

Brief communication (Original)

Reliability, validity, and responsiveness of the short form 36 (SF-36) health survey questionnaire in patients with Graves' disease

Jing Yu^{a,b}, Yi-Yuan Tang^{a,c}

^a*Institute of Neuroinformatics, Dalian University of Technology*, ^b*Department of Nuclear Medicine, No.2 Affiliated Hospital of Dalian Medical University, Dalian 116027, China*, ^c*Department of Psychology, Texas Tech University, TX 79409, USA*

Background: The SF-36 as a generic instrument has been used widely to evaluate health-related quality of life in both healthy subjects and groups of patients. However, no data can be found on its performance in patients with Graves' disease. Hence, this study aimed to assess the reliability, validity, and responsiveness of the SF-36 in a Chinese population of patients with Graves' disease.

Methods: A total of 325 patients with Graves' disease completed interviews that included the SF-36. Internal consistency was measured by Cronbach's α and item-scale correlations. The validity of the SF-36 was studied by means of factor analysis and the association of this scale with sociodemographic and clinical variables. A standardized response mean was used to assess the responsiveness of the SF-36 to change.

Results: Cronbach's α coefficients surpassed the 0.70 criteria for seven subscales and 0.63 for the SF subscale indicating good internal consistency. The correlation coefficients between items and the remainder of the same subscale ranged from 0.51 to 0.96, which were all higher than with other subscales. A principal components factor analysis with varimax rotation identified eight factors. Hamilton Anxiety Rating Scale (HARS) scores correlated negatively to scores at the physical functioning, general health, social functioning, vitality, and role limitations due to emotional problems SF-36 subscales, while HARS scores correlated negatively only to scores on the mental health SF-36 subscale. No correlations were found between SF-36 scores and the levels of thyroid hormones. Responsiveness to improvements in health status was acceptable overall (standardized response means 0.33–0.88).

Conclusion: The Chinese (mainland) version of the SF-36 is a suitable instrument that can be used in patients with Graves' disease.

Keywords: Graves' disease, reliability, responsiveness, SF-36, validity

Besides the classical thyrotoxic symptoms, patients with Graves' disease often have various neuropsychiatric complaints, which include insomnia, dysphoria, anxiety, paranoia, depression, and deterioration of cognitive functions [1-3]. Some of these complaints will have existed for a long time before the patients are diagnosed with Graves' disease, and they may persist for many years even after euthyroidism is achieved [4, 5]. Currently, the therapeutic decisions for these patients are usually based on the serum levels of thyroid hormones without considering the health status subjectively perceived by the patients.

As a multidimensional concept reflecting the overall subjective condition of the individual, health-related quality of life (HRQL) has been used widely either in the assessment of life quality in the general population and patients with chronic disease, or to evaluate the effect of different therapeutic modalities. The SF-36 is a generic HRQL instrument and its usefulness in patients with various diseases has been verified. Several previous studies have demonstrated the detrimental impact of thyroid dysfunction on HRQL [6-9]; however, no data can be found on the feasibility of the SF-36 for patients with Graves' disease.

The purpose of this study was to ascertain the reliability, validity and responsiveness of the SF-36 in a population of patients with Graves' disease.

Correspondence to: Jing Yu, Department of Nuclear Medicine, No.2 Affiliated Hospital of Dalian Medical University, Dalian 116027, China. E-mail: jingyucnn@yeah.net

Materials and methods

Sample

The study was performed in Dalian Medical University Second Affiliated Hospital, China. Data were collected between May 2005 and January 2007. A total of 325 patients with newly diagnosed, untreated Graves' disease were included. They were at least 18 years of age and were able to complete the questionnaire by themselves; informed content was obtained from all patients.

The mean age of participants was 46 ± 13 (SD) years. The majority of participants were women (68%). The diagnosis of Graves' disease was based on increased peripheral thyroid hormones, suppressed thyroid-stimulating hormone (TSH), positive TSH receptor antibodies (TRAb), and diffuse distribution on a Tc-99m pertechnetate scintigraphy of the thyroid gland. The laboratory variables (peripheral thyroid hormones, TSH and TRAb) were measured by standard methods. Patients with any comorbidity or history of psychiatric disorders were excluded from the study.

Instruments and analysis

The Chinese (mainland) version of SF-36 was used to assess the patients' quality of life under the approval of the Institute of Social Medicine and Family Medicine, Zhejiang University. The instrument consists of 36 items that belong to eight health-related aspects of life: physical functioning (PF), role limitations due to physical problems (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role limitations due to emotional problems (RE), mental health (MH), and one single item subscale on health transition. Each of the eight subscales generates a score from 0 to 100, with "0" implies worst health status and "100" implies best health status. Some revisions were made in the Chinese adaptations of the SF-36 [10], e.g., mopping the floor and practicing Tai Chi were used as complementary examples of moderate activities, one mile was translated into 1500 meters, and one block was translated into the distance between two street crossings. The translated version has been tested among 1000 households in 18 communities of Hangzhou (the capital of Zhejiang Province, southeast of mainland China) in terms of its scaling assumptions and it was concluded that the Chinese (mainland) version of the SF-36 functioned similarly to the original American population tested [11, 12].

The Hamilton Depression Rating Scale (HDRS) was used to evaluate depression levels and changes in severity. The reliability and validity of its Chinese version has been confirmed. In this study, the 17-item version was used. The Hamilton Anxiety Rating Scale (HARS) was used to evaluate levels of anxiety and changes in severity. The reliability and validity of its Chinese version has been confirmed. It includes 14 items examining both physical and psychological symptoms.

The SF-36 was administered in a quiet, well-lit room providing an atmosphere in which patients could concentrate on the questionnaires without being disturbed. The first questionnaire was used to collect information on sociodemographic variables and disease-related variables (e.g., family history, duration of disease and symptoms). Afterwards, patients were asked to complete the SF-36 by themselves. A psychiatrist interviewed the patients, and the HDRS and HARS questionnaires were completed.

Internal consistency was used to assess the reliability of the SF-36. Reliability is considered to be acceptable for group comparisons when Cronbach's α exceeds 0.7 [13].

The hypothesis adopted in the analysis of the construct validity was that the HRQL in patients with Graves' disease would be associated inversely with the patient's age, with levels of thyroid hormones and with scores in depression or anxiety. Moreover, according to other studies, it was considered that the HRQL would be more severely impaired in female patients.

A principal component factor analysis with varimax rotation was performed to determine the structure of the SF-36 in our population. We expected that the subscales, originally defined by its authors, would emerge from a factor analysis and items relating to a particular subscale would be grouped together within a single factor. A factor should be considered relevant only if its eigenvalue exceeds 1.1 [14].

After the initial assessment, 293 patients underwent antithyroid drug treatment and 32 patients underwent I-131 treatment. At the time euthyroidism was achieved, a standardized response mean (SRM) [15] was calculated for each SF-36 subscale to assess the responsiveness of the SF-36 to change.

Statistical analysis

A statistician performed statistical analysis with the Software Package for Scientific Statistics (SPSS)

version 13.0. Significance for all statistical tests was set at the $p = 0.05$ level.

Results

Data quality

A total of 347 patients were identified for recruitment, of whom 22 declined to take part and 325 (93.7%) were recruited to the study. The completion rates for the SF-36 individual items were consistently high, ranging from 96.0% to 99.7%. All responses to SF-36 questions were within possible answers and all response choices were used for all questions. All 325 participants completed the SF-36 in 8–29 min, with 85% completing the questionnaire in 14 min or less. The average time required to complete the questionnaire was 10.5 min. The percentage of respondents was calculated according to the standard proportion procedures recommended for the SF-36 [16]. The percentage of missing subscale scores ranged from 0.2% (physical function scores) to 2.8% (social function scores). Noncompletion rate was higher for patients of an age over 50 years. Women had higher noncompletion rates than men in the general health subscale; otherwise there were no major differences.

Reliability

Cronbach's α coefficients for the eight subscales of the SF-36 ranged from 0.63 to 0.92, satisfying the internal consistency of the subscales (Table 1).

The correlation coefficients between items and the remainder of the same subscale ranged from 0.51 to 0.96, and they were all higher than with other subscales, providing further evidence of internal consistency.

Construct validity

The older patients showed significant lower scores ($p > 0.05$) in PF, BP, and GH subscales. In other

subscales such as RE and VT, younger patients showed lower scores, but these reductions were not statistically significant. Women scored significantly lower in most subscales ($p > 0.05$) except RE, SF, and BP.

Most patients had mild to moderate depression or anxiety symptoms, and the psychiatric scores of these patients as measured with HDRS and HARS were significantly high. HDRS score correlated negatively to scores at the SF-36 subscales of PF, GH, SF, VT and RE, while HARS score correlated negatively only to score at the SF-36 subscales of MH. No correlations were found between SF-36 scores and the levels of thyroid hormones.

Factorial analysis identified eight factors: PF (factor 1 = 10 items), RE +SF (factor 2 = 5 items), MH (factor3 = 5 items), VT (factor 4 = 4 items), RP (factor 5 = 4 items), GH1, 3, 5 (factor 6 = 3 items), BP (factor 7 = 2 items), and GH2, 4 (factor 8 = 2 items). Factor 1 was the strongest accounting for 35.9% of the variance, while items of the RE and SF subscales were contained in the same factor, namely the second factor. Eigenvalues were between 1.17 and 11.01, and all items had strong loadings of at least 0.57 across all the subscales of the SF-36 (Table 2).

Responsiveness to change

A total of 302 patients were available for the review, others had withdrawn because they left the city or were unwilling to continue. Scores at the SF-36 subscales improved significantly, except for that of general health perception and bodily pain. The SF-36 subscales of RP have an SRM > 0.8 , representing large clinical change. Moderate clinical change, i.e. SRMs between 0.5 and 0.8, was found for three subscales: PF, VT, and RE. The remaining SF-36 subscales reported SRMs between 0.2 and 0.5, representing small to moderate change (Table 1).

Table 1. Reliability and responsiveness of the SF-36

Subscale	Cronbach's α (n = 325)	SRM (n = 302)
Physical functioning	0.92	0.86
Role-physical	0.88	0.88
Bodily pain	0.84	0.33
General health perception	0.76	0.35
Vitality	0.85	0.57
Social functioning	0.63	0.38
Role-emotional	0.84	0.76
Mental health	0.83	0.42

Table 2. Factor construct of the SF-36 (n = 325)

Subscale and its items		1	2	3	4	5	6	7	8
PF	PF1	0.65							
	PF2	0.79							
	PF3	0.71							
	PF4	0.78							
	PF5	0.81							
	PF6	0.73							
	PF7	0.82							
	PF8	0.87							
	PF9	0.86							
	PF10	0.57							
RE	RE1		0.78						
	RE2		0.8						
	RE3		0.81						
SF	SF1		0.66						
	SF2		0.71						
MH	MH1			0.75					
	MH2			0.74					
	MH3			0.73					
	MH4			0.68					
	MH5			0.7					
VT	VT1				0.72				
	VT2				0.65				
	VT3				0.59				
	VT4				0.6				
RP	RP1					0.86			
	RP2					0.86			
	RP3					0.85			
	RP4					0.81			
GH	GH1						0.61		
	GH2								0.77
	GH3						0.79		
	GH4								0.62
	GH5						0.73		
BP	BP1							0.85	
	BP2							0.87	
Eigenvalue		11.01	4.12	2.86	2.03	1.67	1.45	1.34	1.17

Discussion

The SF-36 is considered as a generic measure of HRQL with the potential for widespread application to healthy subjects and groups of patients [17-25]. In recent years, this scale has been used widely because it was recommended by the International Quality of Life Assessment Committee. The SF-36 has been translated for different clinical conditions [26], makes it one of the most frequently used scales in quality of life evaluation. However, the assessment of the performance of the scale is necessary because it has not been evaluated in patients with Graves' disease.

The feasibility of an instrument can be reflected mainly by its completion rate and the time required to complete it. The present study shows that the SF-36 is a concise instrument, which can be accepted by Chinese patients with Graves' disease. All patients completed the questionnaire in a relatively short time, indicating that the instrument could be well-tolerated. The percentage of missing data ranged from 0.3% to 4.0% at the item level and 0.2% to 2.8% at the subscale level, which is consistent with those reported in the original Medical Outcomes Study in the United States [14].

Cronbach's α produced by the seven subscales of the SF-36 exceeded 0.70, satisfying the recommended criterion for internal consistency [13]. The SF subscale was the least satisfactory in the reliability testing, which is consistent with the results from the general Chinese population and could be explained by the few items in this subscale and lower homogeneity [10].

Depression and anxiety have been reported prevalent in patients with thyrotoxicosis [2, 3, 27], in most cases, these psychiatric symptoms remit with the treatment of thyroid dysfunctions. One prospective study [6], examining HRQL with the SF-36 in 30 Graves' patients, showed that lower SF-36 scores correlated significantly with the level of depression and anxiety symptoms. The negative correlation between HDRS or HARS and the SF-36 subscales found in the present study demonstrated the widespread impact of depression and anxiety on HRQL, and provided further evidence for the capacity of SF-36 to identify changes in depression and anxiety.

Regarding the factorial structure of the scale, results of the principle components analysis are close to those found by the original authors in five subscales. The items in the RE and SF subscales grouped under the same factor (factor 2), and the five items of the subscale GH were separated into factor 6 and 8. The results were similar to those obtained in a Chinese study of another autoimmune disease (SLE) [28].

Construct validity is considered to be acceptable when predictive items have loadings ≥ 0.45 [14] and the eigenvalue of any factor exceeds 1.1. In the present study, factor loadings on all SF-36 factors were no less than 0.57 and eigenvalues ranged between 1.17 and 11.01, thus satisfying the criteria recommended for factorial analysis.

A well-developed HRQL instrument should have the ability to measure changes in health status over time or after treatment. If the size of SRM is taken as a measure of responsiveness, the ability of the SF-36 to detect response to change in Graves' patients is acceptable overall, with the exception that no significant change can be found in GH and BP. Former studies showed that the BP subscale of SF-36 was associated with the physical dimension of health more than the mental dimension [29], and it was not significantly impaired with hyperthyroidism [6, 7], whereas the pain scale of NHP picked up the distress generated by this disease more precisely. Small SRMs may imply a lack of instrument responsiveness, or

reflect a minor change in health quality. The SF-36 mental health subscale was less responsive than the HDRS and the HARS, which is similar to findings by Ruta [30] in a RA population, showing the necessity of the combined use of these instruments.

In the present study, no correlations were found between SF 36 scores and the levels of thyroid hormones. This is consistent with a recent Swedish study [31] in which SF 36 scores were not correlated to TSH levels or associated with suppressed TSH. However, the history of GD did impair the long-term HRQL compared with a healthy, age-matched reference population, especially with regard to mental performance and 'vitality' [32]. Even slight changes in the thyroid state such as subclinical thyrotoxicosis or hypothyroidism may lead to reduced HRQL scores [33].

GD is a complex autoimmune disease. Besides the classical thyrotoxic symptoms, patients with Graves' disease often have various neuropsychiatric complaints. Some of these complaints may have existed for a long time before the diagnosis of Graves' disease, and they may persist for many years even after euthyroidism is achieved. Abnormal thyroid hormone level is only part of the disease. The lack of correlation between SF36 scores and thyroid hormone does not invalidate the usefulness of SF 36 in GD.

In conclusion, the results obtained in data quality, internal consistency, construct validity, and responsiveness to change were satisfactory overall, proving that the Chinese (mainland) version SF-36 will be a practical, reliable, valid and responsive instrument for use in patients with GD.

Acknowledgements

The authors thank to the Institute of Social Medicine and Family Medicine, Zhejiang University, for permission to use the SF-36 Chinese (Mainland) version. This work was supported by 973 Program 2012CB518200. None of the authors have any conflict of interest to report.

References

1. Geffken GR, Ward HE, Staab JP, Carmichael SL, Evans DL. Psychiatric morbidity in endocrine disorders. *Psychiatr Clin North Am.* 1998; 21:473-89.
2. Kathol RG, Delahunt JW. The relationship of anxiety and depression to symptoms of hyperthyroidism using operational criteria. *Gen Hosp Psychiatry.* 1986; 8: 23-8.

3. Trzepacz PT, McCue M, Klein I, Levey GS, Greenhouse J. A psychiatric and neuropsychological study of patients with untreated Grave's disease. *Gen Hosp Psychiatry*. 1988; 10:49-55.
4. Fahrenfort JJ, Wilterdink AM, van der Veen EA. Long-term residual complaints and psychosocial sequelae after remission of hyperthyroidism. *Psychoneuroendocrinology*. 2000; 25:201-11.
5. Scheffer C, Heckmann C, Mijic T, Rudorff KH. Chronic distress syndrome in patients with Graves' disease. *Med Klin*. 2004; 99:578-84.
6. Elberling TV, Rasmussen AK, Feldt-Rasmussen U, Hording M, Perrild H, Waldemar G. Impaired health-related quality of life in Graves' disease. A prospective study. *Eur J Endocrinol*. 2004; 151:549-55.
7. Bianchi GP, Zaccheroni V, Solaroli E, Vescini F, Cerutti R, Zoli M, et al. Health-related quality of life in patients with thyroid disorders. *Qual Life Res*. 2004; 13:45-54.
8. Gulseren S, Gulseren L, Hekimsoy Z, Cetinay P, Ozen C, Tokatlioglu B. Depression, anxiety, health-related quality of life, and disability in patients with overt and subclinical thyroid dysfunction. *Arch Med Res*. 2006; 37:133-9.
9. Razvi S, Ingoc IE, McMillan CV, Weaver JU. Health status in patients with sub-clinical hypothyroidism. *Eur J Endocrinol*. 2005; 152:713-17.
10. Li L, Wang HM, Shen Y. Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalization. *J Epidemiol Community Health*. 2003; 57:259-63.
11. Hemingway H, Stafford M, Stansfeld S, Shipley M, Marmot M. Is the SF-36 a valid measure of change in population health? Results from Whitehall II study. *BMJ*. 1997; 315:1273-9.
12. Hemingway H, Nicholson A, Stafford M, Roberts R, Marmot M. The impact of socioeconomic status on health functioning as assessed by the SF-36 questionnaire: the Whitehall II study. *Am J Public Health*. 1997; 87:1484-90.
13. McHorney CA, Ware JE, Rachel Lu JF, Sherbourne CD. The MOS 36-item Short-form Health Survey (SF-36): III. Test of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care*. 1994; 32:40-66.
14. Jolliffe IT. *Principal Component Analysis*. New York: Springer; 1986. p. 62-70.
15. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis*. 1987; 40:171-8.
16. 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.
17. Hobson P, Bhowmick B, Meara J. Use of the SF-36 questionnaire in cerebrovascular disease. *Stroke*. 1997; 28:464-5.
18. Chen AY, Frankowski R, Bishop-Leone J, Hebert T, Leyk S, Lewin J, et al. 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; 127:870-6.
19. 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; 306:1437-40.
20. Ware JE, Sherbourne CD. The MOS 36 item short form health survey (SF-36). *Med Care*. 1992; 30:473-83.
21. Pinar R. Reliability and construct validity of the SF-36 in Turkish cancer patients. *Qual Life Re*. 2005; 14:259-64.
22. Failde I, Ramos I. Validity and reliability of the SF-36 Health Survey Questionnaire in patients with coronary artery disease. *J Clin Epidemiol*. 2000; 53:359-65.
23. Koh ET, Leong KP, Tsou IY, Lim VH, Pong LY, Chong SY, et al. The reliability, validity and sensitivity of the Chinese version of SF-36 in oriental patients with rheumatoid arthritis. *Rheumatology*. 2006; 45:1023-8.
24. Riley ED, Bangsberg DR, Perry S, Clark RA, Moss AR, Wu AW. Reliability and validity of the SF-36 in HIV-infected homeless and marginally housed individuals. *Qual Life Res*. 2003; 12:1051-8.
25. van der Heijden PG, van Buuren S, Fekkes M, Radder J, Verrips E. Unidimensionality and reliability under Mokken scaling of the Dutch language version of the SF-36. *Qual Life Res*. 2003; 12:189-98.
26. Jenkinson C, Peto V, Coulter A. Making sense of ambiguity: evaluation in internal reliability and face validity of the SF-36 questionnaire in women presenting with menorrhagia. *Qual Health Care*. 1996; 5:9-12.
27. Bunevicius R, Velickiene D, Prange AJ Jr. Mood and anxiety disorders in women with treated hyperthyroidism and ophthalmopathy caused by Graves' disease. *Gen Hosp Psychiatry*. 2005; 27:133-9.
28. Liu J, Shu Q, Wang JZ. Assessing the quality of life of patients with systemic lupus erythematosus using SF-36. *Chin J Public Health (Chin)*. 2004; 11:1383-4.
29. McHorney CA, Ware JE, Raczek AE. The MOS 36-items Short-form Health Survey (SF-36) II. Psychometric and clinical test of validity in measuring

- physical and mental health constructs. *Med Care*. 1993; 31:247-63.
30. Ruta DA, Hurst NP, Kind P, Hunter M, Stubings A. Measuring health status in British patients with rheumatoid arthritis: reliability, validity and responsiveness of the short form 36-item health survey (SF-36). *Br J Rheumatol*. 1998; 37:425-36.
31. Abraham-Nordling M, Wallin G, Lundell G, Töring O. Thyroid hormone state and quality of life at long-term follow-up after randomized treatment of Graves' disease. *Eur J Endocrinol*. 2007; 2:173-9.
32. Abraham-Nordling M, Töring O, Hamberger B, Lundell G, Tallstedt L, Calissendorff J, Wallin G. Graves' disease: a long-term quality-of-life follow up of patients randomized to treatment with antithyroid drugs, radioiodine, or surgery. *Thyroid*. 2005; 11: 1279-86.
33. Biondi B, Palmieri EA, Klain M, Schlumberger M, Filetti S, Lombardi G. Subclinical hyperthyroidism: clinical features and treatment options. *Eur J Endocrinol*. 2005; 152:1-9.