800 randomly selected subjects between the ages of 18–64 inclusive.**

***Outcome measures*—Scores for the eight dimensions of the UK SF36-II and the PCS and MCS summary scores.**

***Results*—The survey achieved a response rate of 64.4% (n=8889). Internal consistency of the different dimensions of the questionnaire were found to be high. Normative data for the SF-36 are reported, broken down by age and sex, and social class. Factor analysis of the eight domains produced a two factor solution and provided weights for the UK SF36-II.**

***Conclusion*—The SF36-II domains were shown to have improved reliability over the previous version of the UK SF36. Furthermore, enhancements to wording and response categories reduces the extent of floor and ceiling effects in the role performance dimensions. These advances are likely to lead to better precision as well as greater responsiveness in longitudinal studies.**

(*J Epidemiol Community Health* 1999;**53**:46–50)

The SF-36[1] is a generic measure of health status, providing scores on eight areas of functioning and well being as well as two broad areas of subjective well being, namely physical health and mental health.[2 3] Data from the SF-36 have been suggested as appropriate for the evaluation of a wide variety of medical interventions.[4] The measure has been widely adopted around the world and over 300 articles were published between 1988–95 that included data gained from the SF-36.[5] It has been translated into 40 languages, including English, German, French, Polish, Swedish, Spanish, Icelandic, Japanese, and Portugese. User manuals for the questionnaire have been produced by the developers in the USA and also in the UK.[6 7] However, despite the widespread use of the measure criticism has been forthcoming concerning the layout and wording of some of the items.[8] Consequently, the developers have produced a modified instrument, the SF-36 Version 2 (SF36-II), which is a direct descendent of the SF-36 Developmental Form and the SF-36 Mark 1 Standard Form. This paper introduces the UK SF36-II and assesses the internal consistency reliability and construct validity of the measure. It provides normative data, from a large scale social survey, the Third Oxford Health and Lifestyles Survey (OHLS-III), for the measure and outlines the derivation of the Physical Component Summary Score (PCS) and the Mental Health Component Summary Score (MCS) for the SF36-II.

Methods

The SF-36 is a 36 item questionnaire that measures eight multi-item dimensions of health: physical functioning (10 items) social functioning (2 items) role limitations due to physical problems (4 items), role limitations due to emotional problems (3 items), mental health (5 items), energy/vitality (4 items), pain (2 items), and general health perception (5 items). There is a further unscaled single item asking respondents about health change over the past year. For each dimension item scores are coded, summed, and transformed on to a scale from 0 (worst possible health state measured by the questionnaire) to 100 (best possible health state). Two standardised summary scores can also be calculated from the SF-36; the physical component summary (PCS) and the mental health component summary (MCS). Version 2.0 of the SF-36 Health Survey is a product of eight years of research and the experience documented in a wide variety of publications.[6] Relative to the standard SF-36[1] improvements in the content and layout of Version 2.0 include: improvements in some instructions and questions to make the wording less ambiguous, most importantly the SF-36 has been accused of containing an item with a double negative,[9] and this item has been reworded; greater comparability with translations widely used in the US and in developed countries; five level response sets in place of dichotomous response choices for seven items in the two role functioning scales.

Minor modifications to the wording of six items on the SF-36 have been made to make it acceptable in the British context. These changes have been endorsed by the SF-36 developers. Version 2.0 includes algorithms for

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

Correspondence to: Dr C Jenkinson.

Accepted for publication 15 June 1998

interval level scoring for all eight scales ranging from 0 (for worse health) to 100 (best possible health as measured by the questionnaire) as well as the same standardised scoring (mean = 50, standard deviation = 10) for the SF-36 summary scores (PCS and MCS).

Evidence from numerous focus groups and empirical studies in the US, Canada, the United Kingdom, Australia and New Zealand support the improvements in question wording adopted for Version 2.0 (see fig 1).[10] These improvements make the English language versions of the SF-36 easier to understand and administer and also improve the comparability of results across translations of the SF-36. There is evidence to suggest that five level response scales improve response rates over dichotomous response categories, such as "yes/no". Consequently the two SF-36 role functioning scales have been changed from dichotomous scales to five point response categories thereby increasing score precision without increasing respondent burden. Specifically, version 2.0 achieves a fourfold increase in the number of scale levels, and is intended to produce a substantially smaller standard deviation, as well as to reduce both ceiling and floor effects for both SF-36 role scales.

STATISTICAL ANALYSIS

Normative data are provided in the form of means and standard deviations broken down by age and sex, and social class. Not all completed questionnaires had every item on the SF-36 answered. The developers of the SF-36 suggest a method of gaining scores for missing values, but because of the sample size, and as the purpose of this paper is to provide normative data, it was decided not to use this protocol: hence missing data are not substituted. The results are based upon the eight multi-item scales of the SF-36. Internal consistency—that is, the extent to which there is correlation between items on a scale—was assessed by Cronbach's alpha, an inter-item correlation statistic, with a value range of 0–1.[11] Higher values indicate items on a dimension are correlated and, therefore, that the scale is tapping an underlying single dimension on the questionnaire. Construct validity—that is, the extent to which the questionnaire supports pre-defined hypotheses—was assessed by determining the extent to which scores on different dimensions reflected the expected distribution of health status for and between certain groups. Similar criteria used to assess the original form of the SF-36 were used to assess the construct validity of the measure,[12] namely that scores might be expected to be lower, reflecting poorer perceived health, for women than men, for those in lower social classes, and for those that report a long standing illness compared with those who did not. *t* Tests were used to analyse whether SF-36 scores differed significantly between these groups.

Recent work has suggested that two summary scales can be derived from this measure: the Physical Component Summary Score (PCS) and the Mental Component Summary Score (MCS).[2 3] It has been suggested that summary scales derived from the eight dimensions of the SF-36 make it possible to reduce the number of statistical comparisons and reduce the role of chance in testing hypotheses about health outcomes, especially in the arena of clinical trials.[2] The Physical and Mental Summary Scales were derived from the dataset presented here using the procedures recommended by the developers.

To gain the physical and mental summary scores the data, which are effectively population norms for the SF-36, were factor analysed in accordance with the recommendations of the MCS/PCS developers, using principal components analysis and orthogonal rotation.[2] Each of the eight individual scales of the SF-36 was then standardised using a z score transformation using means and standard deviations calculated for all respondents in the OHLS-III dataset. Each z score was calculated by subtracting the OHLS mean from each individual respondents scale score and dividing the difference by the corresponding scales 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 standardised to a *t* score where the mean was set to 50 and the standard deviation to 10. The construct validity of the summary scores was assessed using the same criteria applied to the eight dimensions, namely comparing scores for men and women; manual and non-manual social class and reporting of long term illness or not.

The results reported here are based upon data gained from the third Oxford Healthy Life Survey (OHLS-III), undertaken during 1997. This was a postal survey in which the SF-36 together with questions on lifestyle and demographics were incorporated into a booklet. A covering letter, explaining the purpose of the

Item number	Version 1.0 *was*	Version 2 *is*
3, introduction	items	questions
4, introduction	—	"how much of the time" added
4, response choices	yes/no	all of the time/most of the time/some of the time/a little of the time/none of the time
5, introduction	—	"how much of the time" added
5, response choices	yes/no	all of the time/most of the time/some of the time/a little of the time/none of the time
5c	Didn't do work or other activities as carefully as usual	Didn't work or other activities less carefully than usual
9, response choices	Six choices, including "a good bit of the time"	Five choices, "a good bit of the time" dropped
9b	"a very nervous person"	"very nervous"
9h	"a happy person"	"happy"

Figure 1 Changes made to SF-36II from the original SF-36.

*Table 1 Social Class by Standard Occupational Classification, (the Registrar General Scheme) for OHLSIII and the general population (the latter in parentheses; data supplied by the Office of National Statistics, ONS: 10% sample of the 1991 Census). *Not all respondents replied to the question concerning their sex*

	All respondents (valid per cent)	*Men (valid per cent)*	*Women (valid per cent)*
I	5.8 (5.0)	9.5 (7.4)	2.9 (1.9)
II	32.0 (28.9)	34.8 (29.3)	30.0 (28.3)
IIIn	27.5 (23.6)	12.5 (11.3)	39.7 (39.0)
IIIm	15.8 (21.1)	27.2 (32.3)	6.4 (7.1)
IV	14.8 (15.6)	13.2 (14.9)	16.0 (16.3)
V	4.0 (5.8)	2.8 (4.6)	5.0 (7.3)
	n=7707* (n=2 302 728)	n=3419* (n=1 280 053)	n=4220* (n=1 022 675)

Data shown as percentages.

Table 2 Mean (SD) scores for the eight dimenions of the SF-36-II for the sample as a whole, and broken down by sex. Cronbach's alpha internal consistency reliability is reported in the second column

	Total		
Dimension	*Number*	*alpha*	*Mean (SD)*
Physical function	8561	0.92	87.99 (19.65)
Role - physical	8713	0.95	87.17 (22.01)
Pain	8789	0.85	78.80 (23.01)
General health	8620	0.80	71.06 (20.43)
Energy/vitality	8772	0.84	58.04 (19.60)
Social functioning	8776	0.85	82.77 (23.24)
Role - mental	8738	0.92	85.75 (21.18)
Mental health	8775	0.84	71.92 (18.15)

KEY POINTS

- The SF-36 Version II is a generic health status measure based on the widely used SF36.
- The SF36 Version II has improvements in wording and layout over the original instrument intended to improve its precision and reduce floor and ceiling effects.
- The results from this study indicate that dimensions of the SF-36 are more reliable and that floor and ceiling effects have indeed been reduced.
- We recommend that intending users of the SF36 utilise the new version of the instrument in subsequent research.

study, was sent with the questionnaire. For those who did not respond to the initial questionnaire a reminder postcard was mailed approximately three weeks later. If this elicited no response within three weeks then another questionnaire and covering letter was sent. The questionnaire booklet contained, in addition to the SF36-II, questions on the occupational class of the respondent and a question on whether or not the respondent had any longstanding illness.

The questionnaire booklet was mailed to 14 868 randomly selected subjects, unstratifed by age or sex, between the ages of 18–64, inclusive, from the General Practitioner Records held by the Health Authorities for Berkshire, Buckinghamshire, Northamptonshire, and Oxfordshire. A total of 1068 respondents were removed from the denominator because of incorrect postal addresses (n=1004), death (n=6), being outside of the specified age range (n=42) or because they were unable to read the questionnaire (n=16). The final denominator was, therefore, 13 800.

Results

Completed questionnaires were obtained from 8889 of 13 800 people originally contacted, giving a response rate of 64.4%. Of those who did return questionnaires 8801 (99.0%) of respondents answered the question relating to sex, of whom 3863 (43.4%) were male and 4938 (55.6%) were female. This is a slight over-representation of women as population estimates from the last census indicate that 55.6% of the working age population, for whom social class could be calculated, were male and 44.4% female (data provided by the Office for National Statistics[13]).

Social class information was provided by 7707 (86.7%) of the respondents to OHLS-III, and the breakdown of social classes is shown in table 1. The social class distribution broadly mirrors that of the UK population from 1991 Census data (provided by the Office for National Statistics[13]), which is also presented in table 1.

The eight dimensions of the SF-36-II were calculated using algorithms recommended by the developers. Descriptive statistics for the eight dimensions of the SF-36 and internal reliability are reported for the sample as a whole in table 2. All alpha statistics show good internal reliability. Furthermore the internal

Table 3 The eight dimensions of the SF-36 broken down by sex, social class, and reported chronic illness

	Men		*Women*		*Non-manual*		*Manual*		*Longstanding illness*		*No longstanding illness*	
Dimension	*Number*	*Mean (SD)*	*Number*	*Mean (SD)*	*Number*	*Mean (SD)*	*Number*	*Mean (SD)*	*Number*	*Mean (SD)*	*Number*	*Mean (SD)*
Physical function	3727	89.76 (18.78)	4755	86.66 (20.15)	4905	89.63 (17.03)	2546	86.45 (21.60)	3531	79.40 (24.34)	4962	94.03 (12.36)
Role - physical	3800	89.01 (21.09)	4828	85.83 (22.52)	4970	89.13 (19.71)	2607	85.13 (23.98)	3594	77.64 (27.66)	5052	93.89 (13.34)
Pain	3827	81.25 (22.21)	4877	76.97 (23.44)	4988	80.74 (21.27)	2640	76.35 (24.65)	3644	67.08 (25.60)	5078	87.13 (16.56)
General health	3773	70.86 (20.29)	4766	71.28 (20.54)	4913	72.68 (19.36)	2575	69.16 (21.40)	3556	60.77 (21.99)	4999	78.37 (15.61)
Energy/vitality	3823	60.81 (18.93)	4865	55.91 (19.85)	4990	58.48 (19.14)	2630	57.55 (20.31)	3630	51.17 (20.67)	5076	62.97 (17.21)
Social functioning	3810	84.71 (22.56)	4880	81.33 (23.62)	4989	84.33 (22.00)	2625	81.19 (24.40)	3642	75.05 (26.87)	5069	88.33 (18.37)
Role - mental	3811	88.08 (19.91)	4843	84.07 (21.79)	4983	87.09 (19.07)	2613	84.99 (22.75)	3613	80.36 (25.11)	5058	89.64 (16.85)
Mental health	3823	74.32 (17.24)	4868	70.05 (18.65)	4987	72.82 (17.24)	2636	71.16 (19.07)	3638	67.29 (19.69)	5073	75.29 (16.15)

Table 4 Factor score coefficients used to derive PCS and MCS summary scale scores from the SF36 and SF36II

	Old PCS coefficients	*Old MCS coefficients*	*New PCS coefficients*	*New MCS coefficients*
Physical functioning	0.418	−0.213	0.456	−0.227
Role physical	0.334	−0.087	0.362	−0.102
Bodily pain	0.366	−0.125	0.367	−0.130
General health	0.222	0.036	0.199	0.036
Energy/vitality	−0.017	0.286	−0.050	0.278
Social functioning	0.083	0.201	−0.028	0.272
Role emotional	−0.179	0.394	−0.110	0.329
Mental health	−0.200	0.444	−0.256	0.460

reliability statistics for the two role functioning dimensions are higher than those reported in similar studies using the original SF36, as was predicted would be the case by the developers. The developers claimed that the newly configured Role-Emotional and Role-Physical response formats would lead to fewer floor and ceiling effects and this was borne out in this study. One hundred and seventeen (1.32%) respondents scored zero on the Role-Physical dimension and 5004 (56.29%) scored 100. In the previous Oxford Health and Lifestyle Survey[7 12] the figures were 668 (7.16%) and 7066 (n=75.50%) respectively. Similarly the Role-Emotional Scores were 74 (0.83%) scoring zero and 4767 (53.63%) scoring 100 in this survey compared with 758 (8.12%) and 6724 (72.04%) in the previous survey.

Table 3 provide normative data, in the form of means and standard deviations, broken down by sex, social class (manual or non-manual), and whether respondents indicated having a longstanding illness. Overall, women reported poorer health on all dimensions of the SF-36 than men ($p<0.001$) except for general health perception (NS). A significant difference ($p<0.001$) was found on scores on all dimensions of the SF-36 between manual (classes IIIm, IV, and V) and non-manual respondents (classes I, II, and IIIn). The data broken down by those reporting longstanding illness and those reporting no such illness indicated significant differences between the groups, with those with longstanding illness gaining significantly lower scores than those who did not report longstanding illness ($p<0.001$, on all dimensions). These data provide evidence for the construct validity of the instrument. Data from the SF36-II were broken down by 10 year age groups, sex, reporting of chronic illness, and social class. This more detailed breakdown of the data can be accessed at http:\\hsru.dphpc.ox.ac.uk\sf36v2.htm or is available directly from CJ.

The factor analysis of the eight dimensions of the SF36-II in OHLS-III produced, as predicted, a two factor solution. Factor score coefficients are reported in table 4. The factor score coefficients are similar to those gained from the original SF-36 in the previous Oxford Healthy Lifestyles Survey.[14] The two factors accounted for a total of 71.43% of the variance.

Table 5 indicates scores gained on the PCS/MCS summary scales for the sample broken down by sub-groups. The data suggest the summary scores have appropriate construct validity (that is, they detect worse health for women than men, those in manual jobs than those in non-manual jobs and for those reporting a longstanding/chronic illness).

Discussion

The purpose of this paper was to introduce researchers to the new version of the SF36, the SF36-II. Evidence for the internal consistency of the domains of the UK SF-36-II, and the construct validity of the measure has been provided, as well as factor loadings that can be used to weight dimensions when calculating the UK PCS and MCS. The data reported in this paper seem to reflect the social class demographics of the general population and may therefore be of use as preliminary UK population norms for the SF36-II. As well as the data presented in this paper, more detailed breakdown is available on the web at http:\\hsru.dphpc.ox.ac.uk\sf36v2.htm or from CJ. For purposes of comparison this study adopts the same sampling methodology as the previous Oxford Health and Lifestyles Survey. Consequently this study includes only respondents of working age and it is not possible to make claims of the suitability of the measure in older age groups. Further research, specifically to investigate the applicability of the SF36-II in this important and growing part of the population, is required before it can be used confidently in this group.

The developers of the SF36-II suggest that changes to the response categories of the role-emotional and role-physical domains will increase their internal reliability consistency and reduce the floor and ceiling effects that have been reported in the literature. Comparing the data reported in this survey, from the Third Oxford Healthy Lifestyles Survey, with data from the Second Oxford Healthy Lifestyles Survey, which used an earlier version of the SF36, both of the developers claims seem to have been fulfilled. Such improvements are likely to increase the precision of the measure (that is, its ability to differentiate between groups) and its sensitivity to change. In the assessment of outcomes of treatment regimens

Table 5 PCS and MCS scores broken down by sex, social class, and reported longstanding/chronic illness

	Men		*Women*		*Non-manual*		*Manual*		*Longstanding illness*		*No longstanding illness*	
Dimension	*Number*	*Mean (SD)*	*Number*	*Mean (SD)*	*Number*	*Mean (SD)*	*Number*	*Mean (SD)*	*Number*	*Mean (SD)*	*Number*	*Mean (SD)*
PCS	3542	50.63 (9.41)	4442	49.54 (10.40)	4669	50.85 (9.15)	2377	48.93 (10.74)	3256	44.63 (12.16)	4741	53.64 (5.88)
MCS	3542	51.16 (9.34)	4442	49.17 (10.39)	4669	50.25 (9.68)	2377	49.93 (10.38)	3256	48.20 (11.04)	4741	51.28 (9.01)

such changes are to be welcomed, and are likely to further increase the uptake of this instrument in trials and other areas of outcomes research. Copies of the UK SF36-II are available from CJ. Intending users must inform the Medical Outcomes Trust of their intention to use the SF36-II and the purpose(s) of their project. Full details available upon request.

We wish to thank the Directors of Public Health for Berkshire, Buckinghamshire, Northamptonshire, and Oxfordshire Health Authorities for funding this study.
Conflicts of interest: none.

1 Ware J, Sherbourne C. The MOS 36-Item Short-Form Health Survey 1: conceptual framework and item selection. *Med Care* 1992;**30**:473–83.
2 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**:AS264–79.
3 Ware JE, Kosinski M, Keller S. *SF-36 physical and mental summary scales: a user's manual*. Boston, Massachusetts: The Health Institute, New England Medical Center, 1994.
4 Garratt AM, Ruta DA, Abdalla MI, *et al*. The SF 36 Health Survey Questionnaire: An Outcome Measure Suitable for Routine Use within The NHS? *BMJ* 1993;**306**:1440–4.
5 Shiely J-C, Bayliss MS, Keller SD, *et al*. *SF-36 Health Survey annotated bibliography*. 1st ed, 1998–1995. Boston, MA: The Health Institute, New England Medical Centre, 1997.
6 Ware JE, Snow KK, Kosinski M, *et al*. *SF-36 Health Survey: manual and interpretation guide*. Boston, Massachusetts: The Health Institute, New England Medical Center, 1993.
7 Jenkinson C, Layte R, Wright L, *et al*. *The UK SF-36: an analysis and interpretation manual*. Oxford: Health Services Research Unit, 1996.
8 Jenkinson C, McGee H. Patient assessed outcomes: measuring health status and quality of life. In: Jenkinson C, ed. *Assessment and evaluation of health and medical care*. Buckingham: Open University Press, 1997.
9 Jenkinson C. Evaluating the efficacy of medical treatment: possibilities and limitations. *Soc Sci Med* 1995;**41**:1395–401.
10 Ware JE, Kosinski M. Improvements in the content and scoring of the SF-36 Health Survey Version 2. SF-36 Users Site at http://www.SF-36.com/news/SF36–20.html
11 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrica* 1951;**16**:297–334.
12 Jenkinson C, Coulter A, Wright L. Short Form 36 (SF 36) health survey questionnaire: normative data for adults of working age. *BMJ* 1993;**306**:1437–40.
13 Office of Populations Censuses and Surveys. *Social Class and SEG. 1991 Economic Activity*. Vol 2. London: ONS, 1995.
14 Jenkinson C, Layte R, Lawrence K. Development and testing of the Medical Outcomes Study 36-item short form health survey summary scale scores in the United Kingdom. *Med Care* 1997;**35**:410–16.