ORIGINAL ARTICLE

Validation of the Turkish Form of Short Form-12 health survey version 2 (SF-12v2)

Özgül Soysal Gündüz[1] , Senan Mutlu[2] , Arzu Aslan Baslı[2] , Cihan Gül[3] ,
Özgür Akgül[4] , Emel Yılmaz[2] , Ömer Aydemir[3]

[1]Department of Internal Medicine, Division of Rheumatology, Manisa Celal Bayar University, Faculty of Medicine, Manisa, Turkey
[2]Department of Surgical Nursing, Manisa Celal Bayar University, Faculty of Health Science, Manisa, Turkey
[3]Department of Psychiatry, Manisa Celal Bayar University, Faculty of Medicine, Manisa, Turkey
[4]Department of Physical Therapy and Rehabilitation, Division of Rheumatology, Manisa Celal Bayar University,
Faculty of Medicine, Manisa, Turkey

ABSTRACT

Objectives: In this present study, we aimed to perform the reliability and validity analyses of the Turkish Form of Short Form-12 version 2 (SF-12v2) in different groups of patients.

Patients and methods: After the permission for the validation study of the Turkish Form, Optum provided the authors the validly translated Turkish Form of SF-12v2. The study was carried out in rheumatological, psychiatric, and surgical wards of Manisa Celal Bayar University Hafsa Sultan Training and Research Center between September 2019 and June 2020. Taking possible dropouts into consideration, a total of 136 patients (67 males, 69 females; mean age: 43.5±14.4 years; range 19 to 82 years) constituted the study group. In addition to SF-12v2, for concurrent validity, Nottingham Health Profile (NHP) was used as the comparator instrument.

Results: In internal consistency, for the physical component summary score, the Cronbach alpha coefficient was 0.80, and item-total score correlation coefficients were between 0.32 and 0.73. The Cronbach alpha coefficient of the mental component summary score was found to be 0.88 where the item-total correlation coefficients varied between 0.60 and 0.78. Exploratory factor analysis revealed a two-factor solution, representing mental and physical components. For criterion validity, convergent and discriminant validity analyses were performed using NHP with SF-12v2, and domains of SF-12v2 correlated well with the domains of NHP accordingly. In criterion validity, the psychiatric group had the lowest mean score in mental health, vitality, social functioning and role difficulties due to emotional problems, whereas the surgical group had the lowest mean score in bodily pain, role difficulties due to physical problems, and physical functioning.

Conclusion: Our study results show that the Turkish form of SF-12v2 is valid and reliable both in clinical practice and clinical trials.

Keywords: Nottingham Health Profile, quality of life, Short Form-12, validation.

Beginning in the 1960s, the health status of a patient was considered beyond his/her illness and objective outcome criteria such as morbidity, relapse rate, mortality rate.[1] In the new understanding of health status, there were subjective outcome criteria such as functionality, quality of life, life satisfaction, and well-being.

Clinicians began to implement these constructs first in their researches and, then, in their routine clinical practice. Meanwhile, many instruments were developed and provided in this field.[2] First, for research purposes generic and specific instruments of health-related quality of life (HRQOL) were introduced. These instruments

were relatively extensive requiring a long time of administration.[3] Since 2000s, with the rise of personalized medicine and HRQOL measures, patient-reported outcome measures have been implemented in the routine daily practice for each individual patient. Therefore, easy-to-use instruments requiring relatively short amount of time have been developed.[4]

With the Medical Outcome Studies and Health Status Surveys, in the early 1990s, a generic instrument Short Form 36 (SF-36) was developed[5] and adapted into many languages, as well as into Turkish.[6] For practical purposes, SF-12, a shorter version of SF-36, was developed. In the early 2000s, new versions of both SF-36 and SF-12 were developed.[7,8] This new version of SF-12v2 is a 12-item, self-rated, HRQOL instrument and it was validated into many languages as well as for many diseases. It has eight domains represented with one or two questionnaire items; these domains are physical functioning, role participation with physical health problems (role-physical), bodily pain, general health, vitality, social functioning, role participation with emotional health problems (role-emotional), and mental health.

Although SF-12v2 was validly translated into Turkish, the psychometric analyses and validation were not performed. In this present study, we aimed to perform the reliability and validity analyses of the Turkish Form of SF-12v2 in different groups of patients with rheumatological diseases, mental disorders, and surgical conditions.

PATIENTS AND METHODS

This study was conducted at Manisa Celal Bayar University Hafsa Sultan Training and Research Center between September 2019 and June 2020. One of the authors contacted with the Optum holding the copyrights of all of the health status surveys including SF-12v2. After the permission for the validation study of the Turkish Form, Optum provided the authors the validly translated Turkish Form of SF-12v2. In order to cover a wide variety of patients, the study was carried out in the rheumatological ward, psychiatric ward, and in three surgical wards such as orthopedics and traumatology, general surgery, and cardiovascular surgery. It was aimed to collect 50 patients for each category of patients to fulfill

the criterion of having 10-fold subjects for each item of the instrument; that is, at least 120 subjects. Taking possible dropouts into consideration, 60 patients were invited to the study for each group. Inclusion criteria were being between the ages 18 and 65 years, conforming with study protocol and instructions of the instruments, and giving informed consent. Exclusion criteria were severe agitation, disorientation, cognitive impairment due to the underlying disease(s). As a result, 50 patients for the surgical group, 36 patients for the rheumatological group, and 50 patients for the psychiatric group with a total of 136 patients (67 males, 69 females; mean age: 43.5 ± 14.4 years; range 19 to 82 years) constituted the study group. A written informed consent was obtained from each patient. The study protocol was approved by the Manisa Celal Bayar University Ethics Committee for Medical Researches (No: 20.478.486/Date: 13.11.2019).

The SF-12v2 is the main instrument of this present study. It consists of 12 items where two of the items are three-point Likert type, the others are five-point Likert type. During the Medical Outcome Studies, eight comprehensive pools of items regarding the eight domains were evaluated for a brief questionnaire. To overcome the floor and ceiling effects, the response choices are improved noteworthy. In addition to the eight domains, the physical (PCS) and mental (MCS) component summary measures are also calculated. With the 2009 field trial data, the internal consistency coefficients for PCS of SF-12v2 was 0.92 and for MCS it was 0.88.[9] In the construct validity analysis, factor analysis revealed two-component solution indicating the component summary measures. In the convergent and discriminant analyses, the domain scores showed significant correlations with the related constructs. Also, with the domains of SF-36v2, the items of SF-12v2 showed well item-component summary measures correlations. The SF-12v2 domains discriminated well between the healthy subjects and disease groups. For concurrent validity, the SF-36v2 was used as the comparator instrument, and all the correlations were high. In other validation studies for different languages, the SF-12v2 was proven to be valid for a diverse range of cultures.

For concurrent validity, the Nottingham Health Profile (NHP) was used as the comparator

instrument. It was developed by and it was validated into Turkish by Küçükdeveci et al.[10] It contains five domains: pain, emotional reactions, sleep, social isolation, and physical activity. In addition, in the second part, the problems encountered by the individual due to the health status in certain situations are rated. To meet the discriminant and convergent validity analysis criteria, NHP was preferred. In some other studies, the SF-36v2 was used concurrently; however, since the items of SF-12v2 were derived directly from SF-36v2, it would be more item-total score analyses instead of convergent-discriminant analyses.

Statistical analysis

In the reliability analyses, Cronbach alpha coefficients for the domains, Pearson item-scale correlations, and intraclass correlation coefficients were calculated. In construct validity, exploratory factor analysis was performed as principal component analysis with varimax rotation. Factors with eigenvalue greater than 1 and items with factor loadings greater than 0.4 were taken into consideration. For convergent and discriminant validity, Pearson correlation coefficients were calculated for the domains of SF-12v2 and for the domains of NHP. Since the main point behind the convergent and discriminant validity is that it provides how similar domains correlate well with each other, while how distinct domains correlate poorly with each other. Therefore, the correlations of similar domains were hypothesized to be higher than the correlations of the distinct domains. This is an important component of criterion validity.

For known group analyses, the mean domain scores of the three disease groups, as well as sex, age-group scores were compared with analysis of variance (ANOVA) test. Before performing the analyses, normality of distribution was tested with the Levene test. Since the demographic data contained differences between the study groups which would challenge the ANOVA test, we performed one-way multivariate analysis of covariance (MANCOVA) test with Wilk's Lambda method to control the age and education effect on the quality of life parameters. For the multivariate analysis, the F value was found to be 1.12 (p=0.25) and it was not statistically significant. Therefore, ANOVA test was

Table 1. Demographic and clinical features of the study groups

	Surgical Group			Psychiatric Group			Rheumatological Group			
	n	%	Mean±SD	n	%	Mean±SD	n	%	Mean±SD	p
Age (year)			52.0±14.4			35.0±11.5			43.3±10.8	<0.0001
Sex										NS
Male	23	46.0		21	42.0		23	63.9		
Female	27	54.0		29	58.0		13	36.1		
Education (year)			7.1±3.6			12.2±3.0			7.7±3.6	<0.0001
Diagnosis										
Limb surgery	14									
Abdominal surgery	11									
Cancer surgery	11									
Cardiovascular surgery	14									
Bipolar disorder				5						
Major depression				22						
PTSD				1						
OCD				2						
Anxiety disorder				16						
Psychosis				4						
Ankylosing spondylitis							36			

SD: Standard deviation; PTSD: Post-traumatic stress disorder; OCD: Obsessive-compulsive disorder; NS: Not Significant.

Table 2. Exploratory Factor Analysis of SF-12v2

	Factor 1	Factor 2
SF 1	0.47	0.14
SF 2a	0.18	0.76
SF 2b	0.10	0.68
SF 3a	0.32	0.80
SF 3b	0.15	0.84
SF 4a	0.65	0.53
SF 4b	0.60	0.54
SF 5	0.18	0.59
SF 6a	0.83	0.02
SF 6b	0.82	0.11
SF 6c	0.64	0.32
SF 7	0.77	0.28
Eigenvalue	5.58	1.57
Total variance (%)	46.57	13.14

SF-12v2: Short Form-12 version 2.

performed to compare the three study groups in terms of SF-12 domains.

RESULTS

Demographic characteristics of the study groups are given in Table 1. There was no statistically significant difference between the study groups in terms of sex (Chi-square=4.34, p=0.11); however, in terms of the mean age, there was a statistically significant difference (T=23.37, p<0.0001) where the surgical group was older than the psychiatric group and, in terms of education, the psychiatric group was more educated than both the rheumatological and surgical groups (T=32.97, p<0.0001).

For reliability analysis, internal consistency was calculated. Since the domains of SF-12v2 consists of one or two items, it would be inconvenient to calculate internal consistency of every domain. Instead, as in the original development study of SF-12v2, internal consistency of every component summary score was computed. Thus, for the physical component summary score, the Cronbach alpha coefficient was 0.80, and item-total score correlation coefficients were between 0.32 and 0.73. Internal consistency of the mental component summary score was found to be 0.88 where the item-total correlation coefficients varied between 0.60 and 0.78.

In construct validity analysis, exploratory factor analysis was performed with the principal component analysis and varimax rotation. Kaiser-Meyer-Olkin Measure of Sampling Adequacy was found to be 0.83, and Bartlett's test of sphericity was statistically significant (Chi-square=938.30, p<0.0001). Eigenvalue greater than 1.0 and items with factor loading greater than 0.40 were taken into consideration. As a result, a two-factor solution was obtained; Factor 1 with eigenvalue of 5.58 representing 46.57% of the variance, and Factor 2 with

Table 3. Convergent and discriminant validity of SF-12v2 with Nottingham Health Profile using Pearson correlation analysis

NHP SF-12v2	Pain	Emotional reactions	Sleep	Social isolation	Physical mobility	Energy
General health	-0.19	-0.35	-0.17	-0.37	-0.03	-0.34
Physical functioning	-0.54	-0.18	-0.25	-0.10	-0.58	-0.45
Role-physical	-0.42	-0.27	-0.29	-0.22	-0.43	-0.44
Bodily pain	-0.56	-0.31	-0.38	-0.19	-0.62	-0.45
Social functioning	-0.24	-0.54	-0.49	-0.52	-0.21	-0.43
Vitality	-0.12	-0.45	-0.31	-0.41	-0.10	-0.42
Mental health	-0.18	-0.63	-0.49	-0.56	-0.12	-0.52
Role-emotional	-0.28	-0.58	-0.45	-0.46	-0.30	-0.56

SF-12v2: Short Form-12 version 2.

Table 4. Comparison of domain mean scores of the study groups

	Surgical Group	Psychiatric Group	Rheumatological Group	
	Mean±SD	Mean±SD	Mean±SD	p
General health	50.0±17.5	43.5±21.3	47.9±12.5	NS
Role-physical	52.8±25.3	52.8±31.3	76.0±27.0	<0.0001
Physical functioning	50.0±33.9	64.5±31.2	78.5±34.4	<0.0001
Role-emotional	60.8±26.4	38.5±30.4	84.7±23.2	<0.0001
Bodily pain	47.0±27.5	65.0±34.3	88.9±16.3	<0.0001
Mental health	80.5±37.7	44.5±19.7	88.6±37.0	<0.0001
Vitality	46.0±22.2	41.7±25.6	47.9±26.3	<0.0001
Social functioning	64.0±25.3	44.0±34.5	88.6±24.5	<0.0001

SD: Standard deviation.

eigenvalue of 1.57 representing 13.14% of the variance (Table 2). Factor 1 contained all the items belonging to the mental component summary score, whereas Factor 2 consisted of all the items under the physical component summary score. There was only one exception, and the item 1 concerning general health was represented under Factor 1, instead of Factor 2. Also, the items 4a and 4b (items of role difficulties due to the emotional problems) had factor loadings greater than 0.4 in both Factor 1 and 2. However, since the factor loadings in Factor 1 were greater than that in Factor 2, these two items were taken into consideration in Factor 1.

For criterion validity, convergent and discriminant validity analyses were performed using NHP with SF-12v2. Since six domains were concordant in both instruments, Pearson correlation analyses were done concerning these domains. Moreover, physical and emotional domains in general were taken into consideration, when correlation coefficients of SF-12v2 assessed in terms of convergent and discriminant standings (Table 3). Role difficulties due the physical problems and physical functioning were both highly correlated with domains of NHP such as pain, physical mobility, and energy. Bodily pain was highly correlated with pain and, then, with physical mobility and energy. Vitality was equally correlated with physical mobility, energy, and emotional reactions. Role

difficulties due to emotional problems, social functioning, and mental health were highly correlated with all domains of NHP, except for pain and physical mobility. General health was moderately correlated with energy, emotional reactions, and social isolation.

When the mean domain scores were compared among the three study groups (Table 4), the psychiatric group had the lowest mean score in mental health, vitality, social functioning, and role difficulties due to emotional problems, whereas the surgical group had the lowest mean score in bodily pain, role difficulties due to physical problems, and physical functioning. All three groups equally affected general health domain scores.

DISCUSSION

In this present study, the Turkish form of SF-12v2 which is the most widely used brief HRQOL instrument is validated in a diverse group of patients with including rheumatological diseases, psychiatric disorders, and surgical conditions.

In internal consistency analyses, both the physical and mental component summary domains had high coefficients. In the original study, it was found to be 0.92 and 0.88, respectively.[9] In a study in Bengali, it was

reported that the Cronbach coefficients were exceeded 0.90.[11] The Iranian version demonstrated similarly high Cronbach alpha coefficients, 0.87 for PCS and 0.88 for MCS.[12] In the Chinese study, relatively low coefficients (0.67 and 0.60) for Cronbach alpha analysis were found.[13]

Item-total score correlations were statistically significant higher than 0.20, contributing to the reliability of the Turkish form of SF-12v2.

In construct validity, exploratory factor analysis revealed a two-factor solution, as expected. It indicates the physical and mental components well. In the original study, the two-factor construct was demonstrated.[9] In the Bengali version, factor analysis revealed two factors, a physical and a mental component.[11] In the Iranian study, factor analysis revealed two factors a physical and a mental component.[12] In our study, the only exception is that general health was loaded under the MCS instead of PCS. General health item was supposed to rate the perceived level of health of the individual generally. However, since it is reviewing health in general from the individual's point of view, it easily rated as a well-being item[14] and it is converted to a psychological anchor. In this present study, the analyses indicated that general health item was rated in a well-being context.

In convergent and discriminant validity analyses, in this study, the NHP was used as the concurrent instrument, instead of SF-36. As stated in the User's Manual of SF-12vs, when SF-36 is used as the concurrent instrument to perform convergent and discriminant analyses, since the items of SF-12v2 are extracted from SF-36, the analysis becomes more of an item-total scale reliability analysis, instead of convergent and discriminant validity analysis.[8] The domains of NHP can be also considered in physical and psychological contexts. Thus, domains of SF-12v2 under the PCS were correlated well with the physical domains of NHP, whereas domains under the MCS were mostly correlated with psychological domains of NHP. There was only one exception, general health item, which was correlated moderately with PCS domains of SF-12v2. As discussed previously, it was accepted as a well-being construct. In other

validation studies, the SF-36 was the most preferred instrument to demonstrate convergent and discriminant validity, and in the original study[9] and in Chinese[13] and Singapore[15] studies, it is reported that items and domains of SF-12v2 showed well correlation with the related domains of SF-36.

In the criterion validity, the diagnostic groups were used as the distinction among the study groups. Thus, the psychiatric group where psychological disturbances are expected to be more impaired demonstrated relatively lower scores on MCS domains, whereas acute surgical group showed lower scores on PCS domains. As a result, SF-12v2 is able to discriminate areas of disturbances between different individuals.

Since the SF-36 is better known, it would be good to compare and provide the data of the performance of SF-12v2 and SF-36. In the population study with SF-12v2 and SF-36, the same domains of the two instruments showed correlations 0.80 to 0.96, showing that SF-12v2 has the ability to rate HRQOL measures as good as SF-36.[16]

The main strength of the study is the sample size large enough to perform all the psychometric analyses. Also, it contains diverse medical problems and it can represent the routine medical practice of the physicians. The patient group consisted of ordinary individuals without any difference in terms of sex, age, or education; that is, it represents daily routine medical practice. As a limitation, the study is cross-sectional in design which restrains to present the ability to change of SF-12v2.

In conclusion, our study results show that the Turkish form of SF-12v2 is valid and reliable both in clinical practice and clinical trials in Turkey.

Declaration of conflicting interests

Funding

REFERENCES

1. Post MW. Definitions of quality of life: what has happened and how to move on. Top Spinal Cord Inj Rehabil 2014;20:167-80.

2. Ware JE Jr, Brook RH, Davies AR, Lohr KN. Choosing measures of health status for individuals in general populations. Am J Public Health 1981;71:620-5.

3. Stewart AL, Ware JE, editors. Measuring functioning and well-being: The Medical Outcomes Study approach. 1st ed. Durham, NC: Duke University Press; 1992.

4. Halyard MY, Ferrans CE. Quality-of-Life assessment for routine oncology clinical practice. J Support Oncol 2008;6:221-9, 233.

5. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.

6. Koçyiğit H, Aydemir Ö, Fişek G, Ölmez N, Memiş AK. Form-36 (KF-36)'nın Türkçe versiyonunun güvenilirliği ve geçerliliği. İlaç ve tedavi dergisi 1999;12:102-6.

7. Maruish ME. User's manual for the SF-36v2 Health Survey. 3rd ed. Lincoln, RI: Quality-Metric Incorporated; 2011.

8. Maruish ME. User's manual for the SF-12v2 Health Survey. 3rd ed. Lincoln, RI: Quality-Metric Incorporated; 2012.

9. Ware JE, Kosinski M, Gandek B, Sundaram M, Bjorner JB. User's manual for the SF-12v2 Health Survey. 2nd ed. Lincoln, RI: Quality-Metric Incorporated; 2010.

10. Kücükdeveci AA, McKenna SP, Kutlay S, Gürsel Y, Whalley D, Arasil T. The development and psychometric assessment of the Turkish version of the Nottingham Health Profile. Int J Rehabil Res 2000;23:31-8.

11. Islam N, Khan IH, Ferdous N, Rasker JJ. Translation, cultural adaptation and validation of the English "Short form SF 12v2" into Bengali in rheumatoid arthritis patients. Health Qual Life Outcomes 2017;15:109.

12. Montazeri A, Vahdaninia M, Mousavi SJ, Asadi-Lari M, Omidvari S, Tavousi M. The 12-item medical outcomes study short form health survey version 2.0 (SF-12v2): a population-based validation study from Tehran, Iran. Health Qual Life Outcomes 2011;9:12.

13. Lam ET, Lam CL, Fong DY, Huang WW. Is the SF-12 version 2 Health Survey a valid and equivalent substitute for the SF-36 version 2 Health Survey for the Chinese? J Eval Clin Pract 2013;19:200-8.

14. Wettstein M, Eich W, Bieber C, Tesarz J. Pain Intensity, Disability, and Quality of Life in Patients with Chronic Low Back Pain: Does Age Matter? Pain Med 2019;20:464-75.

15. Tan ML, Wee HL, Salim A, Lee J, Ma S, Heng D, et al. Validity of a Revised Short Form-12 Health Survey Version 2 in Different Ethnic Populations. Ann Acad Med Singap 2016;45:228-36.

16. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220-33.