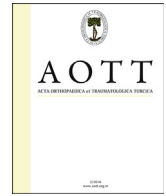




Contents lists available at ScienceDirect

## Acta Orthopaedica et Traumatologica Turcica

journal homepage: <https://www.elsevier.com/locate/aott>

## Short Form Health Survey version-2.0 Turkish (SF-36v2) is an efficient outcome parameter in musculoskeletal research

Derya Çelik <sup>a,\*</sup>, Özge Çoban <sup>b</sup><sup>a</sup> Istanbul University, Faculty of Health Science, Division of Physiotherapy and Rehabilitation, Istanbul, Turkey<sup>b</sup> Yildirim Beyazıt University, Faculty of Health Science, Division of Physiotherapy and Rehabilitation, Ankara, Turkey

## ARTICLE INFO

## Article history:

Received 25 July 2015

Received in revised form

10 October 2015

Accepted 22 January 2016

Available online 17 November 2016

## Keywords:

Musculoskeletal research

SF-36v2

Reliability

Validity

Psychometric properties

## ABSTRACT

**Objective:** Although the Short Form Health Survey version-2.0 (SF-36v2) is widely used since 2000, the researchers and clinicians in Turkey have been still using the original version. However, the original version includes many deficiencies and the SF-36v2 was introduced aiming to correct these deficiencies. The purpose of this study is to indicate differences between SF-36 and SF-36v2 and the present cross cultural adaptation, reliability and validity of the SF-36v2.

**Patients and methods:** The SF-36v2 was cross culturally adapted to Turkish and the measurement properties of the Turkish version of the SF-36v2 were tested in 50 patients (19 males; mean  $\pm$  SD age:  $36.9 \pm 14.6$  years; range: 16–65 years, BMI;  $24.1 \pm 4.6$ ) with a variety of musculoskeletal pathologies. Intra-class correlation coefficients (ICC) were used to estimate the test-retest reliability. Construct validity was analyzed with SF-36v2 and EuroQol Group (EQ-5D). The distribution of ceiling and floor effects was determined.

**Results:** During the cross-cultural adaptation process many changes were made. The Turkish SF-36v2 subscales showed excellent test-retest reliability which was ranged 0.80 to 0.95. The highest correlation was found between SF-36v2-PCS and SF-36v2-PF ( $r = 0.75$ ), the lowest correlation was found between SF-36v2-PCS and SF-36v2-MH ( $r = 0.05$ ). The correlations between EQ-5D and SF-36v2 subscales ranged from 0.10 (SF-36v2 –VT) to 0.46 (SF-36v2 –RE). We observed no ceiling and floor effects.

**Conclusion:** The cultural adaptation of the SF-36v2 was successful. The SF-36v2 has sufficient reliability and validity to measure a variety of musculoskeletal pathologies for Turkish-speaking individuals.

© 2016 Turkish Association of Orthopaedics and Traumatology. Publishing services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

The Short Form 36 (SF-36) is a health survey which includes 36 questions. It involves 8 different assessments of functional health and well-being evaluation and also psychometrically calculated measurements of mental and physical health. It is a general scale which is not specific for any kind of age, disease or treatment group.<sup>1</sup> It is used for more than 200 disease or conditions such as musculoskeletal conditions, neuromuscular conditions, osteoarthritis, psychiatric conditions, spinal injuries, trauma.<sup>2</sup> Translations of the SF-36 have become the topic of 500 publisher's and

researcher's studies from 22 different countries. More than 10 studies have been conducted in 13 different countries.<sup>1</sup>

Although the SF-36 is widely used, it has been reported some deficiencies. The SF-36 version 2 (SF-36v2) has been introduced to correct deficiency of the SF-36. In 1996, after careful quantitative and qualitative studies, it has been implemented by Ware et al.<sup>3,4</sup> Briefly the SF-36v2 involves some important changes. Instructions are improved, questions are shortened, more familiar, less obscure and simplified words are used, design of the form is improved to make it easier to complete, read faster, and reduce missing responses. Fourth and fifth questions that evaluate physical and emotional function, two level response orders of seven items transformed into five level response orders. Ninth question that evaluate mental health and vitality, six level response order shortened into five response in order to simplify the answer.<sup>3,4</sup>

Turkish version of the SF-36 published in 1999 has been widely used in researchers and clinicians.<sup>5</sup> However, the existing Turkish

\* Corresponding author.

E-mail address: [ptderya@hotmail.com](mailto:ptderya@hotmail.com) (D. Çelik).

Peer review under responsibility of Turkish Association of Orthopaedics and Traumatology.

studies that used the questionnaire are not entirely reliable because different linguistic versions have been applied in the studies. We think that is due to the missing of cross-cultural adaptation. However, before using such an outcome measurement in a community, they need to be translated and culturally adapted, given that the majority of these scores reflect the characteristics of the language and the social culture of the community in which they were established. For instance, even though “block” is a measurement unit in United States, it was translated to “sokak” in Turkish version which is not measurement unit in Turkey. In addition to cross cultural adaptation, the reliability and validity of the Turkish version of the SF-36 was only conducted patients with rheumatoid arthritis which is limited for patients with orthopedic pathologies.

There is no available cross culturally adapted SF-36v2 for Turkish population. The purpose of the study is to introduce differences between SF-36 and SF-36v2 and the cross cultural adaptation, reliability and validity of the SF-36v2.

## Patients and methods

Fifty patients with variety of musculoskeletal pathologies were recruited from the Istanbul University, Istanbul School of Medicine, Department of Orthopedics. The inclusion criteria were: (1) 16 years or older; (2) presence of any musculoskeletal problem; and (3) no treatment between the test-retest assessments. The exclusion criteria were: (1) inability to complete the forms as a result of cognitive impairment; (2) illiteracy or lack of understanding of Turkish; (3) the presence of neurologic disorders. Cross cultural adaptation was held by two physical therapists and an orthopedic surgeon. The patients were answered the SF-36v2 and EuroQol Group (EQ-5D) for construct validity and SF-36v2 twice for the test-retest reliability.

### Short Form Health Survey (SF-36) and Short Form Health Survey version-2.0 (SF-36v2)

The SF-36 is a multidimensional questionnaire that assesses eight different aspects of health. It is generic by nature which means that it, as opposed to disease-specific measures, can be used to measure and compare outcomes across different diseases and treatments. The SF-36 is a 36 item questionnaire that measures eight multi-item dimensions of health: physical functioning (10 items) social functioning (2 items) role limitations due to physical problems (4 items), role limitations due to emotional problems (3 items), mental health (5 items), energy/vitality (4 items), pain (2 items), and general health perception (5 items). For each dimension item scores are coded, summed, and transformed on to a scale from 0 (worst possible health state measured by the questionnaire) to 100 (best possible health state). Two standardized summary scores can also be calculated from the SF-36; the physical component summary (PCS) and the mental health component summary (MCS).

In 1996, a new version of the questionnaire (SF-36v2) was introduced which included improvements in the instructions, the wording of some of the items, and the number of response options for two of the eight scales.<sup>3,4</sup> Several general population studies have confirmed the improved precision, reliability, and validity of the SF-36v2 over the original version.<sup>6,7</sup> Version 2.0 of the SF-36 Health Survey is a product of fifteen years of research and the experience documented in a wide variety of publications.<sup>4</sup>

### EuroQol group (EQ-5D)

EQ-5D is a standardized instrument for use as a measure of health outcome. Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status. EQ-5D is primarily designed for self-completion by individuals. It is cognitively simple, taking only a few minutes to complete. Instructions to respondents are included in the questionnaire.<sup>8</sup>

### Statistical analysis

All statistical analyses were performed with the Statistical Package for the Social Sciences (SPSS) ver. 20.0 (SPSS Inc., Chicago, IL, USA). The agreed level of significance was  $p < 0.05$ . Descriptive statistics were calculated for all variables. This included frequency counts and percentages for nominal variables and measures of central tendency (means, medians) and dispersion (standard deviations, ranges) for continuous variables. The Kolmogorov–Smirnov test was used to assess the distribution. The measurement properties were analyzed in this study for test-retest reliability, construct validity as well as ceiling and floor effects.

The test-retest reliability, which is a measure of stability or reproducibility, represents a scale's capability of providing consistent results when administered on separate occasions.<sup>9,10</sup> To determine the test-retest reliability, 50 patients were asked to complete the SF-36v2 3–7 days after the first assessment. To minimize the risk of short-term clinical change, no treatments were provided during this period. Intraclass correlation coefficients (ICCs) were calculated using a 2-way, mixed-model. Values of 0.4 or greater were considered satisfactory ( $r = 0.81–1.0$ , excellent; 0.61–0.80, very good; 0.41–0.60, good; 0.21–0.40, fair; and

**Table 1**  
Patient demographics.

	N (%)
<b>Education</b>	
Primary school	9 (18)
Middle school	4 (8)
High school	20 (40)
University	17 (34)
<b>Profession</b>	
Housewives	17 (34)
Teacher	3 (6)
Student	12 (24)
Blue collar	5 (10)
White collar	7 (14)
Retired	6 (12)
<b>Diagnosis</b>	
Meniscus\ACL\PCL injury	10 (20)
Spinal pathologies	9 (18)
Lateral epicondylitis	2 (4)
Carpal tunnel syndrome	1 (2)
Frozen shoulder\shoulder dislocation	5 (10)
Hallux valgus	4 (8)
Osteoarthritis	3 (6)
Osteosarcoma	1 (2)
Fibromyalgia	2 (4)
Impingement syndrome\rotator cuff tear	2 (4)
Plantar fasciitis	2 (4)
Patella femoral pain syndrome	3 (6)
Pes planus	1 (2)
Fractures	2 (4)
Kienbock's disease	1 (2)
Hip arthroplasty	1 (2)
Bone marrow edema	1 (2)

Abbreviations: ACL: Anterior cruciate ligament, PCL: Posterior cruciate ligament.

0.00–0.20, poor).<sup>11</sup> Validity is represented by the extent to which a score retains its intended meaning and interpretation.<sup>11</sup> In our study, the construct validity of the Turkish SF-36v2 was analyzed based on its correlation with SF-36v2 subscales and the EQ-5D.

## Results

Table 1 provides the demographic and clinical characteristics of the patients. The descriptive statistics for the scores at baseline and at the second administration of the SF-36v2 were given in Table 2. The measurement properties of the Turkish version of the CMS (test-retest reliability, agreement, construct validity, and floor and ceiling effects) were tested in 50 patients (19 males; mean  $\pm$  SD age: 36.9  $\pm$  14.6 years; range: 16–65 years, BMI; 24.1  $\pm$  4.6) with a variety of musculoskeletal pathologies. The duration of patients' symptoms in the group surveyed was 31  $\pm$  26.5 months. The average  $\pm$  SD interval between the 2 assessments was 3.6  $\pm$  2.2 days. The test-retest assessment of the SF-36v2 subscales indicated excellent reliability, with an ICC of 0.80–0.95 (Table 2). The highest reliability was found between SF-36v2-PCS and SF-36v2-PF ( $r = 0.75$ ), the lowest reliability was found between SF-36v2-PCS and SF-36v2-MH ( $r = 0.05$ ). Correlations between EQ-5D and SF-36v2 subscales ranged from 0.10 (SF-36v2 –VT) to 0.46 (SF-36v2 –RE) (Table 3).

## Discussion

The purpose of this paper was to introduce differences between SF-36 and SF-36v2 to the researchers in Turkey and, the cross cultural adaptation, reliability and validity of the SF-36v2. Based on our sample, the cross cultural adaptation of the SF-36v2 was successfully completed and the SF-36v2 demonstrated acceptable levels of reliability and validity, to be used as a generic questionnaire for Turkish-speaking individuals with a variety of musculoskeletal problems.

SF-36v2 compare to SF-36 is easier to understand, administer. Relative to the standard SF-36 improvements in the content and layout of SF-36v2 included; improvements in some instructions and questions to make the wording less ambiguous, double negative item of the SF-36 was reworded, and five level response sets in place of dichotomous response choices for seven items in the two role functioning scales.<sup>12</sup> There is evidence to suggest that five level response scales improve response rates over dichotomous response categories, such as "yes/no". Consequently the two SF-36 role functioning scales have been changed from dichotomous scales to

**Table 2**  
The mean  $\pm$  SD and the test-retest reliability of the SF-36v2.

	Mean $\pm$ SD		Test-retest reliability (ICC)
	Test 1	Test 2	
SF-36v2 (PF)	46.4 $\pm$ 8.9	48.4 $\pm$ 6.7	0.90
SF-36v2 (RP)	40.8 $\pm$ 11.1	43.7 $\pm$ 9.2	0.80
SF-36v2 (BP)	44.1 $\pm$ 9.2	46.1 $\pm$ 9.4	0.81
SF-36v2 (GH)	43.3 $\pm$ 10.6	44 $\pm$ 10.1	0.95
SF-36v2 (VT)	46.6 $\pm$ 9.4	48.6 $\pm$ 10.4	0.89
SF-36v2 (SF)	45.6 $\pm$ 10.3	46.4 $\pm$ 9.7	0.81
SF-36v2 (RE)	38.4 $\pm$ 9.9	38 $\pm$ 11.5	0.84
SF-36v2 (MH)	39.7 $\pm$ 10.7	41.1 $\pm$ 4.7	0.91
SF-36v2 (PCS)	46.4 $\pm$ 9.6	48.7 $\pm$ 7.8	0.90
SF-36v2 (MCS)	40.7 $\pm$ 10.2	40.7 $\pm$ 11.4	0.91

Abbreviations: BP, bodily pain; GH, general health perceptions; MCS, mental component scale; MH, mental health; PCS, physical component scale; PF, physical functioning; RE, emotional role functioning; RP, physical role functioning; SF, social functioning; VT, vitality; SF-36v2; the Short Form Health Survey version-2.0, ICC, intraclass correlation coefficient.

**Table 3**  
The correlation between the SF-36v2 subscales and EQ-5D.

	SF36 (PCS)	SF36 (MCS)	EQ-5D
SF-36v2 (PF)	0.75*	0.10	0.36*
SF-36v2 (RP)	0.71*	0.17	0.25
SF-36v2 (BP)	0.72*	0.16	0.22
SF-36v2 (GH)	0.61*	0.18	0.15
SF-36v2 (VT)	0.36*	0.57*	0.10
SF-36v2 (SF)	0.25	0.57*	0.17
SF-36v2 (RE)	–0.15	0.73*	0.46*
SF-36v2 (MH)	0.05	0.85*	0.24
SF-36v2(PCS)	–	–	0.20
SF-36v2 (MCS)	–	–	0.26

Abbreviations: SF-36v2; the Short Form Health Survey version-2.0, BP, bodily pain; GH, general health perceptions; MCS, mental component scale; MH, mental health; PCS, physical component scale; PF, physical functioning; RE, emotional role functioning; RP, physical role functioning; SF = social function; VT = vitality.

\*Significant ( $<0.05$ ).

five point response categories thus increasing score precision without increasing respondent liability. Specifically, SF-36v2 achieves a quadruple increase in the number of scale levels, and is intended to produce a substantially smaller standard deviation, as well as to reduce both ceiling and floor effects for both SF-36 role scales. In addition, SF-36v2 includes algorithms for interval level scoring for all eight scales ranging from 0 (for worse health) to 100 (best possible health as measured by the questionnaire) as well as the same standardized scoring (mean = 50, standard deviation = 10) for the SF-36 summary scores (PCS and MCS).

During the cross-cultural adaptation process many changes were made. In the present study, "yards" which is not a Turkish unit used to define a distance being adapted to kilometers. However, some patients still were unable to answer this question because they were unaccustomed to describing walking distance. Instead, they preferred to describe walking duration. The patients felt more comfortable explaining distance as a minute to spent walking. Therefore, we included distance and duration in the questionnaire. Past 4 weeks is usually replace with "geçen 4 hafta" in Turkish however, in our experience, we felt that the patients prefer "son 1 aydır" instead of "geçen 4 hafta." In question 3, the patients were being asked if they can participate "in strenuous sports" but there was no explanation or any example of what is strenuous sports. We add "futbol – basketbol" as an example of strenuous sports to make it the question clearer. Moderate activities was described such as "bowling, or playing golf" but these are not very usual activities in Turkey so we used "masa tenisi" and "bilardo" instead of "bowling and golf". In addition, in question 4, "bodily pain" was translated as "vücut ağrısı" but it is uncommon term in Turkish so we only used "ağrı" instead of "vücut ağrısı." Similarly, "physical" was translated as a "bedensel." We changed it as "fiziksel" since it is widely used and a better option for this translation. Lastly, we prefer translating activity as "aktivite" instead of "etkinlik."

While people were filling out the form, they usually had difficulties in understanding and answering at some specific parts. Especially part 4 and part 5 questions which found confusing by patients. Most people were not sure about what means of these questions. In addition, it was taking more time for answering these question compared to the others. Consequently, "Hedeflediğinizden daha azını mı gerçekleştirdiniz?" was changed to "İsteklerinizi gerçekleştirmekte azalma oldu mu?" In order to simplify the statements of question 4 and 5 "İşinizi veya diğer etkinliklerinizi her zamanki kadar dikkatli yapamıyor muydunuz?" was changed to "İşinizde ya da diğer aktivitelerinizde daha dikkatsiz miydiniz?" and "İş veya diğer etkinliklerinizde kısıtlanma oldu mu?" was converted to "İş veya diğer aktiviteleriniz kısıtlandı mı?" People who fill out the form had troubles in understanding

the statements of the question 11 which were “kesinlikle doğru/yanlış” “çoğunlukla doğru/yanlış.” In order to clarify these confused statements, “kesinlikle katılıyorum/katılmıyorum” and “çoğunlukla katılıyorum/katılmıyorum” were used. In addition, the patients found uncertain of the statement which was “Diğer insanlardan biraz daha kolay hastalanıyorum” so it was converted to “Diğer insanlardan daha kolay hastalandığımı düşünüyorum” (Appendix).

Psychometric properties of the SF-36v2 have been shown in many different patients population such as rheumatoid arthritis, tuberculosis, low back pain in the literature.<sup>13–15</sup> We did not include any specific patient population in the study due to the necessity of cross cultural adaptation. Therefore, it basically aim to compare our reliability and validity analysis in the literature. The latest and comprehensive study was reported by Klooster et al, in Duch population with rheumatoid arthritis.<sup>15</sup> In Duch version, the reliability was reported with internal consistency. We presented the test-retest reliability which is considered excellent with all subscales of the SF-36v2 (Table 2). Internal construct validity of the Duch version of SF-36v2 ranged from 0.79 to 0.95 for subscales and the correlation with SF-36v2 PCS and MCS ranges from 0.19 to 0.91. Internal construct validity was reported in this study between SF-36v2 subscales and SF-36v2 PCS and MCS. The highest correlation was found between SF-36v2 PCS and SF-36v2 PF which was normally expected and found te lowest correlation nd between SF-36v2 PCS and SF-36v2 MH. External construct validity was estimated with EQ-5D. The highest correlation was found between EQ-5D and SF-36V2 role emotional ( $r = 0.46$ ). We believe that this can be explained by the content of the questions of EQ-5D which is more related to psychology.

The primary limitations of our study included the untested statistical power and small sample size. However, previous validation studies have used similar numbers of individuals, and the sample size was large enough to reach statistical significance. Nevertheless, the Turkish SF-36v2 should be applied to larger populations to evaluate its reliability, validity, responsiveness, and minimal clinically important differences in patients with various diagnoses.

A cultural adaptation of the SF-36v2 and its reliability and validity were successfully conducted. The original SF-36 has still been

using in Turkey without cross cultural adaptation which was already disused in the world since 15 years. Therefore, we strongly advice of using cross culturally adapted, reliable and valid Turkish SF-36v2 in Turkish population.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.aott.2016.08.013>.

## References

- Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473–483.
- Turner-Bowker DM, De Rosa, Ware JE. SF-36® health survey. In: Boslaugh S, ed. *Encyclopedia of Epidemiology*. Thousand Oaks, CA: Sage Publications; 2008.
- Ware Jr JE. SF-36 health survey update. *Spine*. 2000;25:3130–3139.
- Ware JE, Kosinski M, Dewey JE. *How to score version 2 of the SF-36 healthsurvey*. Lincoln, RI: QualityMetric Incorporated; 2000.
- Koçyigit H, Aydemir Ö, Fişek G, Ölmez N, Memiş A. Reliability and validity of Turkish version of Short Form 36: a study of patients with rheumatoid disorder. *J Drug Ther*. 1999;12:102–106.
- Jenkinson C. Evaluating the efficacy of medical treatment: possibilities and limitations. *Soc Sci Med*. 1995;41:1395–1401.
- Taft C, Karlsson J, Sullivan M. Performance of the Swedish SF-36 version 2.0. *Qual Life Res*. 2004;13:251–256.
- <http://www.euroqol.org/about-eq-5d.html>.
- de Vet HC, Terwee CB, Bouter LM. Current challenges in clinimetrics. *J Clin Epidemiol*. 2003;56:1137–1141.
- Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. *J Clin Epidemiol*. 2003;56:730–735.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
- Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Assessment of the SF-36 version 2 in the United Kingdom. *J Epidemiol Community Health*. 1999;53:46–50.
- Jirarattanaphochai K, Jung S, Sumananont C, Saengnipanthkul S. Reliability of the medical outcomes study short-form survey version 2.0 (Thai version) for the evaluation of low back pain patients. *J Med Assoc Thai*. 2005;88:1355–1361.
- Atif M, Sulaiman SA, Shafie AA, Asif M, Ahmad N. SF-36v2 norms and its' discriminative properties among healthy households of tuberculosis patients in Malaysia. *Qual Life Res*. 2013;22:1955–1964.
- ten Klooster PM, Vonkeman HE, Taal E, et al. Performance of the Dutch SF-36 version 2 as a measure of health-related quality of life in patients with rheumatoid arthritis. *Health Qual Life Outcomes*. 2013;8:11–77.