

Brief communication

Reliability and construct validity of the SF-36 in Turkish cancer patients

Rukiye Pınar

Department of Medical Nursing, College of Nursing, Marmara University, Istanbul, Turkey (E-mail: rukiyePINAR@yahoo.com)

Accepted in revised form 4 March 2004

Abstract

In this study, we assessed the reliability and construct validity of the SF-36, Turkish version on 419 cancer patients. Cronbach's α coefficients surpassed the 0.70 criterions for all subscales indicating good internal consistency. Results of the test–retest method showed that the stability coefficients for the eight subscales of the SF-36 ranged between 0.81 and 0.94. Principal components factor analysis with varimax rotation confirmed the presence of seven factors in the SF-36: physical functioning, role limitations due to physical and emotional problems, mental health, general health perception, bodily pain, social functioning, and vitality. In conclusion, the Turkish version of the SF-36 is a suitable instrument that could be employed in cancer research in Turkey.

Key words: Cancer patients, Construct validity, Reliability, SF-36

Introduction

The word 'health' may have different connotations for an individual with cancer when compared with a healthy person free of illness and disability. The SF-36 is widely used to measure health status or quality of life (QOL) in both healthy and sick populations [1–11]. Aaronson et al. [12] showed that The Dutch language version of the SF-36 would be a practical, reliable, and valid instrument for cancer patients as well. Although numerous articles exploring the use of Turkish versions of SF-36 with different disease groups [3–4, 8–9] have been published, the literature lacks a related study conducted on patients with cancer. We wished, therefore, to determine the suitability of the SF-36 for assessing QOL in patients with cancer.

Methods

Setting

The study was performed in seven outpatient Oncology Clinics at one state and two university hospitals in Istanbul, Turkey.

Sample

A potential sample of 925 cancer patients visiting seven outpatient oncology clinics between 7 March and 11 April, 2003 met the following inclusion criteria. They should: (1) be at least 18 years of age, (2) be able to complete the questionnaire, and (3) not have any other co-morbidities. A random sampling of these 925 patients' hospital chart codes was used to identify 436 patients and they were asked to participate. Four hundred and nineteen (419) patients provided informed consent and were included in the study.

The mean age of participants was 51 ± 12 (SD) years. The majority of participants were men (60%). The sample represented many different types of cancer, the most prevalent being lung and colon. The sample included the following diagnoses: 171 lung cancer (40.8%), 98 colorectal cancer (23.4%), 62 breast cancer (14.8%), 42 gynaecological cancers (10%), and 46 (11%) others. The mean duration of cancer was 14.8 months. At the time of the study, 146 patients (35%) were undergoing chemotherapy, 84 patients (20%) radiation therapy, and 189 patients (45%)

combined treatment including chemotherapy, radiation therapy, and surgical intervention. One hundred and forty-five patients (34.6%) had metastasis.

Instruments

The SF-36 Health Survey (SF-36)

The SF-36, developed by Ware and colleagues, consists of a multi-item scale which assesses eight health concepts: physical functioning (PF), role limitations due to physical problems (RP), bodily pain (BP), general health perception (GH), vitality (VT), social functioning (SF), role limitations due to emotional problems (RE), and mental health (MH). The SF-36 can be completed within 10 min. Extensive background information on the SF-36, as well as standard scoring algorithms and interpretation guides are available in the literature [10, 11, 13]. Fişek (2002, pers. commun.) [14] has translated the original US version of the standard SF-36 into the Turkish using a forward-backward translation method, some revisions were made in the Turkish adaptations of the SF-36, including the changing of miles to kilometres. A few items were also changed to reflect common Turkish activities, e.g., 'playing football' was used to describe vigorous activities instead of 'participating in strenuous sports'. 'Walking' was used to describe to moderately strenuous activities instead of 'bowling or playing golf'. In Fişek's study (2002, pers. commun.), translated version has been tested in a pilot study in terms of its item scaling assumptions and it was concluded that data appeared to meet scaling assumptions in an overall satisfactory manner. This Turkish version of SF-36 has been approved by the Medical Outcomes Trust [15]. In the present study 'SF-36 Standard Turkish Version 1.0', which is available at Quality Metric Incorporated, was used with official permission.

Procedures

Approval for the study was obtained from the Hospitals' Institutional Review Boards for the Protection of Human Subjects in Research.

Instruments were administered in the hospital education room, a quiet, well lighted room providing an atmosphere in which patients could

concentrate on filling out the questionnaires without being disturbed. The first questionnaire was used to collect information on socio-demographic variables (e.g., sex and age) and cancer-related variables (e.g., type of cancer, metastasis and duration of cancer). Afterwards, patients were asked to self-administer the SF-36.

Reliability evaluations of the SF-36 were based on internal consistency (Cronbach's α coefficient) [16] and test-retest stability (intraclass correlation coefficient-ICC-) [17]. The retest procedure was conducted two weeks after the first test of the SF-36. Two weeks was judged to be optimum retest interval; this would be sufficiently long for patients to forget their initial responses to the 36 items, but not so long that most health domain would change substantially. ICC was chosen as test statistic. All of the subjects provided a second assessment. Although test-retest correlation values as low as $r = 0.21$ have been reported as evidence for reliability, $r > 0.5$ is a more realistic measure [18].

Principal components factor analysis with varimax rotation was used for construct validity [19]. We expected that the scales, originally defined by its authors, would emerge from a factor analysis and items relating to a particular scale would be grouped together within a single factor of data of Turkish patients with cancer. Within such an assessment, a factor should be considered relevant only if its 'eigenvalue' exceeds 1.1 [20].

A statistician performed statistical analysis with the Software Package for Scientific Statistics (SPSS) version 10.0. Significance for all statistical tests was set at the $p = 0.05$ level.

Results

Completeness of the SF-36 data

Firstly, we evaluated completeness rate by using the following data quality indicators: (i) 'Average time required to complete the questionnaire', (ii) 'Completion rates for the SF-36 individual items' (for each item, the number of subjects who completed it divided by the total number of the respondents), and (iii) 'percentage of the respondents'. We calculated the percentage of respondents using the standard proportion procedures recommended for the SF-36 (i.e., if half or fewer

items are missing, calculating a person-specific scale score based on the mean value of the non-missing items) [13, 21].

The SF-36 was given to 419 patients. All 419 patients completed the SF-36 in 8–27 min, with 83% completing the questionnaire in 12 min or less. The average time required to complete the questionnaire was 9.8 min. The rates of missing values for the individual items of the SF-36 were consistently low, ranging from 1 to 4.2%. The percentage of missing scale scores ranged from less than 1 (physical function scores) to 3.6% (role emotional scores).

Reliability

Chronbach's α coefficients for the eight subscales of the SF-36 ranged between 0.79 and 0.90, supporting the internal consistency of the subscales. In this study, 419 patients completed both test and retest questionnaires for a single episode. Stability was acceptable for the Turkish Version of the SF-36 with ICC for test–retest scores exceeding 0.81 for the all subscales (see Table 1).

Construct validity

The SF-36 showed satisfactory construct validity. Factor analysis suggested the presence of 7 factors in the SF-36: PF (factor 1 = 10 items), RP + RE (factor 2 = 7 items), MH (factor 3 = 5 items), GH (factor 4 = 5 items), BP (factor 5 = 2 items), SF (factor 6 = 2 items), and VT (factor 7 = 4 items). Eigen values were 11.70, 4.24, 3.10, 2.16, 1.66, 1.30, and 1.18 respectively. Factor 1 was the

strongest accounting for the majority of the variance (34.7%). While items of the PF, MH, GH, BP, SF, and VT scales were grouped separately under the first, third, fourth, fifth, sixth, and seventh factors, respectively, items of RP and RE scales were grouped under the same factor, namely the second factor. All items had strong loadings of ≥ 0.61 across all the subscales of the SF-36 (see Table 2).

Discussion

Completeness of the SF-36 data

Data quality, defined primarily in terms of 'timely manner' and 'percentage of missing data', was proven in the present study. We showed that the SF-36 is a relatively brief and well-tolerated instrument appropriate for use in studies with Turkish cancer patients. All patients completed the questionnaire in a timely manner indicating that the instrument did not pose a burden. The percentage of missing data ranged from 1 to 4.2% at the item level and less than 1 to 3.6% at the scale level. The rates compare favourably with those reported in the original Medical Outcomes Study in the United States (range of missing items = 1.1–5.9%) [7], a cancer sample (1.2%) [12] and, Swedish study (range from 0.9 to 7.5%) [22]. A slight drop in percent complete items could be seen by age, as expected. The drop was comparable with the US sample [7] for the bodily pain and social functioning scales; however, for all other scales a lower percentage completed was seen in our oldest subgroups.

Reliability

Assessing the reliability of the instrument is essential in determining its ability to measure something in the consistent and reproducible manner. In the present study, the reliability assessment of the SF-36 yielded highly satisfactory results. Chronbach's α values produced by all eight subscales of the SF-36 exceeded 0.70, thus, satisfying Nunnally's criterion for satisfactory internal consistency [23]. In practice, however, well-developed and used instruments are ideally expected to have α values in excess of 0.80 [24]. In the present

Table 1. Reliability of the SF-36 (n = 419)

Scale	Test–retest reliability* (ANOVA ICC)	Internal consistency (Chronbach's α)
Physical functioning	0.93	0.90
Role-physical	0.81	0.87
Bodily pain	0.90	0.86
General health perception	0.94	0.79
Vitality	0.86	0.87
Social functioning	0.88	0.84
Role-emotional	0.87	0.82
Mental health	0.84	0.82

*All correlations are statistically significant at $p < 0.001$.

Table 2. Factor construct of the SF-36 (n = 419)

Subscale and its items	1	2	3	4	5	6	7
Physical functioning (PF)							
3a. <i>Vigorous activities</i>	0.69						
3b. <i>Moderate activities</i>	0.77						
3c. Lifting and carrying groceries	0.64						
3d. Climbing <i>several</i> flights of stairs	0.81						
3e. Climbing <i>one</i> flight of stairs	0.87						
3f. Bending, kneeling, or stooping	0.78						
3g. Walking more than a kilometre	0.87						
3h. Walking several hundred meter	0.92						
3i. Walking a hundred meter	0.93						
3j. Bathing or dressing	0.62						
Role-physical (RP)							
4a. Cut down <i>amount of time</i> for work		0.87					
4b. <i>Accomplished less</i> than you would like		0.84					
4c. Limited in the <i>kind</i> of work		0.88					
4d. Had <i>difficulty</i> performing work		0.86					
Role-emotional (RE)							
5a. Cut down <i>amount of time</i> for work		0.62					
5b. <i>Accomplished less</i> than you would like		0.64					
5c. Didn't do work as <i>carefully</i> as usual		0.64					
Mental health (MH)							
9b. Being a very nervous person?			0.73				
9c. Felt so down in the dumps?			0.75				
9d. Felt calm or peaceful?			0.65				
9f. Felt downhearted and blue?			0.61				
9h. Being a happy person?			0.64				
General health (GH)							
1. In general, would you say your health is:				0.63			
11a. Getting sick easier than other people				0.87			
11b. As healthy as anybody				0.82			
11c. Expect health to get worse				0.64			
11d. Excellent health				0.77			
Bodily pain (BP)							
7. How much <i>bodily pain</i> have you had					0.84		
8. How much did <i>pain</i> interfere with your normal work					0.83		
Social Functioning (SF)							
6. What extent your problems interfered with your social activities?						0.74	
10. How much of the time, your <i>health limited your social activities</i> ?						0.73	
Vitality (VT)							
9a. Did you feel full up pep?							0.78
9e. Did you have a lot of energy?							0.73
9g. Did you feel worn out?							0.74
9i. Did you feel tired?							0.69
Eigenvalue	11.70	4.24	3.10	2.16	1.66	1.30	1.18
Percent (%) observed variance	34.70	12.30	8.60	6.00	4.60	3.60	3.00

Note. Some questions are abbreviated in this table to save space.

study, α values with the exception of 0.79 for the GH, ranged from 0.82 to 0.90, thus, mostly satisfying criterion for a well-developed tool.

To our knowledge, test–retest reliability of the SF-36 has not previously been assessed in patients with cancer. The test–retest approach was favoured for the SF-36 as it fitted easily into the study programme, was appropriate for a self-report questionnaire, and has furthermore been presented as stronger than internal consistency.

Construct validity

In this study, the Turkish SF-36 presented acceptable construct validity. Results of the principle components analysis on the 35 items were satisfactory overall, with the exception of the RP and RE scales' items. These items grouped under the same factor (factor 2), while items relating to a particular scale grouped together within a single factor. The same result was obtained in a Turkish study of diabetes [9]. In a postal survey of a broad sample including 1787 patients suffering from four clinical conditions (low back pain, menorrhagia, suspected peptic ulcer, and varicose veins) [25], the principle components analysis identified five relevant factors. Although the PF, GH, and RE scales were grouped separately under the factors one, four, and five, respectively, the second factor represented the MH and VT scales, and the third factor represented the RP, SF, and BP scales. These different study results provided further evidence of possible cultural influences on participant responses. It seems that the seven-dimensional Turkish SF-36 would be practical and reliable instrument for use in cancer research in Turkey.

The present study revealed that factor loadings on all SF-36 factors were equal to or greater than 0.61, thus, satisfying the criteria that predictive items have loadings ≥ 0.45 [20]. These results indicate that all items were strongly related to their factors. Furthermore, a factor is considered relevant if its eigen value (a statistical measure of its power to explain variation between subjects) exceeds 1.1 [21]. In the present study, eigen values ranged between 1.18 and 11.70, thus satisfying this expectation.

Conclusion

The satisfactory results we obtained in data quality, defined primarily in terms of 'timely manner' and percentage of missing data, internal consistency, test–retest reliability, and construct validity prove that SF-36 Standard Turkish Version 1.0 will be a practical, reliable, and valid instrument for use in cancer population surveys in Turkey.

Acknowledgements

The author thanks to Quality Metric Incorporated for permission to use the SF-36-Turkish Version and Associate professor Ayhan Lash, University of Illinois, US, for revising the paper.

References

1. Parker SG, Peet SM, Jagger C, Farhan M, Castleden CM. Measuring health status in older patients: The SF-36 in practice. *Age Ageing* 1998; 27: 13–18.
2. Hobson P, Bhowmick B, Meara J. Use of the SF-36 questionnaire in cerebrovascular disease. *Stroke* 1997; 28: 464–465.
3. Atagoz K. Quality of life among patients with chronic obstructive lung disease. Unpublished master's thesis, University of Dokuz Eylul, Izmir, Turkey, 1998.
4. Aydin S, Yavuz T, Duven H. Effects of coronary by-pass operations on quality of life in early stage in patients who over 65 years old. *Turk J Geriatr* 2002; 5: 64–67.
5. Chen AY, Frankowski R, Bishop-Leone J, Hebert T, Leyk S, Lewin J, Goepfert H. The development and validation of a dysphasia-specific quality-of-life questionnaire for patients with head and neck cancer: The Anderson dysphasia inventory. *Arch Otolaryngol Head Neck Surg* 2001; 127: 870–876.
6. Jenkinson C, Coulter A, Wright L. Short form 36 (SF-36) health survey questionnaire: Normative data for adults of working age. *Br Med J* 1993; 306: 1437–1440.
7. McHorney CA, Ware JE, Lu JF, Sherbourne JD. The MOS 36-item Short Form Health Survey III: Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994; 32: 40–66.
8. Pinar R, Cinar S, Issever M, Albayrak M, Ilhan S. Quality of life in ESRD: Influence of haemodialysis and renal transplantation. *Cinar Hemsire Derg* 1995; 1: 1–5.
9. Pinar R. Self-reported quality of life and effecting factors in patients with diabetes mellitus. Unpublished doctoral dissertation, University of Istanbul, Istanbul, Turkey, 1995.
10. Ware JE, Sherbourne CD. The MOS 36 item short form health survey (SF 36). *Med Care* 1992; 30: 473–483.

11. Ware JE. Measuring patients' views: The optimum outcome measure. *Br Med J* 1993; 306: 1429–1430.
12. Aaronson NK, Muller M, Cohen PD, et al. Translation, validation, and norming of the Dutch language Version of the SF-36 health survey in community and chronic disease populations. *J Clin Epidemiol* 1998; 51: 1055–1068.
13. Ware JE. User's Manual/SF 36 Health Status Questionnaire. Boston, MA: New England Medical Center, The Health Institute, 1989.
14. Fişek G. Personal communication, 2002.
15. Gandek B. Psychometric analysis of the SF-36 Health Survey: Turkish data. Unpublished manuscript, 1997.
16. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297–334.
17. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing interrater reliability. *Psychol Bull* 1979; 36: 420–428.
18. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford: Oxford University Press, 1989.
19. Ozdamar K. *Statistical Data Analysis with Standard Statistical Programs*. 2nd ed. Eskisehir, Turkey: Kaan Publishing, 1999.
20. Jolliffe IT. *Principal Component Analysis*. New York: Springer, 1986.
21. Ware JE, Gandek B. IQOLA Project Group. The SF-36 Health Survey: Development and use in mental health research and the IQOLA project. *Int J Ment Health* 1994; 23: 49–73.
22. Sullivan M, Karlsson J, Ware JE. The Swedish SF-36 Health Survey-I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. *Soc Sci Med* 1995; 41: 1349–1358.
23. Nunnally JC, Bernstein IH. *Psychometric Theory*. New York: McGraw-Hill, 1994.
24. Brink PJ, Wood MJ. *Advanced Design in Nursing Research*. London: Sage Publications, 1989.
25. Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT. The SF-36 health survey questionnaire: An outcome measure suitable for routine use within the NHS? *Br Med J* 1993; 306: 1440–1444.

Address for correspondence: Rukiye Pınar, Tutuncu Mehmet Efendi Cad, Kaptan Ihsan Sk, Seniha Apt. No: 6, D: 1, 80600, Goztepe/Istanbul, Turkey
Phone: +90-216-418-1606/137 (Office) +90-216-360-6326 (Home); Fax: +90-216-418-3773
E-mail: rukiyePINAR@yahoo.com

Copyright of Quality of Life Research is the property of Kluwer Academic Publishing and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.