

# An assessment of the construct validity of the SF-12 summary scores across ethnic groups

Crispin Jenkinson, Tarani Chandola, Angela Coulter and Stephen Bruster

## Abstract

**Background** The aim of the study was to determine the construct validity of the 12-item Short Form health survey questionnaire (SF-12) across ethnic groups in a large community sample of the United Kingdom.

**Methods** A postal survey was carried out in English using a questionnaire booklet, containing the SF-12 and a number of other items relating to experiences of chronic illness and utilization of health care services. The dataset was the National Survey of NHS Patients. The sample consisted of 1000 residents within each Health Authority in England who were randomly selected from the electoral registers, giving an initial sample of 100 000.

**Results** A total of 61 426 (61.4 per cent) questionnaires were returned; 94.3 per cent of respondents classified themselves as white and 5.7 per cent classified themselves as members of other ethnic groups. Construct validity of the SF-12 was assessed by comparing results from the two summary scores (the Mental Health Component Summary (MCS) score and the Physical Health Component Summary (PCS) score) with overall self-assessed health and limiting longstanding illness. Although there were generally consistent patterns of association between overall self-assessed health or limiting longstanding illness and the MCS and PCS scores in all the ethnic groups, there were significant differences between the MCS and PCS scores of Indians, Pakistanis and Bangladeshis who understood English fluently and those who did not. Furthermore, there were differences in the completion rates of the SF-12 between ethnic groups and a reversal of the general pattern of increasing MCS scores with increasing age in Bangladeshis.

**Conclusion** The results indicate that the use of the SF-12 to measure the health of ethnic minorities seems acceptable in most instances, but may prove problematic in those instances where respondents complete the questionnaire via an untrained translator, such as a friend or family member. The systematic differences in MCS and PCS scores between ethnic minorities who understood English fluently and those who did not suggest that the meaning of specific SF-12 items may change when informally translated. Future research using the SF-12 to measure the health status of ethnic minorities in the United Kingdom via postal surveys must include questions on whether respondents completed the questionnaires via informal translations. In general, those wishing to measure the health of members of ethnic groups who are unable to read English might consider using different techniques to gain the information from these groups.

**Keywords:** SF-12, ethnic groups, self-assessed health status, validity

## Introduction

The measurement of health status from the perspective of the patient has become a major aspect of health services evaluation.<sup>1</sup> Health surveys of the population,<sup>2</sup> surveys of patient experiences of health provision<sup>3</sup> and randomized controlled trials of treatments<sup>4</sup> now include measures of self-rated health. However, some researchers have expressed concern that subjective accounts of health are influenced by so many personal and cultural factors that they are likely to be unreliable measures.<sup>5</sup> Most notably, caution has been advocated in the interpretation of self-rated health when comparing different ethnic or cultural groups.<sup>6</sup> However, in large-scale surveys and trials it is inevitable that some respondents will be members of ethnic minority groups and it is important to be able to measure their responses accurately.

One solution to this problem is to translate questionnaires into the relevant languages taking care to retain the meaning, if not exact terminology of the original instrument. Indeed, the International Quality of Life Assessment (IQOLA) group has translated the SF-36 health survey into numerous languages using this approach, and has found such translations to be reliable and valid.<sup>7</sup> However, health surveys in the United Kingdom are usually conducted in English even though it is highly

<sup>1</sup>Pickering Institute Europe, King's Mead House, Oxpens Road, Oxford OX1 1RX.

<sup>2</sup>Health Services Research Unit, University of Oxford, Institute of Health Sciences, Old Road, Headington, Oxford OX3 7LF.

<sup>3</sup>Department of Epidemiology and Public Health, Royal Free and University College Medical School, University College London WC1E 6BT.

<sup>4</sup>Department of Public Health, University of Oxford, Oxford OX3 7LF.

**Crispin Jenkinson**, Director of Research<sup>1</sup> and Director of Health Care Outcomes Research<sup>2</sup>

**Tarani Chandola**,<sup>3</sup> Senior Research Fellow

**Angela Coulter**, Chief Executive<sup>1</sup> and Visiting Professor of Health Services Research<sup>4</sup>

**Stephen Bruster**,<sup>1</sup> Country Manager

Address correspondence to Crispin Jenkinson.

E-mail: Crispin.Jenkinson@PickeringEurope.ac.uk

likely that large-scale social surveys of the UK population will include respondents from ethnic minorities whose first language may not be English. It is therefore important to assess whether such respondents are likely to complete questionnaires sent to them in English in a systematically different way from that of the general population. If such differences were to be found this might reflect different interpretations between ethnic groups of questions and response options. On the other hand, if respondents complete questionnaires in a similar way to the dominant ethnic group then this may obviate the need for specific translations, and indicate that surveys that do not specifically target ethnic minorities with specifically translated questionnaires can still produce valid and reliable information. The Fourth National Survey of Ethnic Minorities<sup>8,9</sup> used similar analyses to demonstrate the validity of using self-assessments of morbidity across ethnic groups.

The 'Short Form' measures developed from the RAND Health Insurance Experiment and the Medical Outcomes Study are probably the most popular generic health status measures currently in use. The SF-12 is a 12-item questionnaire that can be used to measure physical and mental health.<sup>10-13</sup> The instrument produces two summary scores – the Physical Component Summary (PCS) and the Mental Health Component Summary (MCS) – and has been validated for use in the United Kingdom and widely used as a short measure of health status in social surveys,<sup>3,14</sup> studies investigating the impact of specific diseases upon patients<sup>15</sup> and treatment trials.<sup>16</sup> For ease of interpretation scores are standardized to population norms, with the mean score set at 50 (SD 10). To date no data have been reported on the operating characteristics of this instrument in different ethnic groups.

The purpose of the analysis reported here is to establish whether the SF-12 follows expected patterns of results for different ethnic groups. Consequently, this paper examines the construct validity of the measure across ethnic groups. Construct validity refers to the ability of a questionnaire to detect previously hypothesized differences or associations between variables. The following hypotheses, which have been extensively tested in the white population for the SF-12 and SF-36, were tested.

- (1) Female respondents would have worse self-reported health on both the PCS and MCS than males.
- (2) Physical health status as reflected in PCS scores would be higher (i.e. better) for respondents aged under 45 as opposed to those 45 and over.
- (3) Overall reports of health status would be reflected in both the PCS and MCS. One of the items of the SF-12 asks respondents to assess their health overall. This item has been used on its own in a large number of surveys, and has been shown to be a valid, although somewhat broad, assessment of health across ethnic groups.<sup>17</sup> Previous evidence has indicated that scores on dimensions of the SF-36, the instrument from which the SF-12 was developed, follow clear linear trends reflecting judgements of overall health.<sup>18</sup>

- (4) Those reporting chronic illness would gain lower scores on both dimensions than those who do not.

Furthermore, the study examines whether those who report having had the survey questions translated for them or whether those who do not have English as their first language report worse or better health than others in the same ethnic group when controlling for self-assessed health or chronic illness, sex and age. This is an important issue, as the informal translation of the SF-12 questions or the non-comprehension of specific SF-12 questions by some members of ethnic minorities who do not have English as their first language may result in inaccurate estimates of the SF-12 scores for those groups.

## Methods

The dataset used for the analysis was the National Survey of NHS Patients – General Practice. The sample was drawn from electoral registers published in 1998. It comprised 1000 names and addresses of electors selected within each of the 100 Health Authorities in England, giving a total sample of 100 000. In early October 1998 each of the selected respondents was sent a questionnaire together with a letter, a stamped addressed pre-paid reply envelope and an information sheet outlining the purpose of the study. The questionnaire, in English, sought information on demographics (including ethnic group), self-reported chronic illness, access and attitudes towards general practitioner services, and subjective health status, the last measured on the SF-12. The measure includes 12 questions including a global question of perceived health status, to which respondents indicate whether they think their overall health is excellent, very good, good, fair or poor. Finally, respondents were asked if they had had any help completing the questionnaire and what language they spoke most often.

## Statistical analysis

SF-12 scores were calculated using published algorithms to weight responses to each of the 12 items on the questionnaire.<sup>19,20</sup> To be able to derive PCS and MCS scores all items must be completed. Data completeness was assessed by frequencies of incomplete responses by item and for the PCS and MCS overall. Differences between scores on the PCS and MCS for male and female respondents were tested for significance using non-parametric Mann-Whitney *U*-tests, as were differences between older and younger respondents on the PCS. Kruskal-Wallis tests were performed on the data to determine whether there were significant trends in both PCS and MCS scores when scores were broken down by the five-point single item of overall self-assessed health, and also on the three-point item on chronic illness. Multiple linear regression analyses with the PCS and MCS scores as the dependent variables, controlling for age, sex and overall self-assessed health, were undertaken for all ethnic minority groups to determine whether those who

reported needing translations of the health survey or who did not have English as their first language had systematically different scores from those who did not need such translations or who had English as their first language. The association of an interaction term of overall self-assessed health and translation/first language with the PCS and MCS scores tests the hypothesis that within the same levels of self-assessed health, there are no significant differences in the PCS or MCS scores between respondents who understand English fluently and those who do not. If the interaction term is significant, then there are significant differences in the PCS or MCS scores between the two types of respondents.

## Results

Table 1 indicates the levels of incomplete data across ethnic groups. There was considerable variation in the proportion of respondents in each of the ethnic groups for whom PCS and MCS scores could be calculated, with a quarter of the Bangladeshi sample leaving at least one item incomplete. Descriptive statistics on the PCS and MCS, broken down by sex and ethnic group, are reported in Table 2. For all groups, except Chinese respondents, differences between the sexes were in the hypothesized direction, with male respondents reporting better self-rated health on both the PCS and MCS than females. Mann–Whitney tests were statistically significant for comparisons ( $p < 0.001$ ) except for Black Caribbean (PCS only) and Chinese groups. Results were broken down into two age groups, those aged 45 or less, and those over 45. PCS scores were lower for all those in the over-45 groups for all ethnic groups, and these results were statistically significant (Mann–Whitney tests,  $p < 0.02$ ) for all groups except Chinese respondents (see Table 3).

Kruskal–Wallis tests indicated that there were clear linear trends on both dimensions of the SF-12 in relation to the single item overall assessment of self-assessed general health (see Table 4). Similarly, clear trends on both the PCS and MCS were found in relation to reported chronic illness, with those reporting limiting longstanding illness gaining lower scores (i.e. indicating worse health) than those simply reporting longstanding illness. Across all ethnic groups the highest scores were gained from those who reported no chronic illness, although it is striking to note that all ethnic minority groups scored below the population norm for the MCS (see Table 5).

Table 6 shows the cross tabulation of whether the respondent required translation of the questions and whether English is the first language of respondents by the different ethnic groups. Overall only 0.4 per cent ( $n = 260$ ) of the entire sample indicated they required translation. There were a small number of 'Whites' who required translation ( $n = 54$ , 0.1 per cent of white respondents) whereas larger proportions of Indians, Pakistanis, Bangladeshis and Chinese required some translation of the questions. No Black Caribbeans and only four Black Africans required some translation of the questions and so they were excluded from further analyses with translation as a

**Table 1** Missing data (%) per item and for the PCS and MCS

	White ( <i>n</i> = 55537)	Black Caribbean ( <i>n</i> = 551)	Black African ( <i>n</i> = 253)	Indian ( <i>n</i> = 1033)	Pakistani ( <i>n</i> = 515)	Bangladeshi ( <i>n</i> = 168)	Chinese ( <i>n</i> = 175)
1. Own assessment of state of health	0.8	0.7	0.7	0.4	0.3	0.6	0.4
2. Limited in undertaking moderate activities	2.0	3.8	3.7	3.5	1.8	6.4	0.9
3. Limited in climbing several flights of stairs	3.2	5.7	4.6	3.7	6.0	8.6	2.0
4. Accomplished less because of physical health	5.2	7.8	4.4	6.0	7.3	7.8	2.6
5. Limited because of physical health	7.7	11.6	5.8	5.4	8.3	14.0	3.4
6. Accomplished less because of emotional problems	4.5	6.9	5.3	5.3	8.0	7.6	3.6
7. Not as careful as usual because of emotional problems	8.7	12.7	6.2	6.7	9.4	10.4	5.2
8. Pain affected work over the last 4 weeks	2.4	3.5	6.1	1.8	1.8	5.4	6.0
9. How often felt calm and peaceful in last 4 weeks	3.1	3.4	5.9	3.1	2.8	3.3	6.2
10. How often had a lot of energy in last 4 weeks	3.3	4.0	5.9	3.8	3.0	3.7	5.3
11. Felt downhearted and low in last 4 weeks	2.8	4.5	5.4	3.0	3.8	4.9	5.7
12. Physical or emotional health interfered with social life	1.8	1.9	4.5	1.5	2.8	2.6	4.8
% for whom PCS & MCS could not be calculated	15.9	20.5	20.0	14.3	18.3	25.0	11.43

**Table 2** PCS and MCS scores (mean and SD) for men and women broken down by ethnic group

	PCS		MCS	
	Male	Female	Male	Female
White	50.41 (10.03)	49.43 (10.96)	50.14 (10.07)	47.74 (10.96)
<i>n</i>	23546	23167	23546	23167
Black Caribbean	50.87 (9.41)	48.69 (11.68)	49.88 (9.90)	46.24 (12.64)
<i>n</i>	189	249	189	249
Black African	51.17 (9.60)	48.89 (10.27)	50.20 (9.60)	44.74 (12.00)
<i>n</i>	103	102	103	102
Indian	49.42 (9.97)	46.82 (11.30)	49.25 (9.97)	45.05 (12.56)
<i>n</i>	498	387	498	387
Pakistani	48.47 (9.35)	45.91 (12.28)	46.85 (10.35)	44.06 (11.81)
<i>n</i>	227	193	227	193
Bangladeshi	49.37 (9.14)	45.88 (10.78)	46.15 (10.35)	44.23 (11.74)
<i>n</i>	67	59	67	59
Chinese	50.40 (7.26)	50.56 (7.97)	49.35 (8.08)	48.33 (10.71)
<i>n</i>	77	78	77	78

**Table 3** PCS scores broken down by age (respondents under 45 and respondents 45 and over) and ethnic group

	18–44		45+	
	Mean (SD)	<i>n</i>	Mean (SD)	<i>n</i>
White	53.04 (7.75)	25649	46.09 (12.08)	20799
Black Caribbean	53.12 (7.65)	292	42.57 (12.76)	142
Black African	51.84 (7.91)	155	43.66 (13.51)	45
Indian	51.14 (8.53)	637	41.37 (11.73)	242
Pakistani	49.63 (9.26)	327	38.49 (11.85)	82
Bangladeshi	50.27 (7.81)	102	36.48 (11.72)	21
Chinese	51.25 (6.92)	106	48.80 (8.61)	46

variable. English is not the language spoken most often by the majority of Indian, Pakistani, Bangladeshi and Chinese respondents. Only 10 Black Caribbeans did not have English as their first language and so Black Caribbeans were excluded from further analysis with this variable.

Table 7 reports the estimates and standard errors of the SF-12 PCS scale (as the dependent variable) in a multiple linear regression with age, sex, self-assessed health and translation as independent variables and an interaction term of self-assessed health with translation. Separate regression models were fitted for different ethnic groups. As shown in Table 3, increasing age is associated with poorer SF-12 PCS scores in all ethnic groups, although the trend for Chinese respondents is not significant. Furthermore, similar to the results in Table 2, female respondents in all ethnic groups except Chinese have poorer SF-12 PCS scores compared with male respondents. As shown in Table 4, respondents from all ethnic groups who report poorer self-rated health have lower SF-12 PCS scores.

In all the ethnic groups, there is no evidence for differences between respondents who required and did not require translations of the questions in terms of their SF-12 PCS. Furthermore,

in all the ethnic groups, there is no evidence that the association between self-rated health and the SF-12 PCS scores was different between those respondents who required translations and those who did not require translations.

Table 8 repeats the analysis in Table 7 for the SF-12 MCS scores. However, unlike the results in Table 7, the results in Table 8 show some evidence that some ethnic minorities show different associations with the SF-12 MCS scores compared with the majority population. For example, unlike the general pattern of increasing MCS scores with age for Whites, younger Bangladeshis aged 25–44 have the highest MCS scores. Furthermore, there is some evidence that Pakistani respondents who required translations of questions systematically completed their SF-12 MCS questions in a different way compared with other ethnic groups. Pakistanis who required translations and who had poorer self-assessed health had lower MCS scores compared with Pakistanis with poorer self-assessed health who did not require translations. This interaction of self-rated health with translation was non-significant in all the other ethnic groups, including Whites.

Similar analyses were carried out to those reported above but

**Table 4** PCS and MCS scores (mean and SD) broken down by overall self-reported health and ethnic group

PCS					MCS					
	Excellent	Very good	Good	Fair	Poor	Excellent	Very good	Good	Fair	Poor
White <i>n</i>	56.60 (4.42) 6623	54.18 (5.78) 18012	49.27 (8.62) 14565	37.62 (11.37) 6007	25.05 (8.28) 1506	53.13 (7.78) 6623	50.72 (9.16) 18012	48.23 (10.65) 14565	44.11 (11.99) 6007	35.65 (12.00) 1506
Black Caribbean <i>n</i> = 54	55.98 (5.12) 54	54.45 (6.04) 151	50.25 (8.49) 137	40.54 (11.70) 77	25.65 (6.92) 19	53.29 (6.28) 6623	49.21 (6.04) 151	48.25 (10.79) 137	44.18 (13.15) 77	32.63 (9.58) 19
Black African <i>n</i> = 50	54.94 (6.12) 50	52.18 (6.85) 68	52.02 (6.77) 52	38.19 (10.87) 27	26.98 (7.90) 7	50.76(7.79) 50	50.24 (10.45) 68	46.58 (11.05) 52	39.54 (11.53) 27	35.79 (10.41) 7
Indian <i>n</i>	56.48 (4.18) 88	53.79 (5.65) 286	49.22 (8.11) 292	39.27 (9.96) 170	27.01 (7.81) 49	52.92 (9.20) 88	49.94 (10.56) 286	46.65 (11.70) 292	45.42 (11.30) 170	34.12 (9.64) 49
Pakistani <i>n</i>	53.50 (6.88) 41	52.23 (7.96) 99	50.06 (7.36) 166	40.10 (10.11) 87	27.04 (9.18) 9.18	52.89 (9.01) 41	46.50 (10.12) 99	46.65 (10.78) 166	43.36 (10.00) 87	32.30 (10.39) 9.18
Bangladeshi <i>n</i>	53.64 (6.57) 10	55.26 (4.09) 36	49.07 (6.91) 45	40.03 (9.02) 24	30.58 (5.93) 12	51.77 (5.82) 10	44.92 (9.98) 36	48.97 (10.88) 45	41.36 (11.35) 24	34.68 (7.25) 12
Chinese <i>n</i>	54.24 (4.71) 13	53.04 (5.80) 51	49.97 (7.34) 67	44.62 (9.29) 25	– 0	53.71 (5.44) 13	52.32 (8.14) 51	46.69 (9.20) 67	44.98 (11.31) 25	– 0

**Table 5** PCS and MCS scores (mean and SD) broken down by reported chronic illness and ethnic group

	PCS scores for those reporting				MCS scores for those reporting			
	limiting long-standing illness	longstanding illness does not limit daily activities	no longstanding illness	no longstanding illness that does not limit daily activities	limiting long-standing illness	longstanding illness does not limit daily activities	no longstanding illness	no longstanding illness that does not limit daily activities
White <i>n</i>	35.13 (11.77) 8816	50.43 (7.69) 10168	54.48 (5.45) 27580	–	43.98 (12.87) 8816	49.64 (10.16) 10168	50.30 (9.38) 27580	–
Black Caribbean <i>n</i>	34.89 (13.36) 87	51.37 (6.16) 86	53.94 (5.57) 264	–	40.69 (14.32) 87	49.25 (11.16) 86	49.69 (9.86) 264	–
Black African <i>n</i>	32.81 (8.52) 33	52.24 (5.54) 51	53.88 (6.28) 119	–	39.24 (12.00) 33	49.43 (9.53) 51	48.82 (10.36) 119	–
Indian <i>n</i>	34.20 (10.35) 166	46.86 (9.09) 146	52.75 (6.76) 568	–	42.04 (11.84) 166	45.93 (12.17) 146	49.28 (10.86) 568	–
Pakistani <i>n</i>	35.48 (11.01) 99	48.43 (8.04) 56	51.48 (7.65) 265	–	39.04 (11.59) 99	44.68 (9.17) 56	48.20 (10.28) 265	–
Bangladeshi <i>n</i>	37.50 (9.88) 22	48.95 (12.88) 23	50.21 (7.12) 80	–	39.97 (6.71) 22	43.54 (9.34) 23	47.19 (11.91) 80	–
Chinese <i>n</i>	41.99 (10.06) 17	47.02 (8.21) 30	52.81 (5.41) 108	–	41.94 (9.47) 17	51.94 (8.56) 30	49.06 (9.25) 108	–

**Table 6** Cross tabulation of whether the respondent required translation of the questions by ethnicity

	White	Black Caribbean	Black African	Indian	Pakistani	Bangladeshi	Chinese
<i>Language spoken most often</i>							
English	99.3	97.6	64.5	41.7	31.5	12.3	44.4
Non-English	0.7	2.4	35.5	58.3	68.5	87.7	55.6
<i>n</i>	46258	423	197	870	400	122	151
<i>Someone translated the questions into my own language</i>							
No	99.9	100.0	98.0	94.6	89.3	86.9	87.4
Yes	0.1	0.0	2.0	5.4	10.8	13.1	12.6

Cells are column percentages.

**Table 7** Estimates and standard errors of the SF-12 PCS scale in a multiple linear regression with sex, age, self-assessed health, translation and an interaction of self-assessed health with translation in different ethnic groups

Explanatory variable	White		Indian		Pakistani		Bangladeshi		Chinese	
	B	SE	B	SE	B	SE	B	SE	B	SE
<i>Age</i>										
Under 25 <sup>a</sup>	0.0		0.0		0.0		0.0		0.0	
25–34	–0.5*	0.1	–0.3	0.8	–1.8	1.1	2.2	1.6	–2.0	1.8
35–44	–1.0*	0.1	–1.8*	0.8	–4.1*	1.4	2.1	2.1	–0.6	1.7
45–54	–2.4*	0.1	–5.6*	0.9	–7.7*	1.7	–7.6*	2.7	–2.9	1.9
55–64	–5.4*	0.1	–8.9*	1.2	–13.2*	1.9	–10.7*	3.3	–0.9	2.3
65–74	–8.0*	0.1	–13.5*	1.5	–7.2*	2.8	–20.3*	5.6	–4.5	4.7
75 and over	–12.4*	0.2	–8.8*	2.5	–19.9*	4.4	9.0	6.3	3.2	7.7
<i>Sex</i>										
Male <sup>a</sup>	0.0		0.0		0.0		0.0		0.0	
Female	–0.4*	0.0	–1.5*	0.5	–1.4	0.9	–1.0	1.3	0.6	1.2
<i>Self-assessed health</i>										
Excellent to very good <sup>a</sup>	0.0		0.0		0.0		0.0		0.0	
Good to poor	–8.7*	0.0	–7.8*	0.6	–5.2*	1.0	–9.1*	1.5	–4.6*	1.3
<i>Someone translated the questions into my own language</i>										
No <sup>a</sup>	0.0		0.0		0.0		0.0		0.0	
Yes	–0.6	1.8	–0.1	3.6	2.1	4.1	–4.8	3.6	–0.4	3.9
<i>Interaction of self-rated health with translation</i>										
Excellent to very good health or no translation <sup>a</sup>	0.0		0.0		0.0		0.0		0.0	
Good to poor health with translation	–3.7	2.5	–4.4	3.9	–7.5	4.4	4.3	4.4	0.9	4.4

<sup>a</sup>Reference category.\* $p < 0.05$ .

using ‘language spoken most often’ as a variable instead of whether the respondent required translation (data not reported but available from C.J.). Indians with poorer self-assessed health and who did not have English as their first language had lower PCS scores compared with Indians with poorer self-rated health who had English as their first language. In all the other ethnic groups, the interaction of self-assessed health with English as a first language was non-significant. Pakistanis with poorer self-assessed health and who did not have English as their first language had lower MCS scores compared with Pakistanis with poorer self-assessed health who had English as their first language. Bangladeshis with poorer self-assessed

health and who did not have English as their first language had higher MCS scores compared with Bangladeshis with poorer self-assessed health who had English as their first language.

The above analyses were repeated using ‘limiting longstanding illness’ as an explanatory variable in the regression models instead of the binary self-assessed health. Similar results were found in that the interaction between limiting longstanding illness and ‘translation required’ or ‘English as a first language’ was significantly associated with the PCS scores for Indians, and was significantly associated with the MCS scores for Pakistanis and Bangladeshi respondents (analyses not shown but available from C.J.).



**Table 8** Estimates and standard errors of the SF-12 MCS scale in a multiple linear regression with sex, age, self-assessed health, translation and an interaction of self-rated health with translation in different ethnic groups

Explanatory variable	White		Indian		Pakistani		Bangladeshi		Chinese	
	B	SE	B	SE	B	SE	B	SE	B	SE
<i>Age</i>										
Under 25 <sup>a</sup>	0.0		0.0		0.0		0.0		0.0	
25–34	0.0	0.1	0.4	1.0	–1.6	1.3	5.6*	2.2	2.1	2.2
35–44	0.0	0.1	1.5	1.1	0.7	1.7	13.6*	3.0	3.7	2.0
45–54	1.4*	0.1	2.8*	1.3	–4.2*	2.0	–2.6	3.8	4.1	2.2
55–64	3.8*	0.1	0.7	1.5	–3.8	2.3	3.8	4.5	6.0*	2.8
65–74	5.4*	0.2	6.1*	2.1	3.7	3.4	–4.0	7.7	12.4*	5.7
75 and over	4.1*	0.2	4.2*	3.4	2.3	5.2	10.4	8.7	12.7*	9.2
<i>Sex</i>										
Male <sup>a</sup>	0.0		0.0		0.0		0.0		0.0	
Female	–2.3*	0.0	–3.6*	0.7	–2.9*	1.1	0.0	1.9	–0.3	1.4
<i>Self-assessed health</i>										
Excellent to very good <sup>a</sup>	0.0		0.0		0.0		0.0		0.0	
Good to poor	–5.9*	0.0	–6.2*	0.8	–2.9*	1.1	–2.2	2.1	–7.4*	1.5
<i>Someone translated the questions into my own language</i>										
No <sup>a</sup>	0.0		0.0		0.0		0.0		0.0	
Yes	–4.4*	2.1	–3.8	4.8	10.9*	4.8	0.4	5.0	1.8	4.7
<i>Interaction of self-assessed health with translation</i>										
Excellent to very good health or no translation <sup>a</sup>	0.0		0.0		0.0		0.0		0.0	
Good to poor health with translation	4.8	3.0	6.8	5.1	–14.3*	5.2	1.1	6.1	1.0	5.3

<sup>a</sup>Reference category.\* $p < 0.05$ .

## Discussion

This study provides some evidence for the validity of the SF-36 in ethnic minority groups. For example, when two dimensions of the SF-12 are compared with a relatively simple and easily understood single item of overall perceived health a linear trend is detected on all dimensions, with patients claiming worse general health gaining lower (i.e. worse health status) scores on the summary scores. This trend is seen across all ethnic groups in the study. In the absence of a ‘gold standard’ to evaluate subjective health status measures such simple single items have been suggested as providing standards by which to judge such questionnaires, and have been used for this purpose in general population surveys.<sup>21</sup> Such simple ‘gold-standards’ may seem to call into question the whole value of utilizing longer form instruments, such as the SF-12. However, the advantages of utilizing more complex measures is that they have greater precision, can provide specific dimension scores and are less prone to floor and ceiling effects, and hence more sensitive to minor differences in health status.<sup>22,23</sup> Nevertheless, although many of the data indicate that the SF-12 operates in a similar fashion for ethnic minorities as it does for English respondents, they also suggest that the SF-12 may not be appropriate for some British ethnic minorities who do not have English as their first language and/or who may require translation of English language health

questionnaires. There were differences in the SF-12 PCS and MCS scores between respondents who reported they did and those who reported they did not understand English fluently for Indians, Pakistanis and Bangladeshis even when controlling for age, sex and self-assessed overall health or chronic illness. This suggests that the meanings of specific items of the SF-12 may change when informally translated from English to languages such as Hindi, Urdu, Punjabi, Sindhi, Pushtu, Bengali or Sylheti. Such systematic variations could undermine the construct validity of the use of the English version of the SF-12 to measure the health status of certain members of British ethnic minorities. Consequently, other strategies for measuring the health of people who do not speak English may need to be explored. Given the relatively small number requiring translation, in-depth interviewing and/or focus groups may be a more reliable way of assessing health in instances where English is not spoken.

Furthermore, the differences in completion rates of the SF-12 by ethnic groups as well as the reversal of the association between age and MCS scores in Bangladeshis (younger Bangladeshis aged 25–44 had the highest MCS scores, which is a reversal of the general pattern of increasing MCS scores with age) contribute to the debate about the validity of the English version of the SF-12 to measure the health status across all British ethnic minorities. It must be borne in mind that the SF-12 originated in the United States, and was primarily developed

on the assumptions of the white US population. Although the instrument has been translated for use in other countries little thought has been given to the use of the instrument with different cultural groups within one country.<sup>24</sup> Terminology and understanding of questions may differ across cultures, and this may account for the differences in, for example, completion rates found in this study. Furthermore, questions in one language may mean different things to different groups of respondents; consequently, it is the job of the questionnaire translator not to simply translate the item wording, but to retain the meaning across cultures. Score differences and variations in completion rates found in this study may thus be due to questions having different meanings to some ethnic groups. As Kleinman<sup>25</sup> has noted, the application of a measure in one group when designed for another may produce data of limited value.

In conclusion, the results indicate that the use of an English language version of the SF-12 to measure the health of ethnic minorities is promising but not unproblematic. Most importantly, the systematic differences in MCS and PCS scores between ethnic minorities who understood English fluently and those who did not suggest that the meaning of specific SF-12 items may change when informally translated. Future research using the SF-12 to measure the health status of ethnic minorities in the United Kingdom may need to consider either translating the health questionnaire into the major South Asian languages using appropriate methods to retain the cultural equivalence of the translated measure with the original, or at the very least include questions determining whether respondents had help in translating the measure. Undertaking social survey research on the population assuming that a questionnaire is understood in the same manner by all potential respondents, as is implicitly the case with many social surveys, means that the accurate health status of individuals may not be measured.

## References

- Jenkinson C, McGee H. *Measuring health status: a brief but critical introduction*. Oxford: Radcliffe Medical Press, 1998.
- Wright L, Harwood D, Coulter A. *Oxford health and lifestyles survey*. Oxford: Health Services Research Unit, 1992.
- Airey C, Erens B. *National survey of NHS patients. General practice, 1998*. London: NHS Executive, 1999.
- Spilker B. *Quality of life and pharmacoeconomics in clinical trials*. Philadelphia, PA: Lippincott, Wilkin and Watkins, 1995.
- Angel R, Gronfein W. The use of subjective information in statistical models. *Am Sociol Rev* 1988; **53**(June): 464–473.
- Jylha M, Guralnik J, Ferrucci L, Jokela J, Heikkinen E. Is self-rated health comparable across cultures and genders? *J Gerontol: B Psychol Sci Social Sci* 1998; **55**(3): S144–S152.
- Gandek B, Ware JE, Aaronson NK, et al. Cross validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA project. *J Clin Epidemiol* 1998; **51**: 1171–1178.
- Modood T, Berthoud R, Lakey J, et al. *Ethnic minorities in Britain: diversity and disadvantage*. London: Policy Studies Institute, 1997.
- Nazroo J. *The health of Britain's ethnic minorities*. London: Policy Studies Institute, 1997.
- Ware J, Sherbourne C. The MOS 36-Item Short-Form Health Survey 1: conceptual framework and item selection. *Med Care* 1992; **30**: 473–483.
- 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.
- 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.
- 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.
- Petersen S, Stewart-Brown S, Peto V. *Health and lifestyles in four counties*. Oxford: Health Services Research Unit, 1998.
- Lyons RA, Lo SV, Littlepage BN. Comparative health status of patients with 11 common illnesses in Wales. *J Epidemiol Commun Hlth* 1994; **48**: 388–390.
- Jenkinson C, Gray A, Doll H, et al. Evaluation of index and profile measures of health status in a randomised controlled trial: comparison of the SF-36, EuroQol and disease specific measures. *Med Care* 1997; **35**: 1109–1118.
- Chandola T, Jenkinson C. Validating self rated health in different ethnic groups. *Ethnicity Hlth* 2000; **5**: 151–159.
- Jenkinson C, Wright L, Coulter A. Criterion validity and reliability of the SF-36 in a population sample. *Qual Life Res* 1994; **3**(1): 7–12.
- Ware J, Kosinski M, Keller SD. *How to score the SF-12 physical and mental summary scores*. Boston, MA: The Health Institute, New England Medical Center, 1995.
- 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.
- Jenkinson C, Wright L, Coulter A. Criterion validity and reliability of the SF-36 in a population sample. *Qual Life Res* 1994; **3**: 7–12.
- Bindman AB, Keane D, Lurie N. Measuring health changes among severely ill patients: the floor phenomenon. *Med Care* 1990; **28**: 1142–1152.
- Ware JE, Nelson EC, Sherbourne CD, Stewart AL. Preliminary tests of a 6-item general health survey: a patient application. In: Stewart AL, Ware JE, eds. *Measuring functioning and well-being: the medical outcomes study approach*. London: Duke University Press, 1992.
- Staniszewska S, Ahmed L, Jenkinson C. The conceptual validity and appropriateness of using health-related quality of life measures in minority ethnic groups. *Ethnicity Hlth* 1999; **4**: 51–63.
- Kleinman A, Eisenberg L, Good B. Culture, illness and care. Clinical lessons from anthropologic and cross cultural research. *Ann Intern Med* 1978: 251–258.

Accepted on 16 February 2001