

Validity and reliability of the Persian translation of the SF-36 version 2 questionnaire

N. Motamed,¹ A.R. Ayatollahi,² N. Zare² and A. Sadeghi-Hassanabadi³

صلاحية ومُعَوَّلِيَّة الترجمة الفارسية للإخراجة الثانية من الاستبيان إس إف-36
نيلوفر معتمد، سيد علي رضا آيت الله، نجف زارع، علي صادقي حسن أبادي

الخلاصة: بغية تقييم صلاحية ومُعَوَّلِيَّة الترجمة الفارسية للإخراجة الثانية من الاستبيان SF-36، كأداة تستخدم لقياس الوعي والإدراك الصحي العام لدى عامة السكان ومرضى العيادات الخارجية، قام الباحثون بإجراء دراسة على عينة مستعرضة تمثل جميع فئات المجتمع وذلك على العاملين في كلية الطب، بجامعة شيراز في شهر تشرين الثاني/نوفمبر من عام 2001. وقد كان معدل الاستجابة 83.6٪، وبلغت نسبة الانسجام فيما بين سلالمة للقياس تستخدم معامل كرونباخ ألفا: 0.87 (الحد الأدنى المطلوب 0.85). كما كانت صلاحية درجة الانسجام في ما بين سلالمة القياس ضمن المعدل المقبول، بالنظر إلى أن الترابط في ما بين سلالمة القياس وبعض الخصائص الاجتماعية الديمغرافية قد بلغ مبلغاً يُعْتَدُّ به إحصائياً. كذلك، كان معامل الترابط في ما بين سلالمة القياس الثمانية والمكونات الأساسية ذات العلاقة، ضمن المعدل المقبول أيضاً. وعلى ذلك، فإن هذه الترجمة تعتبر أداة مناسبة وملائمة لتقييم الإدراك والوعي الصحي لدى عامة السكان.

ABSTRACT To assess the validity and reliability of a Persian translation of the SF-36 questionnaire (version 2) as a tool for use in the general population and outpatients, we carried out a cross-sectional study of the staff of Shiraz Medical School during November 2001. Response rate was 83.6%. Internal consistency for 8 scales using Cronbach's α was 0.87 (minimum recommended 0.85). Construct validity was acceptable as correlation between scales and some sociodemographic characteristics was significant. The correlation coefficient between 8 scales and related principal components was also acceptable. This translation is an appropriate tool for assessing health perceptions of the population.

Validité et fiabilité de la traduction en langue perse du questionnaire SF-36 version 2

RÉSUMÉ Afin d'évaluer la validité et la fiabilité d'une traduction en langue perse du questionnaire SF-36 version 2 en tant qu'instrument pour utilisation dans la population générale et chez les patients des consultations externes, nous avons réalisé une étude transversale auprès du personnel de l'École de médecine de Shiraz durant le mois de novembre 2001. Le taux de réponse était de 83,6 %. L'homogénéité interne des 8 échelles en utilisant l' α de Cronbach était de 0,87 (minimum recommandé de 0,85). La validité du construit était acceptable, la corrélation entre les échelles et certaines caractéristiques sociodémographiques étant significatives. Le coefficient de corrélation entre les 8 échelles et les principales composantes du construit était également acceptable. Cette traduction est un outil approprié pour l'évaluation des perceptions de la population à l'égard de la santé.

¹Department of Community Medicine, Sari Medical School, Sari, Islamic Republic of Iran (Correspondence to N. Motamed: nilo1351@yahoo.com).

²Department of Biostatistics, Shiraz School of Public Health, Shiraz, Islamic Republic of Iran.

³Department of Community Medicine, Shiraz Medical School, Shiraz, Islamic Republic of Iran.

Received: 24/03/03; accepted: 18/02/04

Introduction

It is important to be able to measure the health perceptions of the population to assess the benefit of health care interventions and to target services. Researchers have developed measures to assess the health of people with specific diseases or disabilities, but these are of limited application when studying people with more than 1 condition or comparing perceived health across different groups [1].

The work of a health professional includes determining effective treatment regimens to help patients live better and longer lives. Perhaps more important, it also involves giving patients the tools to make informed decisions to improve their own quality of life. What is required is a comprehensive, and at the same time sensitive, measure of the full range of illness [1]. The SF-36 is specifically designed for use in the general population and among ambulatory patients. This questionnaire is sensitive to differences in health status among patients

who seek care in various outpatient settings [2]. The original general health measure was lengthy, containing 108 items. In an attempt to develop a general health survey that was comprehensive and psychometrically sound, yet short enough to be practical for use in large-scale studies of patients in practice settings, the authors of the SF-36 experimented with several shortened versions. Version 2 is the substantially revised form. Population reference values for this instrument are available from the United States of America and England and it is also being adapted into languages other than English by the International Quality of Life Assessment project [2].

The SF-36 is a self-administered questionnaire containing 36 items. It takes about 5 minutes to complete. It measures health on 8 multi-item scales, covering functional status, well-being and overall evaluation of health (Table 1) [1].

The measurement model underlying the construction of the SF-36 scales and sum-

Table 1 Dimensions of the SF-36 health survey questionnaire, adapted from Brazier et al. and Ware et al. [1,3]

Summary measure	Scale	No. of questions
Physical functioning	Physical functioning	10
	Role limitations (physical)	4
	Pain	2
	General health ^a	5
	Health change ^b	(1)
Mental functioning	Social functioning ^a	2
	Role limitations (emotional)	3
	Mental health	5
	Vitality ^a	4
Total		36

^aHas strong correlation with both components.

^bThis item was not included in the 8 scales, nor was it scored.

المجلة الصحية لشرق المتوسط، منظمة الصحة العالمية، المجلد الحادي عشر، العدد ٣، ٢٠٠٥

mary measures has 3 levels: items; 8 scales that aggregate 2–10 items each; and 2 summary measures, the physical component summary (PCS) and the mental component summary (MCS) measures, that aggregate scales (using factor analysis) (Table 1). All but 1 of the 36 items (self-reported health change) are used to score the 8 SF-36 scales. Each item is used in scoring only 1 scale. Three scales, physical functioning, role limitations (physical problems) and bodily pain, correlate most highly with the physical component and contribute most to the scoring of the PCS measure. The mental component correlates most highly with the mental health, role limitations (emotional problems) and social functioning scales, which also contribute most to the scoring of the MCS measure. Three of the scales (vitality, general health and social functioning) have significant correlation with both components [3].

It has been claimed that items in the SF-36 questionnaire detect positive as well as negative states of health [1]. In 6 of the 8 scales, patients are asked to rate their responses on 3-point or 6-point gradings rather than simply responding yes or no. For each scale, item scores are coded, computed, and transformed on to a scale from 0 (worst health) to 100 (best health).

Internal consistency is the extent to which items within a scale are correlated with each other. This can be examined by several methods but the most widely used is Cronbach's α , based on correlation between items [1].

Validity of a health measure is conceptually difficult to prove without a standard. Types of validity include face validity, where the relevance of the measurement appear obvious to the investigator and the questions supply the information the investigator needs; content validity, if the variable to be measured is a composite one

and all the component elements of the variable (as conceived) are measured; and construct validity, the extent to which a particular measure relates to measures consistent with theoretically derived hypotheses concerning the concepts, or constructs, that are being measured. One way of checking construct validity is to see how well the measure discriminates between groups that there is reason to believe should differ in the characteristics under study [5]. For example women, older people, and people in semi-skilled and unskilled occupations might be expected to perceive their health as being relatively poor; people making use of health services might also be expected to perceive their health as poorer than non-users [1].

While information is rapidly becoming available on the properties of the American and British versions in ambulatory patients, less is known regarding its performance in healthy people, and even less about the performance of adaptations into other languages. Our purpose was to examine the validity and reliability of the Persian language adaptation of SF-36 in the staff of Shiraz Medical School as a particularly healthy group.

Methods

In our study, we used the Persian language version of the SF-36 on staff of Shiraz Medical School. The wording of 4 items was altered slightly due to cultural differences (gardening and ping-pong instead of bowling or playing golf, 1 kilometre instead of 1 mile, and several hundred metres instead of several hundred yards). Translated questionnaires, together with questions on sociodemographic characteristics, including age, sex, marital status, years of marriage, number of children, occupation, years of work, education and salary, were

distributed to all 134 members of staff working in the wards of Shiraz Medical School in November 2001 and collected a few days later.

Crude scores of the 8 scales (0 to 100) were standardized based on mean = 50 and standard deviation (SD) = 10 and mean and SD for the general population [4].

$$\begin{aligned} [(x-\mu_x)/\delta_x] &= [(y_s-50)/10] \\ \Rightarrow y_s &= 50 + 10[(x-\mu_x)/\delta_x] \end{aligned}$$

where:

$\mu_s = 50$ = general population mean

$\delta_s = 10$ = general population SD

y_s = standard score of each person

x = crude score (0–100) of each person

μ_x = norm-based score of each item (for example physical functioning)

δ_x = norm based SD of each item (for example physical functioning).

Thus, scores of different scales are not crude scores of 0–100. This allows us to compare scores across the various scales and summary measures [4]. The statistical package used was *SPSS*, version 10.

The responses to the questionnaire were subjected to Chronbach's α to test reliability. To assess face validity and content validity, after translation of the questionnaire into Persian by an experienced translator, we discussed it with health specialists who were fluent in Persian and English and then incorporated the changes in 2 steps (for example changes in wording) to develop the final questionnaire.

We used the Student *t*-test and the Kruskal–Wallis 1-way analysis of variance to test whether the SF-36 scores differed significantly between groups. In addition we used linear regression and Spearman's rank correlation to assess the association between variables and different scales. We expected that the correlation between each scale and its related summary measure

would be greater than its correlation with the other summary measure. The level of the variables were:

- age (years): < 25; 25–34; 35–44; 45–55; > 55
- sex: female; male
- marital status: single; married; divorced
- years of marriage: < 10; 10–19; \geq 20
- number of children: \leq 3; > 3
- education: did not complete high school; high-school diploma; university degree
- occupation: high; middle; low
- years of work: < 10; \geq 10.

Results

We received completed questionnaires from 112 of 134 individuals, giving a response rate of 83.6%. On 74 questionnaires (66.1%) all 36 items were complete, on 20 (17.9%) at least 1 item was missing and on 18 (16.1%) at least 2 items were missing. The most frequent item that was missing was, "During the past 4 weeks, have you been very nervous?" Only 1 paper had so much data missing that it could not be processed. Therefore, in all, data from 111 questionnaires were analysed.

The age of the participants ranged from 19 years to 57 years, mean 36.08 years (SD 9.57). Fifty per cent were female. Marital status was: single 34 (30.6%), married 76 (68.5%) and divorced 1 (0.9%). The monthly salary was: 78 (70.3%) earned less than 800 000 rials (US \$1 = 1755 rials, November 2001) and 33 (29.7%) earned 800 000–2 000 000 rials. Education was: 62 (55.9%) of the participants had high-school diploma or less and 49 (44.1%) had university education. Family size was 0–3 children for 88.7% of the participants.

We found that internal consistency using Cronbach's α at the level of 8 scales was acceptable; it exceeded the recommended minimum of 0.85 as it was 0.87 in 103 cases [1]. For the 4 scales of the PCS measure (general health, bodily pain, role limitations, physical functioning) Cronbach's α was 0.79 in 106 cases and for the 4 scales of the MCS measure (mental health, role limitations, social functioning, vitality) was 0.79 in 105 cases. Table 2 shows Cronbach's α according to sex, occupation and education level. For mental health, Cronbach's α was higher for women than for men, but for physical health it was higher for men. The comparatively low Cronbach's α scores seen here may in part be related to the small number of scales (4 scales).

Concerning construct validity, and considering the different levels of variables

(Table 3), we found there was a significant linear correlation between physical functioning, PCS measure and age (Spearman's rank correlation coefficient 0.27 and 0.18 respectively) ($P < 0.05$). Physical functioning and the PCS measure decreased with increasing age. Even after the introduction of sex into the age equation, the relationship between physical functioning and the PCS measure was significant. There was also a significant linear correlation between sex and role limitations for physical problems and role limitations for emotional problems (Spearman's rank correlation coefficient 0.21 and 0.23 respectively). Women had significantly higher scores than men in role limitations for physical problems and in role limitations for emotional problems ($P < 0.005$). After adjusting for sex according to age, this was still true.

Table 2 Cronbach's α for the different scales according to sex, occupation and education level

Characteristic	4 scales, mental health		4 scales, physical health		8 scales	
	Cronbach's α	No.	Cronbach's α	No.	Cronbach's α	No.
Sex						
Female	0.82	52	0.77	53	0.86	51
Male	0.79	53	0.81	53	0.88	52
Occupation^a						
Low	0.76	18	0.70	18	0.85	18
Middle	0.86	25	0.82	25	0.90	23
High	0.77	61	0.81	62	0.86	61
Education						
Did not complete high school	0.79	19	0.67	19	0.85	19
High-school diploma	0.83	38	0.84	37	0.91	36
University	0.74	47	0.78	49	0.81	47

^aOccupation levels: high = permanent staff; medium = temporary staff on yearly contract; low = hired (and paid) by the day.

Table 3 Correlation coefficient of the 8 scales and 2 summary measures: comparison of values obtained in Iran (present study), France [2] and the United States of America [3]

Scales	Correlation coefficient					
	Iran		France		USA	
	PCS	MCS	PCS	MCS	PCS	MCS
Physical functioning	0.71	0.09	0.74	-0.06	0.85	0.12
Role limitations, physical problems	0.67	0.22	0.63	0.33	0.81	0.27
Bodily pain	0.82	0.20	0.72	0.27	0.76	0.28
General health	0.60	0.41	0.63	0.40	0.69	0.37
Vitality	0.42	0.60	0.41	0.75	0.47	0.65
Social functioning	0.29	0.72	0.36	0.73	0.42	0.67
Role limitations, emotional problems	0.27	0.58	0.01	0.82	0.16	0.78
Mental health	-0.002	0.85	0.2	0.84	0.17	0.87

PCS = physical component summary measure.

MCS = mental component summary measure.

There was also a significant linear relationship between years of marriage and physical functioning and the PCS measure (Spearman's rank correlation coefficient 0.25 and 0.26 respectively), as the longer the participant had been married, the lower the score for physical functioning and the PCS measure, although after the introduction of age that was not significant. Kruskal-Wallis 1-way analysis of variance revealed a significant relationship between marital status and physical functioning and the PCS measure ($P < 0.05$) as married individuals had lower scores in these.

Individuals with a maximum of 3 children had significantly higher scores in social functioning and general health ($P < 0.05$), although Spearman's rank correlation coefficient for number of children and physical functioning was not significant ($P = 0.16$).

The higher the occupational level, the lower the role limitations (physical prob-

lems) score (Spearman's rank correlation coefficient 0.23; $P = 0.017$).

The only scale that was different for the 2 income levels was social functioning. In other words, in the lower income group mean score was significantly lower ($P < 0.05$). Also, people who had worked for a maximum of 10 years had significantly lower scores in physical functioning ($P < 0.05$): the more years a participant had worked, the lower the score of physical functioning (Spearman's rank correlation coefficient 0.21; $P = 0.026$).

There was a significant positive relationship between education level and the PCS measure and role limitations (physical problems) (Spearman's rank correlation coefficient 0.22 ($P = 0.026$) and 0.20 ($P = 0.036$) respectively).

Table 3 shows correlation coefficients between 8 scales and 2 summary measures for our study (Islamic Republic of Iran)

along with data from France and the United States of America [2,3] for comparison. Of the 8 scales, physical functioning and bodily pain had the strongest correlation with perception of physical health, similar to values recorded for the United States of America and France. Furthermore, in all 3 studies mental health had the strongest correlation with the MCS measure.

Discussion

We achieved a response rate of 83.6%, suggesting that the SF-36 questionnaire is an acceptable tool for measuring health perception in the general population. Either the present Persian translation of the questionnaire is successful or the properties of the original instrument are robust enough to withstand our translation into Persian. Non-response (attrition) rates could perhaps be improved by a better translation. Our findings support the developers' claims of internal consistency [3] for the SF-36.

With rare exceptions, published reliability statistics have exceeded the minimum standard of 0.7 recommended for measures using in-group comparisons; most have exceeded 0.8 [3]. In our study, Cronbach's α for each summary measure was 0.79. In addition, Cronbach's α for the 8 scales was 0.87.

While studies of subgroups indicated slight declines in reliability for the more-disadvantaged respondents, reliability coefficients consistently exceeded recommended standards for group level analysis [3]. In our study, Cronbach's α exceeded 0.85 for the 8 scales (except university education) and 0.70 for each of the 4 mental and 4 physical scales (except for those who did not complete high school in physical health) (Table 2). It is noteworthy that in

both occupation and education, people in the middle groups had the highest Cronbach's α . Cronbach's α for the PCS measure was higher in men and for the MCS measure it was higher in women. This may be the result of differing concerns men and women have regarding their physical and mental health.

The evidence for construct validity of the SF-36 was substantial. The distribution of scores observed by sociodemographic characteristic showed that the questionnaire is more valid for physical than for mental health, although we did not expect validity for mental health to be significantly different between the sexes or with increasing age (borne out by our findings). McCabe et al. found a significant difference in the mean scores for men and women, but did not find any correlation between age and scores [6].

Although in the study by Brazier et al., men perceived themselves to be significantly healthier than women, except on the general health scales [1], in our study, only in 2 scales (role limitations physical problems and role limitations emotional problems) were men and women different, with the women's score being higher. Brazier et al. found a significant correlation between age and physical functioning and bodily pain [1]. In our study we also found a correlation between age and physical functioning and the PCS measure (decreasing significantly with increasing age, even after controlling for sex). In addition, on examining sociodemographic characteristics such as education, occupation, number of children, income and marital status, we saw considerable differences in scores. For example, as in the study of Regidor et al., where perceived health status declined with decreasing educational level [7], in our study, scores for role limitations (physical

problems) and for the PCS measure increased with increasing level of education.

Another way to assess the construct validity is to assess correlation between individual scales and principal components because if we find a meaningful correlation, we can be sure that the summarization of those scales to components is correct. In the American and French studies, physical functioning correlated most strongly with the PCS measure and mental health with the MCS measure (Table 4). The only difference in our study was that bodily pain was more strongly correlated with the PCS measure than physical functioning. There is no obvious explanation for this at present. Also, as in the other 2 studies, vitality and general health correlated moderately with both principal components.

In attempting to be comprehensive, existing general health questionnaires may be too long or require interviews, or both. In

primary care or community settings, contact time with the patient is often short and thus, to be practical and acceptable to the population, the questionnaire should be brief, easy to use and preferably self-administered. These features are also important for researchers. The SF-36 questionnaire seems to meet these criteria, even when used in a particularly healthy population and after translation into Persian.

In summary, we believe that this version of SF-36 is suitable for its intended purpose, i.e. to measure health in the general population, although some items that are related to mental health require more revision. We consider that with further work to normalize this instrument, and after conducting research with a larger sample size, we could use the Persian language version of the SF-36 questionnaire as an inexpensive and rapid tool for the population of the Islamic Republic of Iran.

Table 4 Spearman's linear regression coefficient (r_s) between scales and characteristics of participants

Characteristic	Scale	r_s	P-value
Age	Physical functioning	-0.27	0.004
	Physical component summary measure	-0.18	0.064
Sex	Role limitations, physical problems	0.21	0.024
	Role limitations, emotional problems	0.23	0.016
Years of marriage	Physical functioning	-0.25	0.011
	Physical component summary measure	-0.26	0.010
No. of children	Physical functioning	-0.15	0.160
Occupation level	Role limitations, physical problems	-0.23	0.017
Marital status	Physical functioning	-0.21	0.023
Years of work	Physical functioning	-0.21	0.026
Education level	Physical component summary measure	0.22	0.026
	Role limitations, physical problems	0.2	0.036

المجلة الصحية لشرق المتوسط، منظمة الصحة العالمية، المجلد الحادي عشر، العدد ٣، ٢٠٠٥

References

1. Brazier JE et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *British medical journal*, 1992, 305(6846):160–4.
2. Perneger TV et al. Validation of a French-language version of the MOS 36-Item Short Form Health Survey (SF-36) in young healthy adults. *Journal of clinical epidemiology*, 1995, 48(8):1051–60.
3. Ware JE Jr, Gandek B. Overview of the SF-36 health survey and the International Quality of Life Assessment (IQOLA) project. *Journal of clinical epidemiology*, 1998, 51(11):903–12.
4. Ware JE Jr. *SF-36® health survey update*. Quality Metric Incorporated (<http://www.sf-36.org/tools/sf36.shtml>, accessed 29 December 2004).
5. Abramson JH, Abramson ZH. *Survey methods in community medicine*, 5th ed. Philadelphia, Churchill Livingstone, 1999:171–204.
6. McCabe CJ et al. Measuring the mental health status of a population: a comparison of the GHQ-12 and the SF-36 (MHI-5). *British journal of psychiatry*, 1996, 169(4):517–21.
7. Regidor E et al. Association between educational level and health related quality of life in Spanish adults. *Journal of epidemiology and community health*, 1999, 53(2):75–82.

Third Regional Conference of Medical Journals in the Eastern Mediterranean Region

The Third Regional Conference of Medical Journals in the Eastern Mediterranean Region will be held in Shiraz, Islamic Republic of Iran from 26 to 29 January 2006. More information and the Conference website will be announced soon.