

# Validation of the Chinese Version of the SF-36 Health Survey Questionnaire in People Undergoing Physical Examinations

Tsai-Chung Li<sup>1,2</sup>, Chiu-Shong Liu<sup>3</sup>, Cheng-Chieh Lin<sup>3</sup>, Yih-Dar Lee<sup>4,6</sup>,

Jim-Shoung Lai<sup>5</sup>, Chia-Ing Li<sup>2</sup>

<sup>1</sup>Institute of Chinese Medicine, <sup>5</sup>Department of Occupational Safety and Health, China Medical University;

<sup>2</sup>Department of Medical Research, <sup>3</sup>Family Medicine, China Medical University Hospital, Taichung; <sup>4</sup>School of Medicine, National Cheng Kung University, Tainan; <sup>6</sup>Lilly Taiwan, Eli Lilly and Company, Taipei, Taiwan.

**Objectives.** The SF-36 is a generic measure of health status and has gained popularity as a measure of outcome in a wide variety of patient groups and surveys. The Chinese version of SF-36 was introduced in 1995 and its reliability and validity in the general population has been reported. However, the reliability and validity has never been reported in clinical settings. Therefore, we provide estimates of the reliability and test validity of the Chinese SF-36 in people undergoing physical examinations.

**Methods.** A cross-sectional study was conducted among individuals who underwent physical examinations in the China Medical University Hospital in 1996. A total of 434 individuals over the age of 18 years were recruited. The overall completion rate was 68.7%. Outcome was evaluated by the Short Form 36 (SF-36), a short questionnaire with 36 items measuring eight multi-item variables: physical functioning, social functioning, role limitation due to physical problems, role limitation due to emotional problems, mental health, vitality, pain and general health.

**Results.** All variables passed tests for item-internal consistency and item-discriminant validity. Analyses of internal consistency coefficients ranged from a low of 0.66 for the bodily pain variable to a high of 0.89 for social and physical functioning variables. Validation by factor analysis yielded results remarkably similar to those proposed by the authors who developed the SF-36. We found significant differences in physical-related variables such as physical functioning, and bodily pain between different age groups, and significant differences in all variables except for bodily pain in different categories of minor psychiatric morbidity.

**Conclusions.** In general, the Chinese version of the SF-36 is reliable and valid for measuring health status in people undergoing physical examination. ( *Mid Taiwan J Med* 2005;10:8-17 )

## Key words

health examination, SF-36 health survey, validity

## INTRODUCTION

Because of the increasing emphasis on the importance of subjective accounts of health in monitoring medical outcomes, tremendous effort

Received : 31 August 2004.

Revised : 18 November 2004.

Accepted : 15 December 2004.

Address reprint requests to : Tsai-Chung Li, Institute of Chinese Medicine, China Medical University, 91 Hsueh-Shih Road, Taichung 404, Taiwan.

has been made in the past decade to develop health questionnaires that measure the perception of health of the population [1,2]. These questionnaires measure the potential benefits of health care intervention that can influence a wide number of variables such as physical mobility and functioning, mental health, social life, and overall well-being [3]. The SF-36 is a generic

measurement of health status and has gained popularity as a means of evaluating outcome in a wide variety of patient groups and surveys, especially primary care practice in Taiwan.

The Chinese-version of the SF-36 was introduced in Taiwan in 1995. Studies have indicated that inadequate language translation may have led to a reduction in the validity of its content [4,5]; therefore, it is necessary to validate the Chinese version of the SF-36 in different settings before it is widely used. Although psychometric testing of the Chinese version SF-36 in the general population has been validated [6-8], this kind of study has never been performed in a primary care setting. If the instrument is validated in a primary setting and it performs as we would expect, we will have confidence in its validity in primary care settings. Therefore, the specific aim of this study was to test the reliability and validity of a Chinese-Language version of the MOS 36-item short form health survey (SF-36) for measuring the health status in a sample of individuals who underwent physical examinations in a medical center in Taiwan.

## SUBJECTS AND METHODS

### Study subjects

Six hundred and thirty-two consecutive individuals who underwent physical examinations in the China Medical University Hospital in Taichung were recruited. All individuals were administered questionnaires, which included a Chinese version of SF-36, Chinese Health Questionnaire, questions about sociodemographic factors, life events, and medical history. All individuals who wanted to complete the self-rating questionnaires were included in the study while those who had cognitive problems were excluded. Among the 632 consecutive individuals, 434 completed the questionnaires. Of those, 200 (46.1%) were older than 65 years, 246 (56.7%) were male, and 214 (49.3%) had more than 12 years of education. The overall completion rate was 68.7%.

### Measurement

*Sociodemographic factors.* Age, gender,

and years of education were collected in the questionnaire.

*SF-36.* The SF-36 is a short questionnaire with 36 items measuring eight multi-item scales: physical functioning (10 items), social functioning (2 items), role limitations due to physical problems, herein abbreviated as role-physical (4 items), role limitations due to emotional problems, herein abbreviated as role-emotional (3 items), mental health (5 items), energy and vitality (4 items), pain (2 items), and general perception of health (5 items). For each scale, item scores are coded, summed, and transformed from 0 (worst possible health state measured by the questionnaire) to 100 (best possible health state). For the SF-36, high scores indicate better perceived health status. The details of the translation process for SF-36 have been reported by Lu et al [7].

*Life events.* This variable was measured by a self-administered questionnaire that consisted of 60 items grouped into 10 problem domains covering housing, work, financial status, legal matters, family status, child-parent interaction, and marital relationship. For each of the 10 domains, the presence of social problems in the past month was determined and the total score was then computed by adding up the number of domains for which social problems were identified. This variable potentially measured the stress in an individual's life.

*Minor psychiatric morbidity (MPM).* For discriminative instruments, construct validity is established by examining the relationship between scores on the instrument and other indices at a single point in time. Chong found that 38.7% of individuals attending a health screening had psychological morbidity [9] which was much higher than in the general population [10]. Therefore, we chose psychological morbidity as one of the indices. MPM was measured by the Chinese Health Questionnaire (CHQ-30). CHQ was administered as a screening test of minor psychiatric morbidity in the community and primary care settings for Chinese. It consisted of 30 items rated on a four-point Likert-type scale ranging from 0 (not at all and the same as usual)

to 1 (more often than usual and much more often than usual). Previous studies have shown that the CHQ is highly reliable (Cronbach's alpha coefficients in community samples and hospital groups were 0.9 and 0.92, respectively) and valid (the sensitivity and specificity were 76% and 77%, respectively) [11,12]. The Receiver Operator Characteristic curve showed that the CHQ-30 was most valid for measuring minor psychiatric morbidity at a cutoff point of 9/10 [12]. The CHQ-30 consisted of questions related to anxiety, depression, insomnia, fatigue, poor concentration and memory, physical health, and family relationship.

**Chronic disease.** Participants were classified as having a chronic medical condition if they had hypertension, diabetes mellitus, heart disease, anemia, incontinence of urine, duodenal ulcer, chronic hepatitis B, hepatitis C, or tuberculosis.

### Statistical analysis

**Scaling assumptions.** Scales should satisfy the assumption that responses to items in each of the scales could be summed without standardization or taking weights. We evaluated this assumption in two ways: equivalence of standard deviations and item-scale correlations. If items of a given scale did not have equal variance, the computation of the total score required standardization of items prior to summation. Item-scale correlation reveals the extent to which each item measures what the scale is intended to measure. If each item contributes roughly equal to the underlying concept, equal weight can be applied to all items in scale scoring. If item-scale correlations are not equal for all items in a given scale, items should be weighted before the scale is computed.

We evaluated two aspects of validity pertaining to item-scale correlations: convergent and discriminant validity. When all items measure the same underlying concept, convergent validity is held. To achieve satisfactory validity with very short scales, each item must correlate substantially with the scale it represents. The correlation must also be corrected for overlap,

which is done by correlating an item with the sum of the other items in the same scale to remove the bias of correlating an item with itself. A correlation  $\geq 0.4$ , after correction for overlap, is considered to be substantial [13]. The overall success rate of convergent validity for a given scale is equal to the number of scaling successes divided by the total number of scaling tests.

Discriminant validity is defined as an item that correlates higher with its own scale (corrected for overlap) than with other scales [13]. We tested discriminant validity by comparing correlation of each item with its own scale with the correlation of each item with the other scales. A success was counted whenever an item correlated significantly higher (two standard errors or more) with its hypothesized scale than with other SF-36 scales [13].

**Reliability.** The reliability of internal consistency was assessed in this study by Cronbach's alpha coefficient, which measures the overall correlation between items within a scale. Cronbach's alpha should exceed 0.7 to be considered acceptable for group comparison [14].

**Validity.** Factor analysis [15], a technique of psychometric validation that assesses the agreement between hypothetical factors which go to make up the measure and the scales designed to assess those factors, was used to test the dimensions of SF-36 by extracting principal components from the correlations among their items. Each principal component was a linear combination of 35 items of the SF-36. The extracted components were orthogonal to each other. The components were rotated using the varimax method. If the Chinese version of the SF-36 is a valid measure, the items of the same scale defined by the authors of SF-36 should load on a given factor in this primary care sample, i.e., within such an assessment, a factor should be considered relevant only if its eigenvalue (a statistical measure of its power to explain variation between subjects) exceeds 1 [16].

We also assessed evidence of construct validity for the Chinese version of the SF-36 by following the logic proposed by Carmines and Zeller [17]: construct validity is the extent to

**Table 1. Hypothesized differences in SF-36 variable scores among groups of people who vary in sociodemographics and self-reported morbidity**

Physical component*	Hypothesized differences <sup>†</sup>								
	●	●	●	◐	◐	◐	●	●	
Mental component*	○	○	○	◐	◐	●	○	○	
Characteristics	PF	RP	BP	GH	VT	SF	RE	MH	
Age	Younger people have higher mean scores than older people			Younger people have higher mean scores than older people			No strong difference		
Gender	Men have higher mean scores than women			Men have higher mean scores than women			Men have higher mean scores than women		
Level of education	People with higher education have higher mean scores than those with less education			People with higher education have higher mean scores than those with less education			People with higher education have higher mean scores than those with less education		
Number of life events	People with less life events have higher mean scores than those with more life events; differences greater for mental health variables than for physical health variables			People with less life events have higher mean scores than those with more life events			People with less life events have higher mean scores than those with more life events; differences greater for mental health variables than for physical health variables		
Chronic disease	People having chronic condition(s) have lower mean scores than healthy people; differences greater for physical health variables than for mental health variables			People having chronic condition(s) have lower mean scores than healthy people			People having chronic condition(s) have lower mean scores than healthy people; differences greater for physical health variables than for mental health variables		
Minor psychiatric morbidity	People having minor psychiatric morbidity have lower mean scores than healthy people; differences greater for mental health variables than for physical health variables			People having minor psychiatric morbidity have lower mean scores than healthy people			People having minor psychiatric morbidity have lower mean scores than healthy people; differences greater for mental health variables than for physical health variables		

\*The associations of SF-36 variables with physical and mental components were based upon Ware et al [13]. <sup>†</sup>Hypothesis based on Ware et al [13]. ●: strong association ( $r \geq 0.7$ ). ◐: moderate to substantial association ( $0.3 < r < 0.70$ ). ○: weak association ( $r \leq 0.3$ ). PF = physical functioning; RP = role physical; BP = bodily pain; GH = general health; VT = vitality; SF = social functioning; RE = role emotional; MH = mental health.

which a particular measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts (or constructs) that are being measured. We hypothesized differences in SF-36 variable scores between groups who varied in age, gender, level of education, number of life events, chronic disease and minor psychiatric morbidity (shown in Table 1). The theoretical relationship between aging and health status was used as a guide to assess the construct validity of the Chinese version of the SF-36. Thus, we hypothesized that older individuals would have lower SF-36 scores than younger individuals. We also hypothesized that scores would vary in a predictable manner between individuals having different numbers of life events, minor psychiatric morbidity, and chronic condition. If the scores varied in a predictable manner, it would provide evidence that the SF-36 is a discriminative instrument.

Student's *t* test compared the means of 8

variables of the SF-36 when the independent variables had 2 categories while analysis of variance (ANOVA) compared the means between more than 2 groups. Multiple linear regression model tested the independent effect of a particular independent variable on 8 variables of SF-36 by controlling for the other independent variables in the model.

## RESULTS

Table 2 shows missing data, Cronbach's alpha coefficients, and results of the scale assumption test: item-discriminant validity and item-convergent validity. Overall, the percentage of missing data was less than 10% in all variables. Cronbach's alpha coefficients ranged from 0.66 to 0.89. Minimum standards of reliability for purposes of group comparisons ( $\geq 0.7$ ) were satisfied for all SF-36 variables in this outpatient sample except for bodily pain. Perfect success rates were achieved across 8 SF-36 variables for

**Table 2. Scaling properties of the 8 variables of the SF-36 in a sample of people undergoing health examinations in a medical center**

Variable	Items' missing rate (%, range)	Cronbach's alpha	Items' standard deviation (range)	Correlations of items with own scale (range)	Correlations of items with other scales (range)	Item-convergent validity*	Item-discriminant validity <sup>†</sup>
Physical functioning	6.17 – 8.13	0.89	0.33 – 0.71	0.43 – 0.80	0.07 – 0.49	10/10 (100%)	80/80 (100%)
Role-physical	9.40 – 10.80	0.85	0.48 – 0.50	0.65 – 0.88	0.31 – 0.63	4/4 (100%)	32/32 (100%)
Bodily pain	6.59 – 6.73	0.66	1.31 – 1.32	0.81	0.36 – 0.57	2/2 (100%)	16/16 (100%)
General perception of health	3.23 – 10.52	0.83	0.99 – 1.31	0.56 – 0.85	0.29 – 0.58	5/5 (100%)	40/40 (100%)
Vitality	7.15 – 8.98	0.83	1.20 – 1.26	0.61 – 0.90	0.25 – 0.61	4/4 (100%)	29/32 (90.6%)
Social functioning	5.61 – 7.15	0.89	0.93 – 1.08	0.51	0.27 – 0.56	2/2 (100%)	11/16 (68.8%)
Role-emotional	9.96 – 10.24	0.85	0.49	0.67 – 0.92	0.28 – 0.60	3/3 (100%)	24/24 (100%)
Mental health	7.29 – 8.70	0.83	1.19 – 1.34	0.51 – 0.83	0.15 – 0.64	5/5 (100%)	37/40 (92.5%)

\*Correlation between an item with its own scale  $\geq 0.40$ . <sup>†</sup>Correlation between an item with its own scale was significantly greater than the item with other scales.

**Table 3. The factorial structure and factor loadings of the SF-36 questionnaire in people undergoing health examinations in a medical center**

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Name of factor	Physical functioning I	Role-emotional and role-physical	Mental health and vitality I	General perception of health	Mental health and vitality II	Bodily pain	Physical functioning II
	PF2 (0.78)	RE2 (0.80)	MH3 (0.82)	GH5 (0.71)	VT1 (0.73)	BP1 (0.72)	PF10 (0.82)
	PF1 (0.73)	RE1 (0.76)	VT3 (0.75)	GH2 (0.65)	VT2 (0.69)	BP2 (0.71)	PF9 (0.71)
	PF4 (0.72)	RE3 (0.73)	VT4 (0.71)	GH3 (0.65)	MH5 (0.61)		PF5 (0.57)
	PF7 (0.70)	RP2 (0.64)	MH2 (0.71)	GH4 (0.63)	MH3 (0.43)		
	PF8 (0.68)	RP1 (0.63)	MH1 (0.64)	GH1 (0.61)			
	PF6 (0.59)	RP3 (0.46)	SF2 (0.41)				
	PF3 (0.88)	RP4 (0.42)					
Eigenvalue	12.71	3.44	1.87	1.59	1.19	1.11	1.04
Proportion of variance explained (%)	36.3	9.8	5.3	4.5	3.4	3.2	3.0

PF = physical functioning; RE = role emotional; MH = mental health; VT = vitality; SF = social functioning; GH = general perception of health; BP = bodily pain; RP = role-physical.

item-discriminant validity (column 7) and in 5 out of 8 for item-convergent validity (column 8). In 267 comparisons out of 280, the correlation between an item and its hypothesized variable exceeded correlations with all others variables by more than 2 standard errors. In addition, all items satisfied the criterion set a priori for convergent validity, i.e., a correlation between an item with its own variable  $\geq 0.4$ . Thus, the success rate for discriminant validity was 95.4%, and for convergent validity, 100.0%.

### Validation by factor analysis

Table 3 shows the results of factor analysis with items having coefficients greater than 0.4. Factor analysis identified 8 relevant factors, with

eigenvalues ranging from 1.0 to 12.7 and with proportions of total variance ranging from 3.0% to 36.3%. The physical functioning variable was separated into 2 factors (factors 1 and 7). Factor 2 was formed by items of role-physical and role-emotional variables. Items of mental health and vitality variables were combined and then separated into two factors (factors 3 and 5). One item of social functioning was also combined with factor 3. The other 2 factors corresponded exactly to 2 variables of the SF-36: general health perception and bodily pain.

### Construct validation

Lower scores on the SF-36 reflect poorer health status. Table 4 shows means and standard

**Table 4. Means (standard deviations) of the SF-36 scores for specific subgroups in people undergoing health examinations in a medical center**

Scale	Physical functioning	Role-physical	Bodily pain	General health perception	Vitality	Social functioning	Role-emotional	Mental health
Age (yr)								
18 – 34	87.9 (11.9)***	61.8 (37.5)***	63.1 (25.6)*	44.8 (22.8)**	47.2 (19.8)	72.6 (21.9)	60.8 (41.5)	55.6 (19.2)
35 – 49	85.5 (15.1)	68.1 (38.1)	67.3 (22.8)	49.9 (22.0)	51.3 (19.2)	77.1 (19.3)	67.1 (41.7)	59.3 (18.1)
50 – 64	72.8 (24.4)	53.4 (42.6)	60.5 (25.2)	41.4 (20.3)	46.3 (20.3)	70.7 (22.7)	60.6 (42.7)	57.1 (21.6)
≥ 64	61.0 (23.3)	37.2 (40.2)	55.2 (27.4)	39.4 (21.7)	45.5 (20.0)	69.6 (23.8)	47.5 (45.4)	58.5 (19.5)
Gender								
Male	82.4 (19.9)***	65.4 (38.7)***	66.0 (24.6)**	49.0 (21.4)***	50.7 (19.6)**	75.2 (20.3)*	66.8 (40.3)**	61.7 (18.4)***
Female	74.1 (22.2)	48.4 (41.8)	58.3 (25.0)	39.3 (21.3)	44.6 (19.8)	70.3 (23.2)	54.1 (44.7)	52.6 (20.1)
Education (yr)								
≤ 9	68.2 (25.0)***	45.0 (41.4)***	53.7 (24.5)***	37.0 (19.6)***	43.1 (19.1)***	68.9 (22.8)**	54.6 (44.5)*	54.0 (19.2)**
9 – 12	80.1 (18.5)	60.3 (38.6)	64.4 (21.3)	49.7 (17.2)	51.5 (17.7)	74.7 (18.0)	63.0 (39.3)	61.1 (17.8)
> 12	86.3 (14.6)	67.5 (38.4)	69.3 (24.2)	49.6 (22.8)	51.1 (20.2)	75.8 (21.3)	65.9 (41.6)	59.8 (20.1)
Number of life events								
≤ 1	79.6 (20.1)	59.5 (42.0)	62.4 (27.0)**	45.2 (22.9)	49.5 (21.3)*	76.3 (21.2)***	65.8 (43.6)***	60.0 (19.9)*
2 – 5	79.1 (21.9)	60.7 (39.7)	65.8 (23.5)	46.3 (21.2)	49.0 (18.5)	74.3 (20.9)	65.5 (40.1)	58.1 (19.5)
≥ 6	75.5 (22.1)	47.8 (40.7)	54.3 (23.2)	39.3 (20.8)	42.3 (19.7)	62.9 (22.5)	39.8 (41.9)	51.5 (18.7)
Minor psychiatric morbidity								
No	83.9 (17.6)***	71.5 (37.5)***	71.0 (23.8)***	53.3 (20.7)***	56.4 (17.8)***	79.6 (16.9)***	77.4 (35.4)***	65.8 (16.5)***
Yes	71.1 (23.8)	38.6 (37.6)	51.0 (22.0)	32.7 (17.4)	36.5 (16.4)	63.9 (24.3)	38.7 (41.8)	46.5 (18.3)
Having chronic disease								
No	84.0 (17.6)***	69.1 (36.5)***	69.5 (23.3)***	52.2 (20.3)***	53.4 (18.5)***	78.0 (18.1)***	67.9 (40.9)**	61.2 (18.7)***
Yes	73.9 (23.2)	48.5 (42.1)	56.6 (24.9)	38.2 (21.2)	43.4 (19.8)	68.8 (23.7)	55.6 (43.4)	54.5 (20.0)

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

deviations, broken down by age, gender, education, life event, minor psychiatric morbidity and chronic disease. Overall, older individuals scored significantly lower on physical functioning, role-physical, bodily pain, and general health than younger individuals, indicating older individuals have poorer health on these domains ( $p < 0.001$  for physical functioning and role-physical;  $p < 0.05$  for bodily pain and  $p < 0.01$  for general health). Women scored significantly lower on all variables than men. Individuals with higher levels of education scored higher on all variables. Individuals with a higher number of life events scored lower on bodily pain, vitality, social functioning, role-emotional, and mental health. Individuals who had minor psychiatric morbidity scored lower in all variables than those without (all  $p < 0.001$ ). Individuals with chronic disease had significantly lower scores on all variables than those without (all  $p < 0.001$  except for role-emotional  $p < 0.01$ ).

Multiple linear regression was employed to examine the independent effects of the significant

predictive factors on 8 variables of SF-36 by adjusting for the confounding effects of the other variables in people receiving health examinations (Table 5). The percentages of the variation of each of the 8 variables explained by age, gender, level of education, life event, MPM, and chronic conditions ranged from 19.5% to 34.6%, for social functioning having the lowest percentage and for general perception of health having the highest percentage. The estimated effects of age were significantly negative on physical functioning and positive on bodily pain while the estimated effects of being female were all significantly negative on physical functioning, role-physical, general health, and mental health. The estimated effects of education were all significantly positive on physical functioning, role-physical, bodily pain, general health, vitality, and mental health. After controlling for the mental illness variable of MPM and other factors in the model, the significant effects of life event were only observed on social functioning and role-emotional. The estimated effects of MPM on

**Table 5. The estimated parameters ( $\beta$  (SE)) of sodiodemographic factors, chronic conditions, life events, and minor psychiatric morbidity in people undergoing health examinations in a medical center**

Variable	Physical functioning	Role-physical	Bodily pain	General health perception	Vitality	Social functioning	Role-emotional	Mental health
Intercept	93.3 (3.8)***	77.5 (7.6)***	61.4 (4.5)***	51.1 (3.8)***	54.6 (3.4)***	83.1 (4.2)***	85.1 (8.1)***	65.0 (3.5)
Age (yr)								
35 – 49	-2.6 (2.5)	4.3 (5.0)	5.0 (3.0)	4.5 (2.5)	3.0 (2.3)	3.3 (2.8)	2.2 (5.3)	2.2 (2.3)
50 – 64	-11.0 (3.0)***	-2.1 (5.9)	7.9 (3.5)*	3.0 (3.0)	3.1 (2.7)	0.1 (3.3)	-1.6 (6.3)	4.3 (2.7)
≥ 65	-21.2 (3.8)***	-14.3 (7.6)	6.4 (4.5)	4.7 (3.8)	4.8 (3.5)	0.5 (4.2)	-11.5 (8.2)	7.0 (3.5)*
Gender								
Female	-5.3 (1.9)**	-11.1 (3.7)**	-2.3 (2.2)	-4.5 (1.9)*	-1.8 (1.7)	-2.2 (2.1)	-7.2 (4.0)	-5.4 (1.7)**
Education (yr)								
9 – 12	4.1 (3.2)	5.4 (6.5)	9.8 (3.9)*	8.8 (3.3)**	7.1 (3.0)*	3.0 (3.6)	3.7 (6.9)	6.4 (3.0)*
> 12	6.5 (2.5)*	10.3 (5.0)*	15.0 (3.0)***	9.3 (2.5)***	6.4 (2.3)**	3.0 (2.8)	1.4 (5.4)	5.0 (2.3)*
Life event								
2 – 5	-3.4 (2.0)	-1.5 (3.9)	1.7 (2.4)	-0.2 (2.0)	-1.4 (1.8)	-2.9 (2.2)	-2.2 (4.2)	-2.4 (1.8)
≥ 6	-2.6 (2.7)	-4.4 (5.4)	-2.7 (3.2)	-0.6 (2.7)	-1.9 (2.5)	-9.8 (3.0)***	-17.7 (5.7)**	-3.4 (2.5)
Minor psychiatric morbidity	-10.7 (1.9)***	-27.7 (3.7)***	-16.7 (2.3)***	-17.8 (1.9)***	-18.1 (1.7)***	-12.9 (2.1)***	-33.9 (4.0)***	-17.2 (1.7)***
Having chronic disease	-4.4 (1.8)*	-12.6 (3.7)***	-9.0 (2.2)***	-10.4 (1.8)***	-7.2 (1.7)***	-6.8 (2.0)***	-4.9 (3.9)	-4.3 (1.7)*
$R^2$	31.6%	26.4%	26.9%	34.7%	31.1%	19.5%	24.3%	28.9%

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

all variables were all significantly negative. For chronic conditions, the estimated effects were all significantly negative on all variables except for role-emotional. By examining these effects of age across these 8 variables, age had the greatest impact on physical functioning; gender on role-physical; education on bodily pain; life event and MPM on role-emotional; and chronic conditions on role-physical.

## DISCUSSION

The SF-36 is a brief and easy-to-use questionnaire. Our study showed that the Chinese version of the SF-36 was reliable and valid and therefore appropriate for self-administration. The questionnaire took about 10 minutes to complete and had a high completion rate (96.77% to 89.20%). Therefore, the Chinese version of the SF-36 questionnaire appears to be an acceptable outcome measure in people receiving health examinations in a medical center.

Our findings supported the claims that the domains of the Chinese version of the SF-36 are internally consistent with the domains proposed by the authors of the SF-36 and also confirmed that its psychometric assumptions have remained intact. For example, success rates were high for convergent and discriminant validity.

Validation by factor analysis yielded results remarkably similar to those proposed by the authors who developed SF-36. Three main differences from the hypothetical construct were observed in our sample. First, the items of vitality closely correlated with those of mental health scale, which is similar to the results of Garratt et al [18]. The items of these 2 scales formed 2 factors in our study, but only 1 factor in the study by Garratt et al. Second, the items of physical functioning were separated into 2 factors in our study while only one factor in the study by Garratt et al. Third, the items of role-emotional and role-physical were closely correlated in our study, but the items of bodily pain formed an independent factor in the study by Garratt et al. However, in the study by Garratt et al, the items of role-physical problems, bodily pain and social functioning were clustered, but the items of role-emotional formed an independent factor. The items of bodily pain and general health precisely corresponded to their hypothetical scales in our sample. Such precise correspondence between factors and scales is rare in factor analysis and thus confirms the validity of the SF-36 in a primary care setting in Taiwan.

All estimates of internal consistency for the

SF-36 scales exceeded accepted standards for measures used for group comparisons. These results support the use of the SF-36 scales in studies of health status that are based on group-level analyses. All but one of the published coefficients exceed the minimum standard of 0.70 for individual comparison suggested by Nunnally [14].

Previous studies have confirmed empirically that the SF-36 was constructed to represent two major dimensions of health: physical and mental [19,20]. We observed significant effects of age on physical functioning and bodily pain, which were strongly associated with the physical component of health being hypothesized by McHorney et al [21]. Life event was observed to have a significant effect on social functioning, and role-emotional, which were strongly or moderately associated with the mental component of health. MPM and chronic conditions were significantly associated with physical and mental components of health. All scores varied in a manner consistent with the relationship proposed in the literature.

A number of limitations should be noted in interpreting the results of this study. The individuals who participated in this study were from a medical center in central Taiwan. They may not be representative of those undergoing physical check-up at other medical centers, different types of clinical settings, or different locations in Taiwan. In addition, SF-36 asks how respondents have been feeling during the past 4 weeks and therefore considers the status during this period. Those with low or high scores at the time of measurement may have been influenced by events before the measurement. This kind of measurement error might be random or differential. For example, if women were more sensitive to the effects of a life event, overestimation of SF-36 scores for females may have occurred.

In conclusion, results of psychometric tests provide initial support for the validity and reliability of the Chinese version of the SF-36 in a clinical setting. The significant differences between age, gender, life event, minor psychiatric

morbidity, and comorbidity status support the discriminatory ability of the instrument. A longitudinal study for responsiveness needs to be conducted.

#### ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation for the hard work of the research assistants. We are also indebted to all respondents for their kind hospitality and cooperation during the study period. The whole project was supported by a grant from the China Medical University Hospital, Taiwan, Republic of China for 1 year (project no DMR-87-022).

#### REFERENCES

1. Geigle R, Jones SB. Outcomes measurement: a report from the front. *Inquiry* 1990;27:7-13.
2. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
3. Brazier JE, Harper R, Jones NM, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992;305:160-4.
4. Berkanovic E. The effect of inadequate language translation on Hispanics' responses to health surveys. *Am J Public Health* 1980;70:1273-6.
5. Deyo RA. Pitfalls in measuring the health status of Mexican Americans: comparative validity of the English and Spanish Sickness Impact Profile. *Am J Public Health* 1984;74:569-73.
6. Li TC, Lin CC, Lee YD. Validation of the Chinese (Taiwan) Version of SF-36 Health Survey in a Random Sample of Central Taiwan Population. *Qual Life Res* 2000;9:1060.
7. Lu JR, Tseng HM, Tsai YJ. Assessment of health-related quality of life in Taiwan (I): development and psychometric testing of SF-36 Taiwan version. *Taiwan J Public Health* 2003;22:501-11.
8. Tseng HM, Lu JR, Tsai YJ. Assessment of health-related quality of life in Taiwan (II): norming and validation of SF-36 Taiwan version. *Taiwan J Public Health* 2003;22:512-8.
9. Chong MY. Psychiatric screening of patients admitted for general health examinations. *Chinese Psychiatry* 1990;4:259-67.
10. Cheng TA. A community study of minor psychiatric morbidity in Taiwan. *Psychol Med* 1988;18:953-68.
11. Cheng TA, Wu JT, Chong MY, et al. Internal



- consistency and factor structure of the Chinese Health Questionnaire. *Acta Psychiatr Scand* 1990;82:304-8.
12. Chong MY, Wilkinson G. Validation of 30- and 12-item versions of the Chinese Health Questionnaire (CHQ) in patients admitted for general health screening. *Psychol Med* 1989;19:495-505.
  13. Ware JE, Snow KK, Kosinski M, et al. SF-36 Health Survey. Manual and Interpretation Guide. New England Medical Center, Boston, 1993.
  14. Nunnally JC. Psychometric Theory. 2<sup>nd</sup> ed. New York: McGraw-Hill, 1978.
  15. Child D. The Essentials of Factor Analysis. 2<sup>nd</sup> ed. London: Cassell, 1990.
  16. Manly BF. Multivariate Statistical Methods: A Primer. 2<sup>nd</sup> ed. Chapman & Hall, 1994.
  17. Carmines EG, Zeller RA. Reliability and Validity Assessment. Beverly Hills, Sage, 1979.
  18. Garratt A, Ruta D, Abdalla MI, et al. The SF36 health survey questionnaire: an outcome measure suitable for routine use in the NHS? *BMJ* 1993;306:1440-4.
  19. Hays RD, Stewart AL. The structure of self-reported health in chronic disease patients. *J Consult Clin Psychol* 1990;2:22-30.
  20. Ware JE, Brook RH, Davies-Avery A, et al. Conceptualization and measurement of health for adults in the Health Insurance Study. Volume I: Model of health and methodology. Santa Monica, CA: The RAND Corporation (publication no. R-1987/1-HEW), 1980.
  21. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31:247-63.

# 中文版SF-36健康量表使用於健康檢查民衆的信效度評估

李采娟<sup>1,2</sup> 劉秋松<sup>3</sup> 林正介<sup>3</sup> 李毅達<sup>4,6</sup> 賴俊雄<sup>5</sup> 李佳霽<sup>2</sup>

中國醫藥大學 中國醫學研究所<sup>1</sup> 職業安全衛生學系<sup>5</sup>

中國醫藥大學附設醫院 醫學研究部<sup>2</sup> 家庭醫學科<sup>3</sup>

成功大學 醫學院<sup>4</sup> 台灣禮來股份有限公司<sup>6</sup>

**目的** SF-36量表可用於測量一般民衆及不同疾病病患的健康狀態，然而，要廣泛使用中文版的SF-36量表之前，評估該工具的信效度是重要的工作。因此本研究目的係為評估中文版SF-36量表使用於健康檢查民衆的信效度。

**方法** 採用橫斷性研究，以1996年至某醫學大學附設醫院接受健康檢查民衆為對象，獲得18歲以上至65歲以下有效樣本共434人，其完訪率為68.7%。主要測量工具為SF-36量表，該量表測量八個構面，八個構面分別為身體活動功能、社交功能、因身體活動功能致角色受限情形、心理健康狀態、因心理健康狀態致角色受限情形、活力狀態、身體疼痛及自覺健康狀態。

**結果** 各構面均符合項目內部一致性(100%符合)及項目鑑別效度(達95.4%符合)的檢測，內部一致性的係數值範圍以身體疼痛構面的0.66最低，社交功能及身體活動功能構面的0.89最高；因素效度分析結果顯示近似於原作者在發展SF-36量表的觀念。另發現不同年齡層其身體活動相關構面如身體活動功能及身體疼痛構面分數有顯著不同；在有無身心症者間所有構面均有顯著不同。

**結論** 使用於民衆健康檢查的中文版SF-36量表有良好的信度及效度。(中台灣醫誌 2005;10:8-17)

## 關鍵詞

SF-36健康調查，健康檢查，效度

---

聯絡作者：李采娟

地址：404台中市北區學士路91號

中國醫藥大學 中國醫學研究所

收文日期：2004年8月31日

修改日期：2004年11月18日

接受日期：2004年12月15日