

Research

Open Access

Cultural Issues in Using the SF-36 Health Survey in Asia: Results from Taiwan

Hsu-Min Tseng*¹, Jui-fen Rachel Lu*¹ and Barbara Gandek²

Address: ¹Health Databank Management Lab, Department of Health Care Management, College of Management, Chang Gung University, TAIWAN and ²Health Assessment Lab, Waltham, Massachusetts, USA

Email: Hsu-Min Tseng* - tsenghm@mail.cgu.edu.tw; Jui-fen Rachel Lu* - rachel@mail.cgu.edu.tw; Barbara Gandek - bgandek@hal-health.org
* Corresponding authors

Published: 26 November 2003

Received: 30 July 2003

Health and Quality of Life Outcomes 2003, 1:72

Accepted: 26 November 2003

This article is available from: <http://www.hqlo.com/content/1/1/72>

© 2003 Tseng et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The feasibility of using the SF-36 in non-Western cultures is important for researchers seeking to understand cultural influences upon health status perceptions. This paper reports on the performance of the Taiwan version of the SF-36, including the implications of cultural influences.

Methods: A total of 1191 volunteered subjects from the general population answered the translated SF-36 Taiwan version, which was developed following IQOLA project protocols.

Results: Results from tests of scaling assumptions and reliability generally were satisfactory. Convergent validity, as assessed by comparing the SF-36 to a mental health oriented inventory, was acceptable. Results of principal components analysis were similar to US results for many scales. However, differences were seen for the Vitality scale which was a stronger measure of mental health than physical health in Taiwan. Results are compared to those from other Asian studies and the U.S.

Conclusion: The results raise important questions regarding cultural influences in international studies of health status assessment. Further research into the conceptualization and components of mental health in Asian countries is warranted.

Background

The health outcomes of a population can be measured in terms of etiology and pathogenesis. Nevertheless, well-developed health outcome measurement systems have expanded the measurement of health beyond the classical endpoints of mortality and morbidity in clinical practice [1,2]. Health-related quality of life has emerged as the new reflection of modern medicine as viewed from biopsychosocial perspectives. With the fast growth in health care expenditures, this concept has been increasingly used as an important attribute in patient care and clinical studies as well as health economic evaluations [3].

Over the past 20 years, health status measures have been widely applied in different medical fields such as oncology, cardiology, arthritis, and psychiatry [4]. Health researchers have begun to evaluate whether common standardized health status measures are technically and conceptually equivalent for various socio-cultural groups. Hence, there is an increasing need for international standards to measure health status in a manner that allows comparisons across countries, but which also are relevant within individual cultures. In particular, the well-recognized differences between Western and Eastern cultures

may well be reflected in health status measurement results.

Encountering diversified cultural backgrounds, researchers hence have to take even more cautious steps in translating well-established standard instruments in Asia [5]. Extensive psychometric testing is also required for the translated instruments. Amongst those instruments which have become standard in the health status field, the SF-36 is one of the most widely accepted, extensively translated and tested instruments around the world. The International Quality of Life Assessment (IQOLA) Project was formed in 1991 and has developed a standard protocol for translating and psychometrically testing the SF-36 in different language versions[6,7].

As one aspect of health status assessment is to measure an individual's physical and mental state, respondents are often sensitive to wording which reflects differences in ethnicity and culture, even if the language used is the same in a broad sense. English and Chinese provide good examples of this. Although most of the words are similar, there are US English and U.K. English versions of the SF-36 Health Survey, reflecting linguistic differences in the two cultures. Similarly, the IQOLA Project has collaborated on the development of several Chinese versions of the SF-36 Health Survey. The published Chinese versions are for the US Chinese population[8] and Hong Kong (the HK translation is being used in some other Southeast Asian countries) [9]. Results of psychometric testing of these Chinese versions suggested that scaling assumptions were generally met and conceptual equivalence was achieved in comparison with forms using Western languages. However, both Chinese-American studies reported less satisfactory psychometric results for the mental health and vitality scales, as evidenced by high correlations between items in both scales. These results suggest that health-related quality of life (HRQOL) measures may need to be interpreted within a cultural framework.

Although these studies have begun to address the need for different versions of Chinese health surveys, one issue is that the samples used in these studies (either Chinese Americans or Hong-Kong Chinese) have been somewhat adapted to Western cultures from an ecological point of view. Thus, results of those studies may be influenced by a mixture of Chinese and Western cultures. Within a cultural framework, people in Taiwan are more influenced by Chinese culture than Western ones as a legacy of historical and political developments. Thus, the Taiwan population provides a distinct culture with which to address influences of Chinese cultural factors upon QOL measures. In this paper, we document the adaptation process of the SF-36 for use in Taiwan, by presenting the translation process for the SF-36 Taiwan version and the results of psycho-

metric tests performed on *three* different groups. The implications of these results are then discussed within a cultural framework.

Methods

Design and Sample

The standard Taiwan version of SF-36 was administered to a total of 1,191 volunteered subjects who participated in three studies of health status surveys. The mean age of all the respondents was 28.2 (S.D. = 14.2) years; 64% were female. The first sample contained data from 614 freshmen students in one private university (mean age = 18.3, S.D. = 0.66; 46.9% female), who responded to the SF-36 Taiwan version (among other scales) in a screening survey of psychological well-being. The second was based on a study that examined the impact of organizational factors upon health. A total of 501 employees of a medical institute were approached and among them 491 employees (mean age = 34.9, S.D. = 10.06; 85.9% female) returned valid SF-36 data for analysis. The third was a group of 76 elderly people taking part in a study on the effects of Tai-Chi practice upon physical strength and balance (mean age = 65.9, S.D. = 5.43; 60.5% female). After excluding questionnaires with invalid data, a total of 1181 SF-36 profiles were available for analysis.

In the student sample, an additional questionnaire, the "Stress Coping Inventory", was administered to test the criterion validity of the *mental health dimension* in the translated SF-36 version. A total of 569 students returned valid data for this analysis.

Measures

SF-36 Health Survey

The SF-36 measures eight health concepts: 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), and mental health (MH), which were selected from the Medical Outcomes Study (MOS) inventory [10]. During its development, extensive and thorough psychometric testing was performed not only on the general population but also on diverse disease groups, and comparisons with other established instruments were also made[11,12]. Before the development of the SF-36 Taiwan standard version in 1996, there were a few researchers who had independently translated the SF-36 into Mandarin, the official language used in Taiwan. The translators of three versions were native Mandarin speakers with a fair amount educational training in the US; most possessed doctoral degrees from major US universities. The first stage of the development of the SF-36 Taiwan version was to reconcile the three existing versions into one single version. All translators and some experienced users of the SF-36 (US version) were invited to

discuss the three versions in a coordination meeting. All participants in the meeting were asked to review the three translated versions and agree on the final translation item by item. The criteria of agreement were based on clarity (*is it clear?*), common language use (*is it easy to understand?*) and conceptual equivalence (*is the concept measured in original US-English version captured?*). In terms of common language use, the convenience of reading the translation into the local dialect, Taiwanese, was especially taken into consideration because Taiwanese is more commonly used in older populations. As an example of conceptual equivalence, the list for moderate activities in item 3 b was slightly modified to substitute "playing Tai-Chi" for "playing golf", since golf is not a common sport among the general population of Taiwan. Other activities were changed with equivalent concepts included "pushing a vacuum cleaner" in item 3 b ("mopping the floor" is used instead) and units for distance (e.g. kilometer is used to substitute for miles) in item 3 g.

After the consolidated version was developed, it was evaluated by a focus group, which was composed of experts in the fields of public health, psychology, psychometrics, nursing, social work and family medicine. In the focus group meeting, members reviewed the questionnaire item by item and evaluated the consolidated translation, based on their training and expertise. Modifications were made when it was deemed necessary, and the members agreed upon a second consolidated version which was then backward translated into English. The Principal Investigator for the Taiwan team and the IQOLA project director had an extensive discussion on problematic items and response choices. The Taiwanese SF-36 standard version then was produced and ready to be field tested.

Stress Coping Inventory

The Stress Coping Inventory was developed by Chen and Wu (1987) [13], based on the bio-psychosocial model. It aims to investigate how personal resources influence the mental health outcome of stress coping among university students. The questionnaire includes 52 items and has two parts: evaluation of individual resources for stress coping and investigation of various psychological symptoms. Higher scores in the resources dimension represent persons with more resources (either personal or social) to cope with stress. On the contrary, higher psychological symptoms mean worse mental health. Personal resources are composed of three subscales, namely "self-esteem," "friendship support," and "family support". Two subscales labeled as "anxiety reaction" and "depression reaction," were related to psychological symptoms. The questionnaire has acceptable internal consistency (Cronbach's α ranged from 0.74 to 0.95).

Data Analysis

Psychometric analyses included tests of assumptions underlying the construction of SF-36 scales, analysis of principal components to test the hypothesized structure of the SF-36, and tests of construct validity using criterion-based and construct approaches. Following the IQOLA procedure, the Multitrait Analysis Program-Revised (MAP-R) for Windows was used to test whether the scores satisfied summated-rating scaling assumptions[14]. Assumptions underlying the scoring of SF-36 data from each of these studies were evaluated to determine if it was appropriate to use the method of summated ratings to score the SF-36 scales, following standard SF-36 scoring algorithms[15]. Internal consistency reliability for each scale score was estimated using Cronbach's alpha coefficient. Because higher levels of reliability increase statistical power, a minimum reliability of 0.70 for measures used in-group comparisons has been recommended[16]. In addition, the percentages of subjects achieving either the highest score (ceiling) or lowest score (floor) were calculated, because a large ceiling or floor effect will limit the ability of SF-36 to detect change over time.

Validity was tested using construct and criterion-based approaches. The SF-36 was constructed to represent two major dimensions of health – physical and mental. A second-order factor analysis (using principal component analysis) of the 8 SF-36 scale scores was carried out to test the assumption that there were two underlying factors in the SF-36. Two factors with eigenvalues greater than 1 were extracted and rotated to orthogonal simple structure using the varimax method. To interpret the two components, we then examined the strength of their correlations with the eight scales. In addition, the results were compared to published results from HK, US, and Japan[9,17,18]. An additional test of construct validity was conducted by comparing SF-36 scale scores to scores from the Stress Coping Inventory, which was constructed to investigate the concept of mental health. It is therefore anticipated that scales of the Stress Coping Inventory within the resources part will positively correlate with SF-36 scales related to the mental health dimension, while a reverse relationship will be found for those within the psychosomatic symptoms part. In addition, a stronger association is anticipated with the mental health dimension than with the physical health dimension in the SF-36 because the construction of the Stress Coping Inventory is more mentally oriented. Criterion-based validity was tested by comparing the elderly group with the university student one. It is anticipated that subscales of the physical health dimension (e.g. PF, RP, BP) will decline with increasing age, while those in the mental health dimension (i.e. MH, RE) will be less influenced by age.

Table 1: Descriptive Statistics for SF-36 Scales

	Mean	SD	Floor (%)	Ceiling (%)
PF – Physical Functioning	92.60	11.47	0.0%	45.3%
RP – Role–Physical	83.56	28.88	5.9%	67.9%
BP – Bodily Pain	82.38	16.75	0.2%	31.1%
GH – General Health	67.49	18.21	0.1%	2.3%
VT – Vitality	65.32	15.15	0.0%	0.9%
SF – Social Functioning	79.35	16.00	0.1%	19.5%
RE – Role–Emotional	71.32	36.98	14.1%	55.4%
MH – Mental Health	68.43	14.67	0.0%	1.6%

Results

The psychometric properties of the translated version are presented in terms of data quality, reliability, and validity.

Data Quality and Descriptive Statistics

A total of 1,181 records were available for psychometric testing of the SF-36 Taiwan version. Table 1 presents scale means and standard deviations, and the percentage scoring at the ceiling and floor for the SF-36 scales. The results of the item descriptive statistics indicate that SF-36 Taiwan version have a high rate of data completeness (Table 2). The rates of missing values on the item level were consistently low, ranging from 0.0% (GH1) to a high of 2.7% (GH2). As would be expected for a sample that is primarily composed of healthy respondents, response distributions tended to be skewed in the direction of positive health. This is especially evidenced by the results that substantial ceiling effects were more frequently encountered in scales measuring functional limitations (e.g. PF, RP, and RE). Conversely, the percentage of respondents scoring at the lowest scale level (i.e., floor effect) was minimal in that floor effects were observed in less than 1% of the sample for all but the two role functioning scales (RP and RE).

Test of Scaling Assumptions

To evaluate scaling assumptions underlying scoring of the SF-36 scales, item variability, item-internal consistency and item discriminant validity were assessed. Table 2 summarizes the results of the item descriptive statistics and the Pearson item-scale correlations between each item and scale. All of the correlations between each item and its hypothesized scale (i.e. item-scale internal consistency) corrected for overlap exceeded 0.40, ranging from 0.40 (MH1) to 0.68 (GH5). However, these are low for the SF scale. Item-scale correlations were roughly equivalent within each scale, although correlations were slightly lower for PF10 (bathing and dressing), GH4 (expect health to get worse), RE3 (didn't do work or activities as carefully as usual), and MH1 (nervous person). Item means and standard deviations generally were roughly

equivalent within a scale, with some exceptions previously noted in other studies[19]. The mean values of VT4 (tired) and MH1 (nervous person) were lower than expected, however.

Test of item discriminant validity focus on the integrity of hypothesized item groupings relative to the health concepts hypothesized. According to the IQOLA protocol[7], an item was considered to have "succeeded" in the test of item discriminant validity if the correlation between an item and its hypothesized scale is statistically and significantly higher (i.e. > 2 standard errors) than the correlations between that item and all scales other than its hypothesized scale. All items passed the test for discriminant validity except the VT and SF subscales which had rates below 100%. Almost all the items in the VT and SF subscales overlapped with those in the MH subscale. In the vitality scale, VT4 which assessed the lack of energy had the same high correlations with the mental health scale (0.53) as its hypothesized scale. In the social functioning scale, SF2 was more highly correlated with the mental health and vitality scale (0.42 and 0.43, respectively) than its hypothesized scale.

Reliability

Internal consistency reliability statistics for the eight SF-36 scales are presented on the diagonal of Table 3. All scales met or exceeded the 0.70 level recommended for group comparisons, with the exception of the SF scale (Cronbach's alpha = 0.57). Inter-scale correlation analysis also revealed that the scale constructs for the translated Taiwan SF-36 version were generally distinct. Most of the inter-scale correlation coefficients were medium to low, and higher coefficients were found between scales which represented similar constructs (e.g. vitality and mental health) than those with competing constructs (e.g. role emotional vs. physical functioning). However, the correlation between the MH and VT scales (0.69) was nearly as high as the reliability of the two scales.

Table 2: Item Descriptive Statistics and Pearson Item-Scale Correlations Corrected for Overlap

Scale (Choice Range)	Item	Missing	Mean	SD	PF	RP	BP	GH	VT	SF	RE	MH
Physical Functioning (1–3)	PF1	1.1%	2.44	0.62	0.51^a	0.40	0.31	0.34	0.22	0.2	0.14	0.13
	PF2	1.1%	2.87	0.38	0.65^a	0.33	0.29	0.25	0.17	0.18	0.05	0.11
	PF3	1.1%	2.91	0.32	0.59^a	0.29	0.28	0.18	0.11	0.14	0.07	0.06
	PF4	1.0%	2.75	0.48	0.58^a	0.32	0.28	0.23	0.17	0.15	0.11	0.12
	PF5	1.2%	2.96	0.22	0.60^a	0.25	0.25	0.18	0.16	0.18	0.13	0.10
	PF6	1.0%	2.88	0.36	0.58^a	0.29	0.32	0.22	0.15	0.16	0.11	0.11
	PF7	0.8%	2.85	0.39	0.53^a	0.33	0.2	0.23	0.23	0.21	0.19	0.18
	PF8	0.8%	2.92	0.30	0.61^a	0.36	0.22	0.17	0.17	0.22	0.17	0.10
	PF9	0.9%	2.98	0.17	0.51^a	0.19	0.16	0.12	0.07	0.10	0.09	0.05
	PF10	0.8%	2.98	0.16	0.42^a	0.18	0.12	0.12	0.10	0.13	0.13	0.10
Role Physical (1–2)	RP1	0.9%	1.83	0.37	0.33	0.63^a	0.29	0.28	0.17	0.29	0.26	0.08
	RP2	1.0%	1.80	0.40	0.31	0.58^a	0.27	0.33	0.21	0.29	0.35	0.14
	RP3	1.0%	1.84	0.37	0.43	0.63^a	0.35	0.32	0.18	0.28	0.22	0.12
	RP4	1.0%	1.87	0.34	0.42	0.54^a	0.32	0.31	0.26	0.30	0.23	0.19
Bodily Pain (1–6)	BP1	0.2%	4.88	1.02	0.33	0.31	0.60^a	0.40	0.28	0.29	0.17	0.25
	BP2	1.1%	4.53	0.65	0.37	0.41	0.60^a	0.37	0.28	0.42	0.26	0.23
General Health (1–5)	GH1	0.0%	3.05	0.97	0.28	0.33	0.33	0.56^a	0.41	0.24	0.17	0.30
	GH2	2.7%	3.85	1.07	0.19	0.30	0.34	0.56^a	0.34	0.26	0.21	0.27
	GH3	2.5%	3.94	0.96	0.20	0.23	0.25	0.54^a	0.37	0.21	0.16	0.32
	GH4	2.6%	3.86	1.08	0.29	0.28	0.28	0.45^a	0.32	0.27	0.16	0.28
	GH5	2.5%	3.80	0.92	0.27	0.31	0.37	0.68^a	0.46	0.30	0.21	0.40
Vitality (1–6)	VT1	1.2%	4.34	1.09	0.18	0.23	0.21	0.42	0.58^a	0.34	0.26	0.52
	VT2	1.2%	4.14	1.09	0.20	0.15	0.23	0.44	0.63^a	0.33	0.25	0.59
	VT3	1.2%	4.47	0.90	0.23	0.24	0.24	0.33	0.46^a	0.32	0.26	0.44
	VT4	1.0%	4.11	0.90	0.16	0.17	0.27	0.35	0.53^a	0.35	0.23	0.53
Social Functioning (1–5)	SF1	0.1%	4.44	0.69	0.18	0.31	0.33	0.28	0.29	0.41^a	0.39	0.41
	SF2	2.5%	3.91	0.84	0.25	0.32	0.32	0.31	0.43	0.41^a	0.30	0.42
Role Emotional (1–2)	RE1	0.3%	1.74	0.44	0.17	0.31	0.18	0.20	0.27	0.37	0.61^a	0.32
	RE2	0.2%	1.71	0.45	0.17	0.32	0.19	0.21	0.26	0.33	0.64^a	0.29
	RE3	0.3%	1.69	0.46	0.11	0.22	0.19	0.20	0.29	0.30	0.50^a	0.32
Mental Health (1–6)	MH1	1.2%	4.07	1.10	0.11	0.12	0.21	0.28	0.34	0.26	0.22	0.40^a
	MH2	1.5%	4.69	0.92	0.15	0.12	0.17	0.29	0.52	0.38	0.29	0.61^a
	MH3	1.2%	4.31	1.02	0.14	0.12	0.24	0.34	0.54	0.37	0.32	0.56^a
	MH4	1.2%	4.53	0.91	0.11	0.12	0.19	0.31	0.54	0.41	0.28	0.63^a
	MH5	1.0%	4.50	1.07	0.11	0.13	0.17	0.34	0.60	0.37	0.27	0.61^a

^a denotes correlation between an item and its hypothesized scale

Principal Component Analysis (PCA)

The SF-36 was constructed to represent two major dimensions of health – physical and mental. To test this in the Taiwan data, two factors with eigenvalues greater than 1 were extracted, which accounted for 60% of the total variance in SF-36 scale scores. As shown in Table 4, correlations between SF-36 scales and the two components generally were consistent with hypotheses in Taiwan. The PF, RP, and BP scales had the highest correlations with the physical component ($r = 0.80, 0.80,$ and 0.64 respectively) and the lowest correlations with the mental component. In addition, the MH scale and SF scale had

moderate associations with the mental component ($r = 0.90$ and $0.61,$ respectively) as anticipated.

However, correlations between the two components and the three other SF-36 scales were less consistent with hypotheses. The RE scale did not show as strong an association with the mental component ($r = 0.54$) in Taiwan as observed in Western countries. The GH and VT scales, originally hypothesized to measure both physical and mental health components, appeared to represent more a mental health concept ($r = 0.56$ and 0.84 respectively) than a physical one ($r = 0.46$ and 0.16), particularly for the Vitality scale.

Table 3: Reliability Statistics and Correlations between SF-36 Scales

Scale	PF	RP	BP	GH	VT	SF	RE	MH
PF – Physical Functioning	(0.83)							
RP – Role-Physical	0.47	(0.79)						
BP – Bodily Pain	0.39	0.39	(0.70)					
GH – General Health	0.34	0.40	0.43	(0.78)				
VT – Vitality	0.25	0.26	0.31	0.51	(0.75)			
SF – Social Functioning	0.26	0.37	0.38	0.35	0.44	(0.57)		
RE – Role-Emotional	0.18	0.34	0.23	0.25	0.33	0.41	(0.75)	
MH – Mental Health	0.17	0.17	0.27	0.43	0.69	0.49	0.38	(0.78)

Note. Scale reliability was represented on the diagonal. Internal consistency reliability lower than recommended level of 0.70 were shown with underlined entries.

Table 4: Factor Loadings of the 8 SF-36 scales using rotated principal components in Taiwan, Japan, Hong Kong and the US

Scale	Expected	Rotated Principal Components ^a									
		Physical					Mental				
		Taiwan	Japan	HK	US	Taiwan	Japan	HK	US		
PF	●	0.80	0.75	0.78	0.85	○	0.09	0.17	-0.08	0.12	
RP	●	0.80	0.86	0.63	0.81	○	0.19	0.19	0.34	0.27	
BP	●	0.64	0.51	0.61	0.76	○	0.28	0.52	0.23	0.28	
GH	◐	0.46	0.37	0.70	0.69	◐	0.56	0.66	0.10	0.37	
VT	◐	0.16	0.21	0.48	0.47	◐	0.84	0.88	0.51	0.64	
SF	◐	0.38	0.45	0.20	0.42	●	0.61	0.60	0.74	0.67	
RE	○	0.30	0.69	0.05	0.17	●	0.54	0.34	0.74	0.78	
MH	○	0.02	0.13	0.10	0.17	●	0.90	0.89	0.77	0.87	

● Strong association ($r \geq .70$) ◐ Moderate to substantial association ($.30 < r < .70$) ○ Weak association ($r \leq .30$) ^a Correlations between each scale and rotated principal component. Source: Data for US is from Ware, J., Snow, K., Kosinski, M., & Gandek, B. (1993). *SF-36 Health Survey: Manual & Interpretation Guide*. MA: Boston: Nimrod Press. Data for Japan is from Fukuhara S, Ware JE, Kosinski M, Wada S, Gandek B (1998). Psychometric and Clinical Tests of Validity of the Japanese SF-36 Health Survey. *J Clin Epidemiol*, 51, 1998; 51: 1045–1053 Data for HK is from Lam, C. L., Gandek, B., Ren, X. S., & Chan, M. S. (1998). Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey. *J Clin Epidemiol*, 51, 1139–1147.

Cross-cultural comparisons of the principal component analysis for Taiwan, Japan, Hong Kong, and the United States are also presented in Table 4. Consistent with results in previous studies, the pattern of correlations was similar across countries for the PF, RP, SF, and MH scales. For the BP scale, Taiwanese results appear to be similar to those from the US, but the BP scale is a little less physical. As evidenced by the correlation patterns, the GH and VT scales in Taiwan and Japan represent more of a mental concept than a physical one, which is different from US results. The RE scale represents more a mental concept than a physical one in Taiwan, as opposed to the Japanese study, but the association is not as strong as that observed in US.

Validity

Construct validity was examined by comparing SF-36 scores to the Stress Coping Inventory, which was aimed to assess a mental health construct. As expected, the results (Table 5) suggest that SF-36 score profiles correlate with those of the Stress Coping Inventory in expected ways. All subscales of the SF-36 mental health dimension except role-emotional, correlated more highly with related constructs measured by the Stress Coping Inventory than did those of the SF-36 physical health dimension. The results of the correlation analysis suggested an acceptable level of convergent validity for the SF-36 Taiwan version.

Results of criterion-based validity conformed to the original hypothesis, i.e. subscales of the physical health dimension declined with increasing age, while those in the mental health dimension were less influenced by age.

Table 5: Correlations between the SF-36 Health Survey and the Stress Coping Inventory (N = 569)

Stress Coping Inventory SF-36	Resources			Psychological Symptoms	
	Self-Esteem	Family Support	Friendship Support	Anxiety	Depression
PF – Physical Functioning	0.29	0.11	0.22	<u>-0.31</u>	-0.30
RP – Role-Physical	0.18	0.13	0.20	-0.29	-0.21
BP – Bodily Pain	0.15	0.14	0.17	-0.28	-0.26
GH – General Health	<u>0.38</u>	0.23	0.29	<u>-0.39</u>	<u>-0.38</u>
VT – Vitality	<u>0.53</u>	0.26	<u>0.39</u>	<u>-0.49</u>	<u>-0.53</u>
SF – Social Functioning	<u>0.32</u>	0.20	0.27	<u>-0.45</u>	<u>-0.36</u>
RE – Role-Emotional	0.18	0.08	0.15	<u>-0.41</u>	<u>-0.34</u>
MH – Mental Health	<u>0.53</u>	0.22	<u>0.42</u>	<u>-0.6</u>	<u>-0.57</u>

Note: All correlation coefficients in this table are statistically significant ($p < 0.05$). Underlined entries represent Moderate to Substantial associations ($.30 < r < .70$)

As shown in Figure 1, means of subscales in the physical health dimension (e.g. physical function and role physical) decreased with increasing age, while those in the mental health dimension (e.g. mental health or vitality) fluctuated less with age.

Discussion

One important issue in the current study is to determine whether the SF-36 measurement model can be applied in Taiwan. In general, the findings of the current study provide evidence that the concepts embodied in the SF-36 can be conveyed to the Taiwanese people and are feasible to be applied in Taiwan. Most tests of the psychometric properties of the SF-36 Taiwan version were satisfactory according to criteria set by the IQOLA project protocol, suggesting the feasibility of the translated version of the SF-36 for use in Taiwan. Data quality was high across the three study samples. The percentage of missing data ranged from less than 0% to 2.7% at an item level. These rates compare favorably with those reported in the original Medical Outcomes Study in the US [18] and other Western countries [20]. Results of the multitrait scaling analysis basically supported the hypothesized scale structure of the SF-36 in Taiwan and indicated that standard scoring algorithms could be used to score the eight SF-36 scales. The ordering of item means within scales generally were clustered within scales as hypothesized, with two exceptions involving the "felt tired" (VT4) and "felt nervous" (MH1) item. Similar results of lower mean score than expected for the VT4 item were also found in several other countries[19]. Due to the uneven distribution of subjects in terms of age range in the present study, the results of lower means in the VT4 and MH1 item should be interpreted with caution. However, psychometric testing results of the present study also indicate specific areas of the Taiwanese SF-36 in which further refinement and work will be required. Of the SF-36 eight scales, internal consistency reliability was generally acceptable for group-

level comparisons except for the Social Functioning scale. In the Taiwan version, the two SF items were correlated more highly with the MH scale than with their hypothesized scale. In addition, the SF scale had the lowest scaling success rate (87.5%). The suboptimal scaling performance of the SF scale has been observed and reported in a cross-cultural content comparison of SF-36 translations [21]. Due to cultural differences in the concept of social functioning, these items have been reported to be difficult to translate in some other countries [19]. The finding seems to suggest cultural differences in item interpretation. The concept of social functioning may be more westernized and less clear for Taiwanese people. Deeply ingrained in the Confucian ideology of collectivism, it is culturally unacceptable for people in Taiwan to use health problems as an excuse to avoid family or social gatherings[8]. That is, the denial of disturbance of physical and emotional health on social activities is more salient among Asians than Americans. Therefore, a specific family functioning scale may need to be added to generic health status questionnaires used in Taiwan, to acknowledge the impact of health on family life in cultures in which family life may play a more central role in people's lives, one that is distinct from the roles friends and other contacts hold[21].

Differences in the SF-36 profile between young and old adults shows evidence of discriminant validity for the SF-36 Taiwan version. In the student sample, acceptable convergent validity was obtained by the results of the comparison of the SF-36 and the Stress Coping Inventory, in which higher correlation coefficients were obtained for scales measuring similar psychological constructs.

In many countries, the SF-36 has been shown to yield reliable scale scores measuring eight dimensions of health status, which have two underlying measures of physical and mental health [22]. In Taiwan, the results of a principal component analysis lend support to the two compo-

ment models as hypothesized. However, some disparity in the pattern of correlations relative to United States was found. While primarily a physical scale, the Bodily Pain scale does not have as strong an association with the physical dimension as was found in the US. Although primarily a mental scale, the Role Emotional scale was not as purely associated with the mental component as in the US and Western Europe. However, the Role Emotional scale did have a higher loading on the mental factor than the physical factor in Taiwan. In contrast, the Role Emotional scale had a high loading on the physical factor and a low loading on the mental factor in Japan (although this finding did not hold for highly educated Japanese women) [17]. Differences in the factor structure of the Role Emotional scale between Asian and Western countries may reflect a reluctance to attribute limitations in role functioning to emotional states, particularly for the elderly who are less influenced by Western culture than the younger generation[23].

The VT and MH scales were highly intercorrelated, however, VT and MH items did have higher correlations with their hypothesized scales than all other scales. The Vitality scale had a high correlation with the mental component and a low correlation with the physical component in both Taiwan and Japan; these results contrasted with those from US, in which Vitality had a moderate to substantial association with both the mental and physical components. In addition, the Vitality scale in Taiwan was strongly correlated with scales that are related to the construct of mental health in the Stress Coping Inventory. Therefore, the Vitality scale appears to be less valid for measuring physical health in Taiwan than in the West. Such a relationship between Vitality and the Mental Health dimension has also been seen in other studies of Chinese-Americans[8,24] Subjects in the abovementioned samples are all familiar with the culture of traditional Chinese medicine; hence the results may reflect how people perceive their vitality status within a cultural framework. In the traditional theory of Chinese medicine, the phrase ("JingShen") associated with the presence of vitality is used to describe "mental well-being" [25]. It is therefore not surprising for Ren et al[8] to conclude that "vitality is central to the concept of a healthy mental state for Chinese."

Some limitations of the current study should be kept in mind. The reliability of the tool is not fully established in the current study, future assessment of reproducibility and responsiveness would be necessary for this. In addition, the construct validity of the results reported in this article is mainly derived from the student sample. Further research is necessary to replicate the results in different age groups to validate the SF-36 Taiwan version.

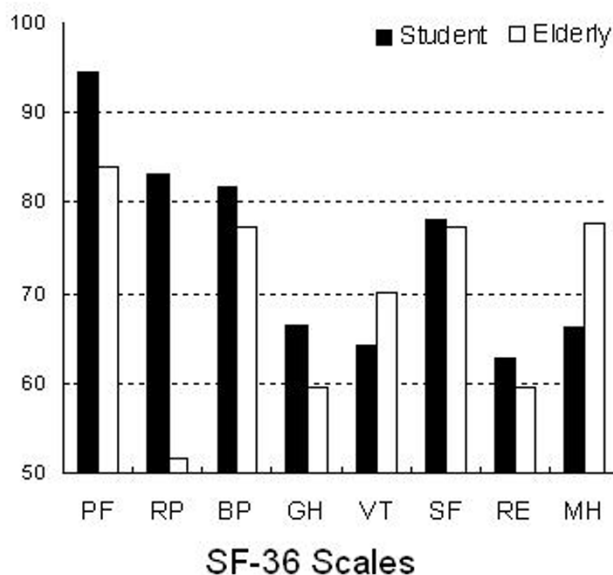


Figure 1
Comparison of SF-36 Scale Profiles between the Student and Elderly Sample.

Conclusions

In conclusion, we have provided empirical data to illustrate the feasibility of translating and validating the SF-36 in an Asian country. The Taiwan version of the SF-36 Health Survey appears to be a practical and reliable instrument in the general population. The finding that the Vitality scale is strongly associated with the mental health component is interpreted in a cultural framework. Previous studies have also found this pattern and raise important questions regarding cultural influences upon illness attribution and perception. Further research into conceptualization of the Vitality and/or Role Emotional scale among Asian countries within a cultural framework is warranted.

Authors' contributions

HMT carried out and participated in the design of the studies and performed the statistical analyses, as well as drafted the manuscript. JRL carried out and finalized the translated Taiwan version of SF-36, and drafted the manuscript. BG participated in the translation process and in the manuscript preparation. All authors read and approved the final manuscript.

Acknowledgements

This research was supported by the National Science Council of Taiwan (Grant No. NSC-92-2416-H-182-012). The following members who contributed to the translation, focus group discussions and pilot testing process in the development of SF-36 Taiwan version, Chung-Fu Lan, Shwu-Chong

Wu, Wen-Liang Liu, Jwo-Leun Lee, Chun-Huei Chi, Yung-Ho Ko, Li-Jen Weng, Shu-Jen Shiau, Kai-Kuen Leung, Sue-Ing Chang, and Grace Kuo are greatly appreciated. The authors are also thankful to H. Irene Wang and Pei-Hsin Lin for providing the datasets.

References

1. Ellwood PM: **Outcome management: A technology of patient experience [Shattuck lecture].** *N Engl J Med* 1988, **318**:1549-1556.
2. Bullinger M: **German translation and psychometric testing of the SF-36 Health Survey: preliminary results from the IQOLA Project.** *Social Science & Medicine* 1995, **41**:1359-1366.
3. Cutler DM, McClellan M, Newhouse JP: **What has increased medical-care spending.** *American Economic Review* 1998, **88**:132-136.
4. Awad AG, Voruganti LN: **Intervention research in psychosis: issues related to the assessment of quality of life.** *Schizophrenia Bulletin* 2000, **26**:557-564.
5. Goh CR, Lee KS, Tan TC, Wang TL, Tan CH, Wong J, Ang PT, Chan ME, Clinch J, Olweny CL, Schipper H: **Measuring quality of life in different cultures: translation of the Functional Living Index for Cancer (FLIC) into Chinese and Malay in Singapore.** *Annals of the Academy of Medicine, Singapore* 1996, **25**:323-334.
6. Ware J. E., Jr., Gandek B: **Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project.** *Journal of Clinical Epidemiology* 1998, **51**:903-912.
7. Bullinger M, Alonso J, Apolone G, Leplege A, Sullivan M, Wood-Dauphinee S, Gandek B, Wagner A, Aaronson N, Bech P, Fukuhara S, Kaasa S, Ware J. E., Jr.: **Translating health status questionnaires and evaluating their quality: the IQOLA Project approach.** *Journal of Clinical Epidemiology* 1998, **51**:913-923.
8. Ren XS, Amick B., 3rd, Zhou L, Gandek B: **Translation and psychometric evaluation of a Chinese version of the SF-36 Health Survey in the United States.** *Journal of Clinical Epidemiology* 1998, **51**:1129-1138.
9. Lam CL, Gandek B, Ren XS, Chan MS: **Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey.** *Journal of Clinical Epidemiology* 1998, **51**:1139-1147.
10. Steward AL, Ware JE: **Measuring Functioning and Well-Being: The Medical Outcomes Study Approach.** Durham, NC, Duke University Press; 1992.
11. Mchorney CA, Ware JE, Rogers W, Raczek AE, Lu JFR: **The validity and relative precision of MOS short- and long-form health status scales and Dartmouth COOP charts.** *Medical Care* 1992, **30**:Supplement MS253-65.
12. McHorney CA, Kosinski M, Ware J. E., Jr.: **Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey.** *Medical Care* 1994, **32**:551-567.
13. Chen Ching-Yu, Wu En-Chang: **A biopsychosocial model to assess the psychosomatic symptoms in college students of National Taiwan University.** *Chinese Journal of Mental Health* 1987, **3**:89-105.
14. Ware JE, Harris WJ, Gandek B, Rogers BW: **MAP-R for Windows: Multitrait-Multi-Item Analysis Program - Revised User's Guide.** MA: Boston, Health Assessment Lab; 1997.
15. Gandek B, Ware J. E., Jr.: **Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach.** *Journal of Clinical Epidemiology* 1998, **51**:953-959.
16. Nunally JC, Bernstein IR: **Psychometric theory.** NY, McGraw-Hill; 1994.
17. Fukuhara S, Ware J. E., Jr., Kosinski M, Wada S, Gandek B: **Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey.** *Journal of Clinical Epidemiology* 1998, **51**:1045-1053.
18. McHorney CA, Ware J. E., Jr., Lu JF, Sherbourne CD: **The MOS 36-item Short-Form Health Survey (SF-36): III.** *Medical Care* 1994, **32**:40-66.
19. Gandek B, Ware J. E., Jr., Aaronson NK, Alonso J, Apolone G, Bjorner J, Brazier J, Bullinger M, Fukuhara S, Kaasa S, Leplege A, Sullivan M: **Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project.** *Journal of Clinical Epidemiology* 1998, **51**:1149-1158.
20. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, Sprangers MA, te Velde A, Verrrips E: **Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations.** *Journal of Clinical Epidemiology* 1998, **51**:1055-1068.
21. Wagner AK, Gandek B, Aaronson NK, Acquadro C, Alonso J, Apolone G, Bullinger M, Bjorner J, Fukuhara S, Kaasa S, Leplege A, Sullivan M, Wood-Dauphinee S, Ware J. E., Jr.: **Cross-cultural comparisons of the content of SF-36 translations across 10 countries: results from the IQOLA Project.** *Journal of Clinical Epidemiology* 1998, **51**:925-932.
22. Keller SD, Ware J. E., Jr., Bentler PM, Aaronson NK, Alonso J, Apolone G, Bjorner JB, Brazier J, Bullinger M, Kaasa S, Leplege A, Sullivan M, Gandek B: **Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: results from the IQOLA Project.** *Journal of Clinical Epidemiology* 1998, **51**:1179-1188.
23. Fukuhara S, Ware JE, Kosinski Jr M, Wada S, Gandek B: **Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey.** *Medical Decision Making* 1996, **16**:169-177.
24. Chen AY, Daley J, Thibault GE: **Angina patients' ratings of current health and health without angina: associations with severity of angina and comorbidity.** *Medical Decision Making* 1996, **16**:169-177.
25. Lin KM: **Traditional Chinese medical beliefs and their relevance for mental illness and psychiatry.** *Normal and abnormal behavior in Chinese culture* Edited by: Kleinman A and Ln TY. Boston, MA, Kluwer Boston; 1980:95.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

