A reassessment of factor structure of the Short Form Health Survey (SF-36): A comparative approach

Vida Alizad (1)
Manouchehr Azkhosh (2)
Ali Asgari (3)
Karyn Gonano (4)

(1) Iranian Research Center on Ageing, University of Social Welfare and Rehabilitation Sciences, Koodakyar St., Daneshjou Blvd, Evin, Tehran, Iran.
(2) Counseling Department, The University of Social Welfare and Rehabilitation Sciences, Koodakyar St., Daneshjou Blvd, Evin, Tehran, Iran.
(3) Department of Educational Psychology, the University of Kharazmi, Tehran, Iran.
(4) Academic Language and Learning Services, Queensland University of Technology, Floor 3, X Block, Gardens Point Campus, Brisbane QLD 4000, Australia

Correspondence:
Manouchehr Azkhosh
Counseling Department, The University of Social Welfare and Rehabilitation Sciences, Koodakyar St., Daneshjou Blvd, Evin,
Tehran, Iran.
Email: azkhosh34@yahoo.com

Abstract

Background: The factor structure of the Short Form Health Survey (SF-36) and its application to older people in Eastern countries has been the focus of limited research. Four theoretical and experimental factor structures of the SF-36 were tested and compared here to establish a best-fitting model for Iranian older people.

Methods: A sample of 391 participants (60 -89 years) years completed the Farsi SF-36. A confirmatory factor analysis assessed the fit and viability of the measurement model. Three theoretical and experimental factor structures of the SF-36 were tested using an exploratory principal component analysis to explore the factor solution of the Farsi SF-36.

Results: An exploratory factor analysis identified the two factor solutions (mental and physical) to be the same as the original US model, but the fit indices of the confirmatory factor analysis identified the two and three factor model (mental, physical and well-being) to be the same, making the latter more extensive for use with older people.

Conclusion: This study provides strong evidence that the Farsi SF-36 has the potential to measure well-being status of older people. Such an application is valid if the Vitality items are modified and new items are developed for the Well-being scale.

Key words: Health–related Quality of life, SF-36 Health Survey, older people

Please cite this article as: Alizad V. A reassessment of factor structure of the Short Form Health Survey (SF-36): A comparative approach. *World Family Medicine.* 2017; 15(8):45-51. DOI 10.5742/MEWFM.2017.93088

Introduction

Measuring the health-related quality of life (HRQoL) of older people has a prominent position in gerontology as an indicator for monitoring the health status of older people. Such monitoring forms the basis for clinical decision-making and gerontological research outcome measures. Of the several instruments for measuring HRQoL, the SF-36 Health Survey is the most widely used in health research (1, 2) and is known for its high standard of reliability and validity (3, 4). The SF-36 accentuates both the practical and popular nature of the questionnaire in both clinical settings and research. It has been translated into more than 20 languages (5). The SF-36 enables policy makers to involve older people in the decision making process about their own health with a comprehensive and short instrument (6). The SF-36 has been validated for use with older people, and its applicability and suitability are well documented (5, 7, 8), however, those versions cannot be used with Iranian older people due to lack of cultural equivalence. The existing Farsi SF-36 was translated and validated for use with a general population by an Iranian research team in 2005 (9), however, it cannot be used directly with older people owing to their heterogeneous characteristics (6). There is also controversy surrounding the numbers of underlying dimensions measured by the different translated versions of the SF-36 compared to the US original SF-36 (10-12). Therefore, a lack of a validated Farsi SF-36 for elderly and controversy about the number of underlying dimensions measured by the SF-36 highlights the need to conduct a new psychometric analysis. This study therefore investigates the factor structure of the Farsi version of the SF-36 in older people to find the best-fitting model for this population group.

Materials and Methods

Participants: A sample of 391 participants (197 males and 194 females) were randomly selected from the Tehran population and they ranged in age from 60-89 years. The inclusion criteria were age of 60 years and older and Abbreviated Mental test score ≥ 6. The participants were asked to complete the existing Farsi SF-36 (9), which took about 20 to 30 minutes. The research was approved by the ethics committee of the University of Social Welfare and Rehabilitation Sciences (USWR.REC.7393.162). Written informed consent was obtained from each participant.

Instrument: The SF-36 Health Survey assesses the mental and physical health status and eight generic health concepts including Physical Functioning (PF); Role Limitations due to Physical Health (RP); Bodily Pain (BP); General Health (GH); Vitality (VI); Social Functioning (SF); Role limitations due to Emotional Health (RE); and Mental Health (MH). The SF-36 has been translated for use in several countries as part of the International Quality of Life Assessment (IQOLA) project (13, 14), and has demonstrated reliability and validity across diverse samples (3). The scale has 36 items that are scored and summed according to a standardized protocol and expressed as a score on a 0-100 scale for each of the eight health concepts, with higher scores representing a better health status (15). The psychometric testing of the Farsi version of the SF-36 followed the procedure of the IQOLA project (9).

Data analysis: A confirmatory factor analysis (CFA) on item level using LISREL 8.4 (16) assessed the fit and viability of the measurement model which was developed from the original US model (Model A) (15). The chi-square is significant at $p < .001$ and an adequate fit is < 2.0. The point estimate of the Root Mean Square Error of Approximation (RMSEA) and its upper confidence limit for the model should be less than 0.05 (17). The Expected Cross-Validation Index (ECVI) (6.08) should be less than the ECVI for the saturated model (3.23) (16). An exploratory principal component analysis (PCA) was conducted to explore the factor solution of the Farsi SF-36 with both orthogonal and oblique rotations. In this analysis, three alternative models were examined to explore the best fitting model. These alternative models were a one-factor model (model B), a three uncorrelated second order factor model (model C) based on previous studies (18), and an eight-factor model (model D). Figure 1 depicts the diagrams of these three models and the original US model (Model A).

Results

Descriptive statistics

The test of normality of the scale scores showed the distributions of all study variables were negatively skewed (Table 1 - page 48). The α-coefficient for the VI was very low, and for SF and MH were also below typically accepted standards. On the other hand, Cronbach's alphas were adequate for the GH, RE, Physical and Mental components, and good for PF, RP and BP.

Testing the Sf-36 factor structure models

Four CFA models were developed to confirm the factor structure of the Farsi SF-36. This analysis served to confirm Model A, the original US model, which is a comparison of the three competing models to ascertain the extent to which the Model A would demonstrate a superior fit to the three alternative models. Model B, the first alternative model was developed to load all items of the SF-36 into a single health construct. The second alternative model, Model C, was developed from the three summary measures of mental, physical and well-being (18). The third alternative model, Model D was then developed based on the eight factors that aggregate the 36 items of the SF-36.

Table 2 indicates how Model B provided a poor fit for this data. While a significantly greater model fit was observed for the original US model (Model A) and Model C and D; when compared to Model B, these models did not demonstrate a good fit. Models A, C and D however, provided a relatively better fit for this data. An examination of model fit statistics revealed mixed evidence for a good model fit, where all models did not fit the data well, according to the significant chi-square index, the relative chi-square per degrees of freedom and the RMSEA (17). The Comparative goodness-of-fit and Incremental Fit Index for these models indicated more than an acceptable model fit, but the goodness-of-fit

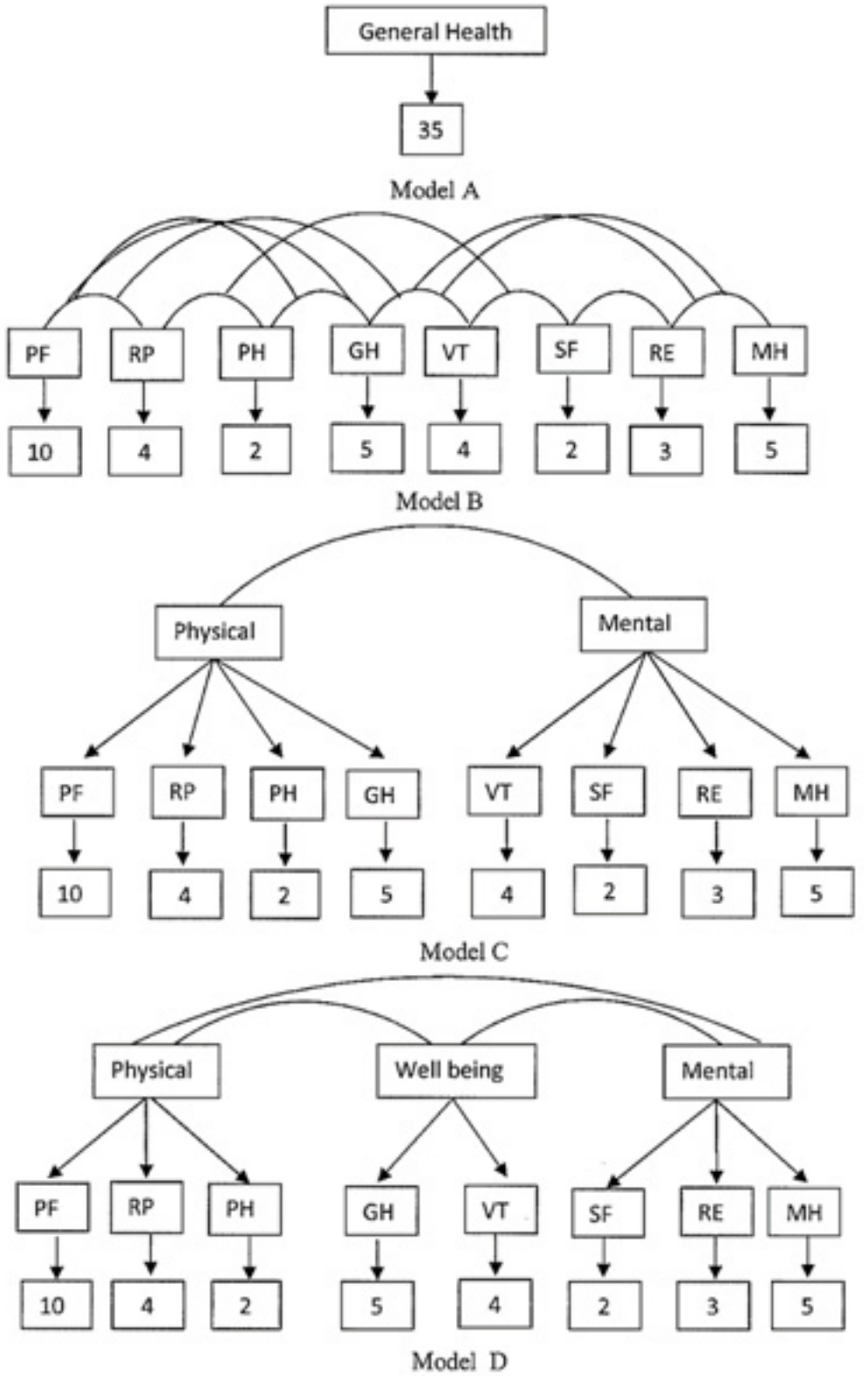


Figure 1. Path diagrams of models A, B and C of the Farsi version of the SF-36.
Note: Digits in the squares indicate the number of items (indicators). Model A, included the items belonging to one general factor (General Health), Model B included items placed in the eight factors (first order CFA), Model C included a hierarchy CFA model based on the original taxonomy model.

Table 1: Descriptive statistics and K-S test of normality of Persian SF-36 scales and summary measures

Subscales	M	SD	Skewness	Kurtosis	D	α
PF	60.3	31.6	-0.48	-1.00	2.5**	0.93
RP	62.1	39.5	-0.49	-1.30	4.9**	0.83
BP	65.3	30.0	-0.32	-1.07	3.7**	0.83
GH	58.9	21.7	-0.29	-0.39	1.28	0.74
VI	59.1	22.7	-0.13	-0.43	1.27	0.49
SF	76.4	28.4	-0.84	-0.05	4.4**	0.69
RE	63.9	39.5	-0.58	-1.22	5.4**	0.76
MH	66.4	21.0	-0.47	-0.20	1.5*	0.66
Physical	61.3	21.7	-0.28	-0.74	1.3	0.74
Mental	64.8	19.7	-0.47	-0.15	1.4*	0.72

PF= Physical Functioning; RP=Role-Physical; BP= Bodily Pain; GH= General Health; VT= Vitality; SF= Social Functioning; RE = Role-Emotion; MH= Mental Health. D = Kolmogorov-Smirnov (K-S) test of normality. α = Cronbach's alpha. ** P<.01. * P<.05

Table 2: Factor loadings, communalities, mean and standard deviation of the scales of the Farsi SF-36, a two factor solution

Scales	Factors				Descriptive statistics		
	Varimax		Oblimin		M	SD	h^2
	I	II	I	II			
PF	.818		.774		60.3	31.6	0.67
BP	.700		.768		65.4	30.1	0.51
GH	.666		.755		58.9	21.7	0.59
RP	.563	.433	.663		62.0	39.4	0.61
RE		.807		.796	63.9	39.5	0.66
SF		.751		.774	76.5	28.4	0.60
MH		.587		.668	66.4	21.0	0.49
VT	.441	.478		.581	63.9	22.7	0.42
Eigenvalue	3.59	.945					
% Variance	44.8	11.8					

Note. PF= Physical Functioning; BP= Bodily Pain; GH= General Health; RP=Role-Physical; RE = Role-Emotion; SF= Social Functioning, MH= Mental Health, VT= Vitality. h2= communality. Factor loading > 0.4. Cross-loaded items > 0.

Table 3: The goodness of fit statistics for CFA modified and non-modified two and three summary measure models of the Farsi SF-36

Models			χ2	df	χ2 / df	CFI	GFI	IFI	NFI	ECVI	RMSEA
First order		Model A	1258.09**	551	2.33	0.91	0.75	0.91	0.89	3.63	0.057
		Model B	5057.67**	560	9.03	0.80	0.45	0.80	0.78	6.08	0.144
		Model C	1256.99**	548	2.29	0.91	0.75	0.91	0.89	3.64	0.058
		Model D	1189.72**	531	2.24	0.92	0.76	0.92	0.90	3.56	0.056
Higher order	Non-modified	Model E	99.2 **	19	5.2	0.95	0.94	0.94	0.94	0.34	0.104
		Model F	103.5**	17	6.08	0.95	0.94	0.95	0.94	0.36	0.114
	modified	Model E	22.3	15	1.48	0.99	0.98	0.99	0.99	0.16	0.035
		Model F	24.7*	13	1.90	0.99	0.98	0.99	0.97	0.18	0.048

Note: Model E included the two summary measures belonging to one general factor (General Health), Model F included the three summary measures belonging to one general factor. CFI= Comparative Fit Index, GFI=goodness of fit index, IFI= Incremental Fit Index, NFI= Non Normed Fit Index, RMSEA=Root Mean Square Error of Approximation, ECVI= Expected Cross-Validation Index. ECVI for Saturated Model=0.18. * P< .05. ** P< .001

index for all models was lower than the accepted criteria. Together, these results indicate that the one-factor did not provide a good fit for this data, and the three alternative models are the same, according to the fit indices, along with the fact that their overall fit did not appear across the model fit statistics.

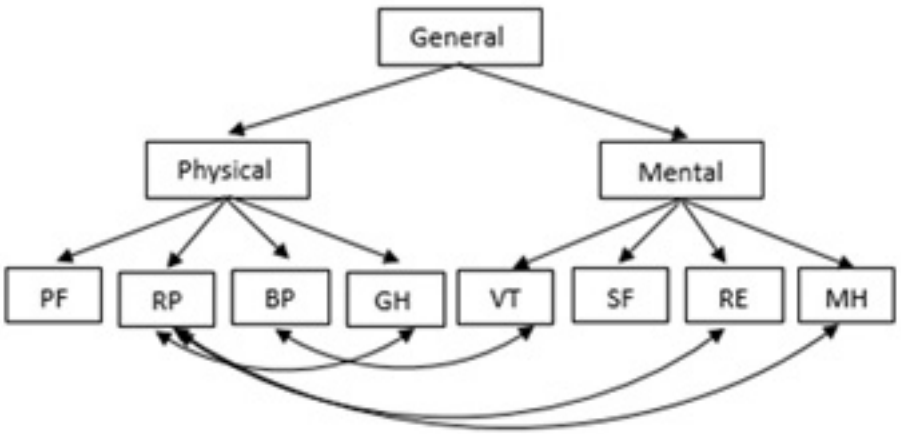
Farsi SF-36 factor structure

A PCA of the eight scales was conducted with orthogonal and oblique rotations to explore the factor structure of the Farsi SF-36, and examined how many summary measures were extracted in the Iranian sample. The analyses demonstrated only one factor could be extracted with eigenvalues over 1, and explained a total 44.9% of the observed variance. The original two measure model and the three uncorrelated measures were examined and a scree-test was used to determine the proper number of several factor solutions. The results showed that the first two factors represented the main sources of variance in the data matrix. However, the results of orthogonal rotation showed the RP and VT had high factor loadings on both factors. The results of the oblimin rotation however, showed a better fit of the original measurement model because all eight scales were loaded on their appropriate factors. The explained variance by the two extracted factors was 56.7%.

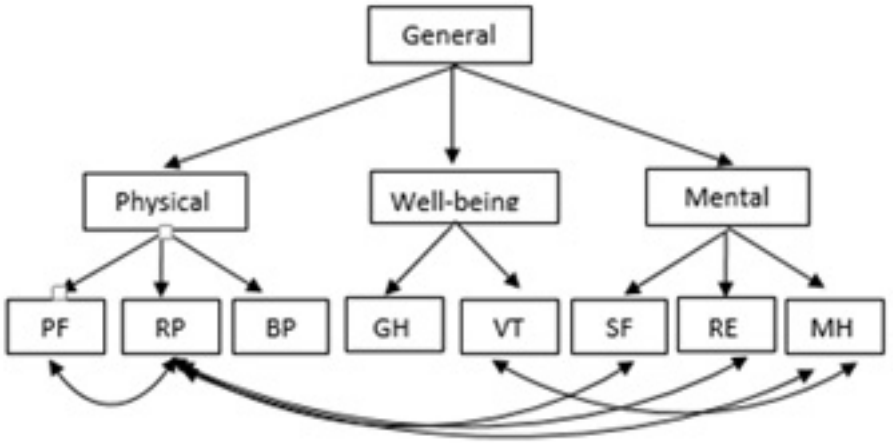
The communality, eigenvalues and factor loadings for the two rotated factors using varimax and oblimin rotations are presented in Table 2, along with the means, and standard deviations for all eight scales.

One General Health Construct

A second order CFA based on the two factor model (Physical and Mental) and three factor model of adding General well-being, were loaded on a one general health conception. This analysis served to explore and determine the underlying latent trait in the SF-36 that aggregates the summary measures (Figure 2).



Model E



Model F

Figure 2. Path diagrams of second order models E and F of the Farsi SF-36. Model E, included the two summary measures belonging to one general factor (General Health), Model F included the three summary measures belonging to one general factor.

Table 3 shows how both models demonstrated an unacceptable fit of this data. The chi squares were significant, with $p < 0.05$, the relative chi square was higher than 2, and the RMSEA was higher than 0.05. However, the goodness-of fit indices were acceptable and confirmed both models.

One strategy to improve the models was to add the correlations between error terms which were specified, when suggested, by high modification indices and also where theoretically defensible. In Model E, the correlated errors were added between RP items with GH, RE and MH, and between BP-VT. The correlated errors in model F were added between RP with PF, SF, RE and MH, and between MH-VT.

An examination of model fit statistics after modification revealed adequate evidence for a good model fit. However, in both Models the RP scale had the most error correlations with the other scales, but its pattern of correlated errors was different in the models with GH, RE and MH.

Discussion

A comparative approach of the four factor structure models of the Farsi SF-36 is presented here.

Reliability

The Farsi SF-36 has shown satisfactory internal consistency reliability (>0.70) for all scales except VI, SF and MH, with the highest value for PF. Similar results for PF have been reported in other countries (19), which was to be expected given that the PF scale has 10 out of 36 items of the SF-36. Such a large set of items increases the Cronbach's α. A very low level of reliability has been reported for the VI in other studies conducted with older people and patients with chronic conditions (10, 20, 21). These studies determined whether the lower internal consistency of the SF-36 VI scale is due to the study sample or cultural differences, although this deserves further study. Additionally, the SF scale results for internal consistency reliability were below typically accepted standards in Iran, suggesting a decreased level of social abilities among older adults. The fact that this result was consistent with other studies (20, 22, 23) raises the question of whether only two items in the SF scale are adequate for assessing the concept of social functioning.

Factor structure

Both two and three factor models were confirmed in the Iranian older population, however a comparison of fit indices of the higher order two-factor model, Model A with two summary measures, and higher order three factor model, Model C with three summary measures, showed no differences between these two models. This result means it is not possible to show a preferred model for older people and suggests the instrument is conceptually equivalent with the original version. However, the question remains about how many scales could be extracted from this concept.

The two factor model (mental and physical) showed PF, RP, BP and GH to correlate with physical health component, and VI, SF, RE and MH with the mental dimension. These results are consistent with other studies (24) and confirmed the Farsi SF-36 met the psychometric standards hypothesized in the original model for physical and mental health.

The results of testing a three second-order factor are consistent with the study conducted across nine countries (18) and Rasch validation of the SF-36 in Korea (25). The third factor, interpreted as general well-being in this data, was the result of clustering GH and VI, as in previous studies, and interpreted for the mixed factor content of GH and VI (18). The reason for such a divergent result between this and previous studies that confirmed the two-factor model may be the difference in separating elderly from the other population group. Another reason for this divergent result is that cultural value plays a role in the interpretation of these differences, and Iranian elderly tend to put more value on items related to well-being compared with other population groups and cultures. Therefore the three-factor model makes the Farsi SF-36 particularly suitable for use in the assessment of older adults, as its three scales make it more appropriate to identify older people's needs. Such data makes it possible to develop a more precise care plan, since the more factors indicate a greater identification of the underlying latent trait. The three factor model will be achieved by slightly modifying items on the VI scale and developing items for the well-being scale.

Conclusion

The Farsi SF-36 has generally accepted psychometric properties, with empirical evidence showing that developing items for the third factor of well-being would be useful to better identify the needs of older people.

Acknowledgement

The authors of this paper acknowledge their gratitude to Dr Emma Caukill and all of the participants in this research project.

References

1. Myint PK, Smith RD, Luben RN, Surtees PG, Wainwright NW, Wareham NJ, et al. Lifestyle behaviours and quality-adjusted life years in middle and older age. Age and ageing. 2011;40(5):589-95.
2. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ. 2002;324(7351):1417.
3. McHorney CA, Ware JE, Jr., 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. Medical care. 1993;31(3):247-63.

4. McHorney CA, Ware JE, Jr., Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Medical care. 1994;32(1):40-66.
5. Mishra GD, Gale CR, Sayer AA, Cooper C, Dennison EM, Whalley LJ, et al. How useful are the SF-36 sub-scales in older people? Mokken scaling of data from the HALCyon programme. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2011;20(7):1005-10.
6. Hazard WR. principles of geriatric medicine and gerontology1999. 1387-403 p.
7. Bohannon RW, DePasquale L. Physical Functioning Scale of the Short-Form (SF) 36: Internal Consistency and Validity With Older Adults. Journal of Geriatric Physical Therapy. 2010;33(1):16-8.
8. Hill R, Mansour E, Valentijn S, Jolles J, van Boxtel M. The SF-36 as a precursory measure of adaptive functioning in normal aging: the Maastricht Aging Study. Aging Clin Exp Res. 2010;22(5-6):433-9.
9. Montazeri A, Goshtasebi A, Vahdaninia M, Gandek B. The Short Form Health Survey (SF-36): translation and validation study of the Iranian version. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2005;14(3):875-82.
10. Meng H, King-Kallimanis B, Gum A, Wamsley B. Measurement bias of the SF-36 Health Survey in older adults with chronic conditions. Quality of Life Research. 2013;22(9):2359-69.
11. Bartsch L, Butterworth P, Byles J, Mitchell P, Shaw J, Anstey K. Examining the SF-36 in an older population: analysis of data and presentation of Australian adult reference scores from the Dynamic Analyses to Optimise Ageing project. Quality of Life Research. 2011;20(8):1227-36.
12. Ngo-Metzger Q, Sorkin DH, Mangione CM, Gandek B, Hays RD. Evaluating the SF-36 Health Survey in Older Vietnamese Americans. Journal of Aging and Health. 2008.
13. Ware JE, Jr., Keller SD, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health status questionnaires. Methods from the IQOLA project. International Quality of Life Assessment. International journal of technology assessment in health care. 1995;11(3):525-51.
14. Ware JE, Jr., Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment Project. Journal of clinical epidemiology. 1998;51(11):903-12.
15. Ware JE KM, Keller SD. . SF-36 Physical and Mental Component Summary Measures—A User's Manual.: Boston: The Health Institute; 1994.
16. Joreskog KG SD. Scientific Software 2003. Available from: http://www.ssicentral.com/.
17. Hu L-t, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999;6(1):1-55.
18. Keller SD, Ware JE, Jr., Bentler PM, Aaronson NK, Alonso J, Apolone G, et al. Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: results from the IQOLA Project. International Quality of Life Assessment. Journal of clinical epidemiology. 1998;51(11):1179-88.
19. Ware JE, Jr., Gandek B, Kosinski M, Aaronson NK, Apolone G, Brazier J, et al. The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. Journal of clinical epidemiology. 1998;51(11):1167-70.
20. ten Klooster PM, Vonkeman HE, Taal E, Siemons L, Hendriks L, de Jong AJ, et al. Performance of the Dutch SF-36 version as a measure of health-related quality of life in patients with rheumatoid arthritis. Health and quality of life outcomes. 2013;11:77.
21. Hoopman R, Terwee CB, Deville W, Knol DL, Aaronson NK. Evaluation of the psychometric properties of the SF-36 health survey for use among Turkish and Moroccan ethnic minority populations in the Netherlands. Quality of life research : international journal of quality of life aspects of treatment, care and rehabilitation. 2009;18(6):753-64.
22. Pappa E, Kontodimopoulos N, Niakas D. Validating and norming of the Greek SF-36 Health Survey. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2005;14(5):1433-8.
23. Tseng HM, Lu JF, 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.
24. Ware JE, Jr., Kosinski M, Gandek B, Aaronson NK, Apolone G, Bech P, et al. The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. Journal of clinical epidemiology. 1998;51(11):1159-65.
25. Kim SH, So WY. Rasch validation of the SF-36 for assessing the health status of Korean older adults. Journal of physical therapy science. 2015;27(3):601-6.