

Brief communication

The Short Form Health Survey (SF-36): Translation and validation study of the Iranian version

Ali Montazeri¹, Azita Goshtasebi¹, Mariam Vahdaninia¹ & Barbara Gandek²

¹*Iranian Institute for Health Sciences Research, Tehran, Iran (E-mail: ali@jdcord.jd.ac.ir);* ²*Health Assessment Lab, Waltham, MA, USA*

Accepted in revised form 22 June 2004

Abstract

This was a large population-based study to develop and validate the Iranian version of the Short Form Health Survey (SF-36) for use in health related quality of life assessment in Iran. A culturally comparable questionnaire was developed and pilot tested. Then, the Iranian version of the SF-36 was administered to a random sample of 4163 healthy individuals aged 15 years and over in Tehran. The mean age of the respondents was 35.1 (SD = 16.0) years, 52% were female, mostly married (58%) and the mean years of their formal education was 10.0 (SD = 4.5). Reliability was estimated using the internal consistency and validity was assessed using known groups comparison and convergent validity. In addition factor analysis was performed. The internal consistency (to test reliability) showed that all eight SF-36 scales met the minimum reliability standard, the Cronbach's α coefficients ranging from 0.77 to 0.90 with the exception of the vitality scale ($\alpha = 0.65$). Known groups comparison showed that in all scales the SF-36 discriminated between men and women, and old and the young respondents as anticipated (all p values less than 0.05). Convergent validity (to test scaling assumptions) using each item correlation with its hypothesized scale showed satisfactory results (all correlation above 0.40 ranging from 0.58 to 0.95). Factor analysis identified two principal components that jointly accounted for 65.9% of the variance. In general, the Iranian version of the SF-36 performed well and the findings suggest that it is a reliable and valid measure of health related quality of life among the general population.

Key words: Health status, Iran, Population-based study, Quality of life, The SF-36

Introduction

The Short Form Health Survey (SF-36) is a well-known generic health-related quality of life instrument that has been developed in the United State of America and translated into a variety of languages. Psychometric analyses of the translated versions provide evidence that the SF-36 is a reliable and valid measure in multiple populations. [1–7]. Although cross-cultural studies are time consuming, it is suggested that there is increasing need for international standards to measure health status in a manner that allows comparison across countries, but which also are relevant within individual cultures. Thus, it was decided to devel-

op the SF-36 Iranian version to respond to this increasing demand and provide a validated generic quality of life instrument. In Iran there are increasing demand by researchers and health professionals to include a validated quality of life questionnaire in their research or clinical investigations. In Asia a few studies have been carried out to translate and culturally adapt the SF-36 into different languages [8–12], but at present there is no an agreed or a validated generic quality of life instrument in Iran. This paper reports the development of the Iranian version of the SF-36 Health Survey and the results of its psychometric testing among the general population in Tehran, the capital of Iran.

Methods

The questionnaire

The SF-36 is a general quality of life instrument that measures eight health related concepts: physical functioning (PF-10 items), role limitations due to physical problems (RP-4 items), bodily pain (BP-2 items), general health perceptions (GH-5 items), vitality (VT-4 items), social functioning (SF-2 items), role limitations due to emotional problems (RE-3 items), and perceived mental health (MH-5 items). In addition a single item that provides an indication of perceived change in general health status over a one-year period (health transition) is also included in the SF-36 [13].

Translation

Permission was asked from the International Quality of Life Assessment (IQOLA) Project and the IQOLA translation methodology was followed to translate the SF-36 from English into Persian, the Iranian language [14]. Two independent health professionals translated the items and two others translated the response categories. Then a consolidated forward version was produced. This questionnaire then was backward translated into English by two professional translators to check for differences between the Iranian version and the original questionnaire. After a careful review and cultural adaptation few changes have been made and the provisional version of the questionnaire was provided. In general there were no difficulties in translating response categories but in items regarding activities bowling and playing golf have been changed to light sport activities, mile has been changed to kilometer, and walking one block or walking several blocks have been changed to walking one alley or several alleys to refer to a similar distance in Iranian language. Subsequently the provisional forward translated questionnaire was pilot tested and administered to a sample of 50 healthy individuals. The mean time to complete the SF-36 was 9.6 (SD = 6.6) min, 70% stated that did not have any difficulties completing the questionnaire and 98% indicated that there were no upsetting questions. The results of pilot testing were reported to the IQOLA Project Director and

after her review a few more changes have been made. There were difficulties in translation of feelings such as 'full of pep', 'felt so down' and 'felt downhearted and blue'. These all carefully converted to Iranian equivalence and the final version was provided and used in this study.

Data collection and statistical analysis

This was a population-based study and the SF-36 was administered to a random sample of healthy individuals aged 15 years and over living in Tehran. To select a representative sample of the general population a stratified multi-stage area sampling was applied. Every household within 22 different districts in Tehran had the same probability to be sampled. A team of trained interviewers collected data and all participants were interviewed in their home. According to the IQOLA Project to test psychometric properties of the SF-36 Iranian version several tests were performed [15, 16]. To test reliability the internal consistency for each scale was estimated using Cronbach's alpha coefficient and alpha equal to or greater than 0.70 was considered satisfactory [17]. Validity was assessed using known groups comparison to test how well the questionnaire discriminates between subgroups of the study sample that differed in gender and age. It was expected that women and old people would have lower scores than men and young people in all measures. In addition convergent validity (to test scaling assumptions) was assessed using the correlation of each item with its hypothesized scale. The Pearson product moment statistic (Pearson correlation coefficient) of 0.40 or above was considered satisfactory [18]. Furthermore the factor structure of the questionnaire was extracted by performing principal component analysis using oblique factor solution.

Results

The study sample

In all 4804 healthy individual were approached and 4163 (87%) were agreed to be interviewed. The mean age of the respondents was 35.1 (SD = 16.0), 52% were female, mostly married (58%), and the mean year of their formal education

Table 1. The characteristics of the study sample (n = 4163)

	No. (%)
<i>Age group (year)</i>	
15–24	1420 (34)
25–44	1614 (39)
45–64	882 (21)
≥65	247 (6)
Mean (SD)	35.1 (16.0)
<i>Gender</i>	
Male	1997 (48)
Female	2166 (52)
<i>Marital status</i>	
Single	1601 (38)
Married	2406 (58)
Widowed/divorced	156 (4)
<i>Educational level</i>	
Primary	755 (18)
Secondary	2595 (62)
Higher	812 (20)
Mean year (SD)	10.0 (4.5)
<i>Employment status</i>	
Employed	1482 (35)
Housewife	1225 (30)
Student	801 (19)
Unemployed	407 (10)
Retired	248 (6)

was 10.0 (SD = 4.5). The characteristics of the respondents are shown in Table 1. Of those who did not participate in the study 230 individuals were female and the remaining 411 were male. The main reason for this was due to the fact that after two approaches most of these individuals were not available in their home. Only a few individuals refused to respond to the questionnaire.

Descriptive statistics and reliability

The descriptive statistics for the eight SF-36 scales for the whole population are shown in Table 2. In addition the Cronbach's α coefficient to indicate item internal consistency reliability for each measure is presented and all measures but vitality showed satisfactory results. All scales met or exceeded the 0.70 level recommended for group comparison with the exception of the vitality scale (Cronbach's $\alpha = 0.65$). In addition the percentage of respondents scoring at the highest level (i.e., ceiling effect) was substantial for scales measuring physical functioning, role physical, bodily pain, social functioning, and role emotional. In contrast the percentage of respondents scoring at the lowest level (i.e., floor effect) was minimal for all scales but role emotional. Inter-scale correlation analysis also indicated that the scale constructs for the SF-36 Iranian version were generally distinct although fairly strong relationships between scales (correlations greater than 0.50) were observed in the expected directions. The results are shown in Table 3.

Known groups comparison

This was to test the scale validity. It was hypothesized that women and older people would have poorer health status than men and the younger respondents. The analysis showed that the females and the older respondents significantly had lower scores in all measures as expected (Tables 4 and 5). This indicated that the SF-36 well discriminated between sub-groups of people who differed in gender and age.

Table 2. Descriptive statistics and reliability statistics for the SF-36 scales

	Mean (SD)	Cronbach's α	Floor (%)	Ceiling (%)
Physical functioning (PF)	85.3 (20.8)	0.90	0.6	38.6
Role physical (RP)	70.0 (38.0)	0.85	0.15	53.5
Bodily pain (BP)	79.4 (25.1)	0.83	1.1	42.7
General health (GH)	67.5 (20.4)	0.71	0.3	4.7
Vitality (VT)	65.8 (17.3)	0.65	0.2	1.8
Social functioning (SF)	76.0 (24.4)	0.77	1.2	32.4
Role emotional (RE)	65.6 (41.4)	0.84	22.5	52.9
Mental health (MH)	67.0 (18.0)	0.77	0.1	1.9

Table 3. Inter-scale correlation for the SF-36 scales

	PF	RP	BP	GH	VT	SF	RE	MH
Physical functioning (PF)								
Role physical (RP)	0.56							
Bodily pain (BP)	0.53	0.55						
General health (GH)	0.46	0.45	0.54					
Vitality (VT)	0.46	0.44	0.52	0.55				
Social functioning (SF)	0.47	0.51	0.61	0.51	0.56			
Role emotional (RE)	0.34	0.52	0.41	0.39	0.41	0.51		
Mental health (MH)	0.33	0.34	0.42	0.051	0.65	0.56	0.46	

Table 4. Comparison of the SF-36 scores for the general population by gender

	Male mean (SD)	Female mean (SD)	<i>p</i> *
Physical functioning (PF)	87.8 (19.0)	82.9 (22.1)	<0.0001
Role physical (RP)	73.8 (36.4)	66.5 (39.1)	<0.0001
Bodily pain (BP)	82.7 (23.4)	76.4 (26.2)	<0.0001
General health (GH)	70.2 (19.6)	65.0 (20.8)	<0.0001
Vitality (VT)	68.9 (16.2)	62.9 (17.8)	<0.0001
Social functioning (SF)	78.0 (23.5)	74.2 (25.1)	<0.0001
Role emotional (RE)	70.1 (39.7)	61.4 (42.4)	<0.0001
Mental health (MH)	69.2 (17.1)	65.0 (18.6)	<0.0001

*The *t*-test result.

Table 5. Comparison of the SF-36 scores for the general population by age groups

	15–24 Mean (SD)	25–44 Mean (SD)	45–64 Mean (SD)	≥65 Mean (SD)	<i>p</i> *
Physical functioning (PF)	93.8 (12.9)	87.9 (17.2)	75.1 (22.4)	54.9 (29.2)	<0.0001
Role physical (RP)	81.0 (30.8)	73.2 (30.3)	56.4 (40.6)	34.2 (40.9)	<0.0001
Bodily pain (BP)	87.2 (19.7)	80.6 (23.8)	71.0 (27.3)	56.9 (30.2)	<0.0001
General health (GH)	75.5 (17.6)	68.3 (18.9)	58.6 (20.0)	49.0 (21.9)	<0.0001
Vitality (VT)	70.2 (15.2)	66.2 (16.8)	61.5 (17.7)	53.0 (20.3)	<0.0001
Social functioning (SF)	82.9 (20.6)	76.5 (23.7)	69.0 (24.9)	58.4 (30.4)	<0.0001
Role emotional (RE)	73.4 (37.3)	65.0 (41.8)	59.1 (43.3)	47.6 (44.2)	<0.0001
Mental health (MH)	71.1 (16.3)	65.9 (18.2)	63.7 (18.7)	62.2 (18.6)	<0.0001

*The result of one-way analysis of variance.

Test of scaling assumptions (convergent validity)

Table 6 presents the item-scale correlation matrix between each item and the eight SF-36 scales. All of the correlations between each item and its hypothesized scale showed satisfactory results suggesting that the items had a substantial association with the scale representing the concept. Each of the eight scales measured a distinct domain of functioning and well being as demon-

strated by higher item-scale correlations. Pearson correlation coefficient exceeded the 0.40 level recommended ranging from 0.58 (GH3) to 0.95 (BP1).

Factor structure

The principal component analysis with oblique rotation solution was performed and as expected a two-factor structure (physical and mental compo-

Table 6. Item-scale correlation matrix for the eight SF-36 measures

Item	PF	RP	BP	GH	VT	SF	RE	MH
<i>Physical functioning (PF)</i>								
PF1	0.65	0.41	0.39	0.41	0.38	0.34	0.25	0.25
PF2	0.74	0.45	0.41	0.41	0.36	0.36	0.28	0.26
PF3	0.71	0.41	0.40	0.35	0.35	0.31	0.24	0.24
PF4	0.80	0.47	0.44	0.36	0.38	0.38	0.28	0.27
PF5	0.77	0.39	0.37	0.27	0.31	0.33	0.24	0.22
PF6	0.75	0.43	0.44	0.34	0.37	0.35	0.29	0.27
PF7	0.81	0.45	0.43	0.38	0.36	0.39	0.28	0.26
PF8	0.82	0.44	0.40	0.32	0.32	0.37	0.26	0.23
PF9	0.77	0.37	0.33	0.26	0.27	0.32	0.21	0.19
PF10	0.59	0.27	0.25	0.20	0.23	0.24	0.19	0.18
<i>Role physical (RP)</i>								
RP1	0.46	0.81	0.45	0.38	0.38	0.40	0.42	0.27
RP2	0.44	0.83	0.44	0.38	0.38	0.41	0.45	0.31
RP3	0.50	0.83	0.47	0.37	0.36	0.44	0.42	0.29
RP4	0.46	0.84	0.46	0.37	0.36	0.43	0.43	0.27
<i>Bodily pain (BP)</i>								
BP1	0.49	0.50	0.95	0.52	0.48	0.56	0.37	0.39
BP2	0.51	0.53	0.91	0.49	0.50	0.60	0.40	0.40
<i>General health (GH)</i>								
GH1	0.47	0.45	0.50	0.71	0.45	0.44	0.35	0.39
GH2	0.25	0.29	0.36	0.69	0.39	0.33	0.26	0.35
GH3	0.22	0.18	0.23	0.58	0.25	0.23	0.16	0.23
GH4	0.24	0.24	0.28	0.66	0.27	0.29	0.24	0.34
GH5	0.41	0.39	0.49	0.78	0.51	0.45	0.33	0.43
<i>Vitality (VT)</i>								
VT1	0.29	0.29	0.32	0.31	0.64	0.33	0.24	0.29
VT2	0.36	0.34	0.39	0.42	0.77	0.41	0.31	0.52
VT3	0.33	0.32	0.41	0.44	0.69	0.42	0.32	0.52
VT4	0.29	0.27	0.33	0.35	0.67	0.38	0.27	0.47
<i>Social functioning (SF)</i>								
SF1	0.43	0.46	0.59	0.45	0.49	0.91	0.47	0.49
SF2	0.41	0.47	0.52	0.47	0.52	0.90	0.45	0.53
<i>Role emotional (RE)</i>								
RE1	0.28	0.44	0.36	0.35	0.35	0.43	0.87	0.40
RE2	0.27	0.44	0.34	0.32	0.35	0.43	0.87	0.39
RE3	0.34	0.47	0.37	0.36	0.37	0.47	0.88	0.39
<i>Mental health (MH)</i>								
MH1	0.16	0.19	0.26	0.32	0.38	0.36	0.28	0.69
MH2	0.22	0.25	0.30	0.41	0.43	0.41	0.36	0.74
MH3	0.24	0.26	0.32	0.36	0.49	0.39	0.30	0.69
MH4	0.24	0.25	0.30	0.35	0.49	0.43	0.38	0.77
MH5	0.31	0.29	0.35	0.39	0.56	0.42	0.32	0.72

nents) was loaded that jointly accounted for 65.9% of the variance. However, this was not exactly similar to the structure seen for the original questionnaire. The results are shown in Table 7.

Discussion

Cross-cultural validation studies although very difficult to be carried out, their results might be

Table 7. Hypothesized association between the SF-36 scales and the actual factor loading obtained in this study

	Hypothesized association		Factorial analysis: Rotated principal component		
	Physical	Mental	Correlation with		Variance explained
			Physical	Mental	
Physical functioning (PF)	+	-	0.81	0.18	0.69
Role physical (RP)	+	-	0.83	0.22	0.73
Bodily pain (BP)	+	-	0.68	0.42	0.65
General health (GH)	*	*	0.45	0.60	0.56
Vitality (VT)	*	*	0.31	0.78	0.71
Social functioning (SF)	*	+	0.50	0.63	0.65
Role emotional (RE)	-	+	0.46	0.50	0.45
Mental health (MH)	-	+	0.10	0.90	0.83

+ Strong association ($r > 0.70$), *Moderate association ($0.30 < r < 0.70$), -Weak association ($r < 0.30$).

considered worthwhile. First, providing standard health measures make health status comparison between countries possible. Second, they provide validated instruments to monitor population health, to estimate burden of disease and to investigate outcomes in clinical practice and to evaluate treatment effects. This study provided evidence that the SF-36 is a valid measure of population health status and quality of life in Iran. As it was suggested the results indicated that the SF-36 version meet at least two of the prerequisites for cross-cultural use of health status questionnaires, namely, cultural appropriateness and comparability of content [19].

The SF-36 was basically designed to be a self-administered questionnaire but it can be completed through an interview in person, computerized administration, or by telephone [20]. To collect data this study used face-to-face interviews. Thus there were no missing data and the problem of illiterate individuals was not encountered. Furthermore the study used a relatively large sample of the general population. Therefore as it has been suggested [16] the result of this study could be considered as Iranian normative data for the SF-36 Health Survey and might be used as a basis for comparison with specific populations in the future studies. However, one might argue that a sample from the urban capital is not necessarily representative of the entire country. In general this is true but since Tehran has become a multicultural metropolitan area it has been suggested that a sample from the general

population in Tehran at least could be regarded as a representative sample of urban population in Iran.

In general all psychometric tests of the SF-36 Iranian version showed satisfactory results. Reliability of the questionnaire as measured by the Cronbach's α coefficient for all eight scales but the vitality scale exceeded the recommended level. The fact that the vitality scale performed differently from other scales might reflect the difficulty encountered in translating vitality items from English to Persian. Even translating the word 'vitality' from English to Persian was somewhat difficult. However in translating vitality items it was decided to use very similar vocabularies to indicate what we are exactly asking for. Known groups analysis also indicated that the SF-36 could discriminate well between sub-group of people that differed in gender and age. The findings showed that women and old people had poorer health as compared to men and the younger respondents. This suggests that in study of health status and using the SF-36 in Iran the contribution of age and gender to the findings should be considered. Studies using the SF-36 in other countries also reported similar results [21].

The multi-item analysis of the SF-36 Iranian version showed promising results. For all eight scales satisfactory results were observed indicating that each item strongly correlated with its hypothesized scale. In addition the result showed that items were stronger measures of their hypothesized constructs than of other constructs.

Items in the bodily pain (BP) and social functioning (SF) scales correlated most strongly with their own scale than with other scales.

The factor analysis of the SF-36 clearly indicated that the questionnaire includes two underlying factors namely physical and mental components, although the findings was not exactly similar to the original hypothesized association. The vitality scale (VT) was strongly correlated with mental component than physical component. Also the role emotional scale (RE-role limitations due to emotional problems) showed a moderate association with physical and mental components while one expects a weak association with physical component and strong association with mental component. It is argued that these might be due to the cultural differences that exist between Asia and Western countries rather than structural defect since the criteria of validity of items and scales were satisfactory [8, 11, 12]. Perhaps this indicates the strength of physical and emotional dimensions of quality of life for Iranians and the fact that general health loads more strongly on mental health than physical health. However, one should be aware that in the present study the factor analysis was carried using oblique rotation method while the instrument developers used the varimax method [22]. In this study a two-factor structure (physical and mental components) jointly accounted for 65.9% of the variance that was a relatively better factor loading reported from Asian countries such as China (56.3%), Taiwan (60%) and Lebanon (62.9%) [10–12].

In conclusion, although the present study does not provide evidence on test-retest reliability or on responsiveness, the findings however provide further evidence that the translation of the SF-36 is feasible in Asia and could be used as a reliable and valid instrument for measuring health related quality of life. The Iranian version of the SF-36 is being used in several academic research projects and further results from its validity are awaited.

References

1. Bullinger M. German translation and psychometric testing of the SF-36 Health Survey: Preliminary results from the IQOLA Project. *Soc Sci Med* 1995; 41: 1359–1366.
2. Razavi D, Gandek B. Testing Dutch and French translation of the SF-36 Health Survey among Belgian angina patients. *J Clin Epidemiol* 1998; 51: 975–981.
3. Bjorner JB, Thunedborg K, Kristensen TS, Modvig J, Bech P. The Danish SF-36 Health Survey: Translation and preliminary validity studies. *J Clin Epidemiol* 1998; 51: 991–999.
4. Leplege A, Ecosse E, Verdier A, Perneger TV. The French SF-36 Health survey: Translation, cultural adaptation and preliminary psychometric evaluation. *J Clin Epidemiol* 1998; 51: 1013–1023.
5. Aaronson NK, Muller M, Cohen PDA, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998; 51: 1055–1068.
6. Wagner AK, Wyss K, Gandek B, Kilima PM, Lorenz S, Whiting D. A Kiswahili version of the SF-36 Health Survey for use in Tanzania: Translation and test of scaling assumptions. *Qual Life Res* 1999; 8: 101–110.
7. Taft C, Karlsson J, Sullivan M. Performance of the Swedish SF-36 version 2.0. *Qual Life Res* 2004; 13: 251–256.
8. Fukuhara S, Bito S, Green J, Hsiao A, Kurokawa K. Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. *J Clin Epidemiol* 1998; 51: 1037–1044.
9. Fuh JL, Wang SJ, Lu SR, et al. Psychometric evaluation of a Chinese (Taiwanese) version of the SF-36 Health Survey amongst middle-aged women from a rural community. *Qual Life Res* 2000; 9: 675–683.
10. Li L, Wang HM, Shen Y. Chinese SF-36 Health Survey: Translation, cultural adaptation, validation, and normalization. *J Epidemiol Community Health* 2003; 57: 259–263.
11. Sabbah I, Drouby N, Sabbah S, Retel-Rude N, Mercier M. Quality of life in rural and urban populations in Lebanon using SF-36 Health Survey. *Health and Quality of Life Outcomes* 2003; 1: 30. [<http://www.hqol.com/content/1/1/30>].
12. Tseng H, Lu JR, Gandek B. Cultural issues in using the SF-36 Health Survey in Asia: Results from Taiwan. *Health and Quality of Life Outcomes* 2003; 1: 72. [<http://www.hqol.com/content/1/1/72>].
13. Ware JE, Sherbourne CD. The MOS 36-Item Health Survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30: 473–483.
14. Bullinger M, Alonso J, Apolone G, et al. Translating health status questionnaires and evaluating their quality: The IQOLA Project approach. *J Clin Epidemiol* 1998; 51: 913–923.
15. Ware JE, Gandek B. Methods for testing data quality, scaling assumptions and reliability: The IQOLA Project approach. *J Clin Epidemiol* 1998; 51: 945–952.
16. Gandek B, Ware JE. Methods for validating and norming translations of health status questionnaires: The IQOLA Project approach. *J Clin Epidemiol* 1998; 51: 953–959.
17. Nunnally JC, Bernstein IR. *Psychometric Theory*, 3rd edn. New York: McGraw-Hill; 1994.
18. Ware JE, Brook RH, Davies-Avery A, et al. *Model of Health and Methodology*. Santa Monica, CA: RAND Corporation; 1980: R-1987/1-HEW.

19. Wanger AK, Gandek B, Aarenson NK, et al. Cross-cultural comparison of the content of SF-36 translations across 10 countries: Results from the IQOLA Project. *J Clin Epidemiol* 1998; 51: 925–932.
20. Ware JE, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *J Clin Epidemiol* 1998; 51: 903–912.
21. Hopman WM, Towheed T, Anastassiades T, et al. Canadian normative data for the SF-36 health survey. *CMAJ* 2000; 163: 265–271.
22. 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 construct. *Med Care* 1993; 31: 247–263.

Address for correspondence: Ali Montazeri, Iranian Institute for Health Sciences Research, P.O. Box 13185-1488, Tehran, Iran
E-mail: ali@jdcord.jd.ac.ir