行政院國家科學委員會專題研究計畫成果報告

測量醫療結果之新工具-中文版健康亮表 SF-36 效度評估之研究

The Emerge of A New Instrument for Medical Outcome - A Study of A Chinese-Version Short Form 36

計畫編號: NSC 87-2314-B-039-023
執行期間: 1997 年 7 月 1 日至 1998 年 6 月 30 日

計畫主持人: 李采娟

執行單位: 私立中國醫藥學院

中華民國八十七年六月三十日

English Abstract

As the objectives of medical care for patients not only in prolonging the duration of life but also in improving the quality of life, achieving a more effective life and preserving function and well-being have been recognized, it has been increasing consensus about the importance of the centrality of the patient point of view in monitoring the quality of medical care outcomes. Therefore, the main efforts of the past decades is using standardized patient surveys to collect information on general health outcomes in order to serve research effectively. A new generic health-status measure, the SF-36, has been introduced and has attracted considerable interest. SF-36 has shown significant reliability and validity in four fields of application: monitoring population health, estimating the burden of different conditions, clinical trials of treatment effects, and monitoring outcomes in clinical practice.

The Chinese-version SF-36 was developed following the guidelines for cross-cultural adaptation. The primary objectives of this study are to provide estimates of reliability and validity of Chinese version SF-36 among individuals selected from general population and primary care attenders as well as to compare their health status between two groups.

A cross-sectional study with two samples was designed to collect information about SF-36, life events, clinical diagnosis, and sociodemographic factors. One sample will be recruited from primary care attenders in a large teaching hospital in Taichung while the other will be selected from a general population in Taichung using multistage sampling method.

Analyses were conducted across 15 subgroups differing in sociodemographic characteristics and chronic conditions. For both samples of each scale, item-completion rates were high across all subgroups (97.6% to 99.8%), but tended to be somewhat lower among the elderly and those with chronic disease. On average, surveys were complete enough to compute scale scores for more than 97% of the sample. For random sample, all scales passes tests for item-internal consistency (100% passed) and item-discriminant validity (98.9%passed). For outpatient sample, all scales pass tests for item-internal consistency (96.4% passed) and item-discriminant validity (94.3%passed). Reliability coefficients ranged from a low of 0.76 to a high of 0.93 across scales for random sample and from a low of 0.61 to a high of 0.93 for outpatient sample. Validation by factor analysis yielded results remarkably similar to those proposed by the authors who developed SF-36 for both

samples. For the comparisons of all scales of SF-36 between random and outpatient samples, subjects from primary care settings reported significantly compromised health status compared to subjects of general population after considering the effect of age, gender, education, and chronic conditions.

Keywords: short form 36 (SF-36); health status; validity; reliability; primary care; general population;

中 文 摘 要

由於醫療照護不僅要延長生命，並且改善生活品質、達到更有效之生活且保有功能和健康支主要目的已被認識，因此，以病人的觀點來偵測醫療照護結果的重要性已漸趨一致。過去十年來，以病人爲對象之研究調查主要的努力乃使用標準化的問卷來收集有關健康狀態的資料，使研究調查更有效率。

一個新的一般性健康狀態量表，SF-36，已被引進且引起各界相當大之興趣，不管在偵測一般民眾之健康狀態、預估不同情形之健康負擔、臨床之療效評估或是偵測臨床之照護結果，SF-36 皆顯示有相當高之信度及效度。

中文版的 SF-36 在遵循不同文化改編之指導下已被翻譯，有兩個主要原因突顯出中文版 SF-36 效度評估之重要性，一爲避免翻譯的不完全所導致降低內容效度，另一爲效 度之研究將提供有關 SF-36 分數所代表之健康意義，因此，本研究之主要目的爲評估中文版 SF-36 在一般民眾與家庭醫學科病人之信度與效度，並比較其健康狀態之差異。

本研究將採橫斷面研究設計來收集相關之資料，研究對象將選取 300 名前來中國醫藥學院附設醫院家醫科門診之病人與 500 名一般民眾，一般民眾選自戶籍設於台中市之市民，抽樣方法爲三階段等機率隨機抽樣方法，所收集之資料包括 SF-36、慢性病數與人口因子等。

本研究之分析依十五個以人口因子及慢性病情形區分之組別做比較，不論一般民眾或家醫科病人其項目完成率在各個組別皆相當高(97.6% 至 99.8%)，其中年紀較高者與有慢性病者之項目完成率稍微較低。平均而言，量表之之完成率不論一般民眾或家醫科病人皆在 95%以上，一般民眾之會聚效度達 100%，鑑別效度則爲 98.9%，家醫科病人之會聚效度爲 96.4%，鑑別效度則爲 94.3%，一般民眾內部一致性之信度則介於0.76至0.93之間而家醫科病人之內部一致性之信度則介於 0.61 至 0.93，以因素分析來評估建構效度之結果顯示不論一般民眾或家醫科病人所形成之因子皆和原作者所提出之量表建構相當的接近，表示有很好之建構效度，在兩樣本的比較方面，在調整年齡、性別、教育及慢性病情況後，八個 SF-36 向量中家醫科病人之分數顯著低於一般民眾，此結果顯示 SF-36 健康量表有能力區分不同的樣本，此提供了 SF-36 健康量表具有建構效度之另一個證據。

關鍵詞：健康量表；健康狀態；生活事件；信度；效度；

Introduction

The Need of Medical Outcome Instrument

As the objectives of medical care for patients not only in prolonging the duration of life but also in improving the quality of life, achieving a more effective life (McDermott, 1981) and preserving function and well-being (American College of Physicians, 1988; Cluff, 1981; Ellwood, 1988; Schroeder, 1987; Tarlov, 1983) have been recognized, it has been increasing consensus about the importance of the centrality of the patient point of view in monitoring the quality of medical care outcomes (Geigle & Jones, 1990; Ware & Sherbourne, 1992).

However, information from patients about their experiences of disease and treatment has never been routinely collected in clinical research or medical practice and this information is not a part of the medical record, either. Therefore, it is unavailable for analysis in the current health care database.

Another reason for the need of medical outcome instrument is that traditional measures of morbidity and mortality are generally agreed to be too narrow to measure the potential benefits of health care interventions, which can influence a wide number of variables such as physical mobility, emotional well-being, social life, and overall well-being (Brazier, 1992). Therefore, many specific questionnaires were developed and were intended to encompass aspects of physical, psychological and social well-being in evaluating the outcomes of different forms of treatment and care (Wilkin, 1992; McDowell, 1987; Bowling, 1991). These consist of both disease-specific measures which are designed to be sensitive to the outcomes of particular disease processes and to characterize the impact of the disease, and generic measures which are designed to be applicable across a wide range of medical conditions. But disease-specific clinical measures do not provide a complete picture of the impact of the disease upon patients. Most importantly, they fail to address the impact of the illness upon subjectively assessed function and well-being of patients (Longstreth, 1992). The inclusion of generic measures has been treated as central to the evaluation of treatment regimens and surveillance of disease progression (Fitzpatrick, 1994). However, a generic questionnaire that is easy to administer, acceptable to patients, and short as well as being fully validated is few. Nottingham health questionnaire was one of the more widely used with its acceptability, and easy administration, but it has been criticized that it is not able to detect low levels of disability, which are important not only clinically but also to respondents (Ware &

Sherbourne, 1992).

Policy analysts also need to use information about functional status, well-being, and other important health outcomes to compare the costs and benefits of competing ways of organizing and financing health care services and managers of health care organizations seek to produce the best value for each health care dollar. Health outcome information will also be utilized by clinical investigators to evaluate new treatments and by practicing physicians and other providers to achieve the best possible patient outcomes. Therefore, the main effort of the past decades is using standardized patient surveys to collect information on general health outcomes in order to serve research effectively (Ware & Sherbourne, 1992).

Advances in Developing Medical Outcome Measures

Significant advances in methods for assessing patient perspectives about functional status, well-being and other important health care outcomes during the past decade are (1) an improved understanding of the major dimensions of health and the validity of specific scales in relation to those dimensions (Hays & Stewart, 1990, Liang, 1986; Ware et al., 1981); (2) demonstration of the usefulness of standardized health surveys in clinical trials (Bombardier et al., 1986; Croog et al., 1986; Fowler et al., 1988); (3) health policy evaluation (Brook et al., 1983; Ware et al., 1986); (4) general population health surveys (Bergner et al., 1981; Stewart et al., 1988, 1989; Ware et al., 1986) and medical practice (Nelson & Berwick, 1989).

The Short Form 36 (SF-36)

A new generic health-status measure, the SF-36, has been introduced and has attracted considerable interest. The SF-36 is referred to as a generic measure because it assesses health concepts that represent basic human values that are relevant to everyone‘s functional status and well-being (Ware, 1987, 1990a). Such measures are called generic not only because they are universally valued but also because they are not age, disease, or treatment specific. Generic health measures assess health-related quality of life outcomes, namely, those known to be most directly affected by disease and treatment.

Generic health measures are not designed to serve as substitutes for traditional measures of clinical endpoints but to test generic health measures in parallel with clinical measures (Ware & Sherbourne, 1992). The potential of such comparisons is illustrated in the profiles of functional status and well-being for patients with different medical and psychiatric conditions and in contrast to the general U.S. population. These comparisons serve at least two important purposes. The comparisons test the

validity of generic health measure scales in describing groups of patients known to differ in functional status and well-being. These comparisons also facilitate understanding among clinicians of the meaning of differences in generic health measures scale scores because these diagnostic groups are familiar.

A long battery of health measures has its excellent characteristics in terms of traditional psychometric standard of reliability, validity, and precision. But its advantages have been cut down by considerable costs of data collection and respondent burden in terms of new psychometric standard of feasibility and practicality. During the stage of developing SF-36, the researcher attempts to achieve reductions in respondent burden with sacrificing measurement precision below the critical level (Ware & Sherbourne, 1992). This reduction was accomplished by constructing scales from more efficient items. In the Health Insurance Experiment (HIE), for example, 25 items were necessary to define seven levels of physical functioning (Stewart et al., 1978). With the SF-36 Physical Functioning, only 10 items are necessary to define 21 level of functioning (Stewart & Kamberg, 1992). In addition, empirical data confirmed its acceptablility, quickness, comprehensibility, the appropriateness of the items and coverage among elderly people, young adult population, and patients with Parkinson‘s disease. (Hayes, 1994). The SF-36 has also been compared in “normal” populations with the Nottingham health profile, and has been reported to be preferable for measuring improvements in health in a population with relatively minor conditions such as in general practice or in the community (Brazier, 1992). This is because more subjects use a wider range of scores, which leads to a greater power to discriminate between groups.

Dimensions of SF-36

Although SF-36 try to reduce respondent‘s burden, it still has 8 categories of operational definitions to measure four health concepts: which includes one multi-item scale measuring each of eight health concepts (a) behavioral functioning, (b) perceived well-being, (c) social and role disability, and (d) personal evaluations (perceptions) of health in general. Table 1 shows the physical and mental health phenomena assumed to be represented by scales, which (1) physical functioning (PF), (2) role limitation due to physical health problems (RP), (3) bodily pain (BP), (4) general health (GH), (5) vitality (energy/fatigue) (VT), (6) social functioning (SF), (7) role limitations due to emotional problems (RE), and (8) mental health (MH) (psychological distress and psychological well-being).

Table 1. Summary of health phenomena captured by SF-36 scales.

	Function		well-being		disability		perception	
	mental	physical	mental	physical	mental	physical	mental	physical
PF		x						
RP						x		
BP				x		x		
GH	x						x	x
VT			x	x				
SF					x	x		
RE					x			
MH			x					

The SF-36 has undergone a considerable amount of testing for reliability and validity among patient population in the USA and shown to detect differences in health status for patients with different types and severity of medical condition (Ware, 1992; Jenkinson, 1993; Brazier, 1992; Garratt, 1993; Jenkinson, 1993; Lyons, 1994; McHorney, 1992; 1993). This provides evidence for the potential value of this measure in identifying sickness-related dysfunction among patients. It has been adapted for use in UK, with the wording of six questions slightly altered, and it has also been demonstrated to achieve high levels of reliability and construct validity among community and patient populations in the UK (Brazier, 1992; Jenkinson C, 1993; Garratt, 1993).

Fourteen reliability assessment studies had reported that all estimates exceeding accepted standards for measures used in group comparisons. For each scale, the median of the reliability coefficients across studies equals or exceeds 0.80, with the exception of the Social Functioning scale (the median for this two-item scale is 0.76). These results support the use of the SF-36 scales in studies of health status that are based on group-level analyses. Only the Physical Functioning scale consistently exceeded the 0.90 standard of reliability, which some consider a minimum requirement for comparisons of scores for individual patients (Ware, 1992).

Table 2 represents the summary information of the SF-36. The scales of validity in Table 2 was ordered according to their validity, from the scale known to be the most valid measure of the physical component of health status, PF, to the last scale in the table, MH, which is the most valid measure of the mental component of health status. Interestingly, MH is the poorest measure of the physical component, and PF is the poorest measure of the mental component. Scales in between PF and MH are ordered according to their validity in measuring physical and mental components.

The SF-36 survey of generic health concepts is a promising tool for monitoring the results of care. Prior to the Sf-36, not only none of health outcome measuring generic functional status and well-being measures had received widespread adoption, but also none had been shown to be suitable for use across diverse populations and health care settings. As a result, the opportunity to describe differences in functioning and well-being for both the sick and the well was lost. Little was known about how patients suffering from one chronic medical or psychiatric condition differed from each other in terms of functional status and well-being. The SF-36 provides a common yardstick to compare those patients with chronic health problems to those sampled from the general population.

In summary, factors limiting the rate of progress in monitoring health outcomes from the patient point of view have included the absence of measurement tools with good psychometric properties that are easily administered and well documented. The SF-36 offers one approach for achieving these objectives. Standardization of SF-36 content and scoring will make meaningful interpretation and comparisons of results across studies possible.

Table 2. Information about SF-36 health status scale.

	no of item	no of level	reliability	Validity P	Validity M	Meaning of Score Low	Meaning of Score High
PF	10	21	0.93	*	-	Limited a lot in performing all physical activities	Performs all types of physical activies
RP	4	5	0.89	*	-	Problems with work or other daily activities	No problems with work or other daily activities
BP	2	11	0.90	*	-	Very severe and extremely limiting pain	No pain or limitations due to pain
GH	5	21	0.81	+	+	Evaluates personal health as poor and believes it is likely to get worse	Evaluates personal health as excellent
VT	4	21	086	+	+	Feels tired and worn out all of the time	Feels full of pep and energy all of the time
SF	2	9	068	+	*	Extreme and frequent interference with normal social activities due to physical or emotional problems	Performs normal social activities without interference
RE	3	4	082	-	*	Problems with work or other daily activities as a result of emotional problems	No problems with other work or other daily activities as a result of emotional problems
MH	5	26	0.84	-	*	Feelings of nervousness and depression all of the time	Feels peaceful, happy, and calm all of time
RHT	1	5	a	a	a	Believes general health is much better now than one year ago	Believes general health is much worse now than one year ago

a: validity is not available.

The Applications of the SF-36

Of the many potential applications of the SF-36, four examples are briefly

discussed below: (1) monitoring the health of the general population, (2) estimating the burden of different conditions, (3) clinical trials of treatment effects, and (4) monitoring outcomes in clinical practice.

Monitoring Population Health

The health of the general population in developed countries cannot be well understood from analyses of treatment survival rates or from population mortality statistics (Elinson & Mattson, 1984). Application of standardized generic measures of physical and mental function and well-being, social and role disability, and general health perceptions will make comprehensive monitoring of the health of the general population possible.

In order to measure the health of the general population and to compare different population groups, we can compare SF-36 profiles for different populations. The trend of the health between different populations will be revealed by the difference of mean scores of each scale. Standardization of the SF-36 for use in all countries will facilitate further study of population differences, specific treatment benefits, and various health care policy issues (Aaronson et al., 1992).

Estimating the Burden of Different Conditions

The SF-36 and other standardized assessment methods offer a number of advantages to providers. By standardizing questions, answers, and scoring, reliable and valid comparisons can be made to determine the relative burden of different conditions by comparing health profiles of each scale.

Clinical Trials of Treatment Effects

The SF-36 has been used to evaluate the burden of specific conditions such as the burden of heart disease and the benefits of heart valve replacement (Ware & Sherboure, 1992). To date, we are aware of more than dozen publications reporting results from clinical studies that included the SF-36 and there are about 150 topics under study in clinical trials using the SF-36 health survey.

Monitoring Outcomes in Clinical Practice

The SF-36 and other patient-based instruments have the potential to serve as "laboratory tests" of functioning and well-being in everyday medical practice (ACP, 1988). Their routine administration would be useful in : detecting and explaining decreased functional capacity and well-being, keeping track of changes in function

over time, making it possible to consider the patient's total functioning in choosing among therapies, guiding the efficient use of community resources and social services, and predicting more accurately the course of chronic disease.

The Needs of Chinese Version of SF-36

In addition to needs of medical outcome measures for monitoring population health, estimating the burden of different conditions, clinical trials of treatment effects, and monitoring outcomes in clinical practice, additional two urgent demands call for the need of Chinese version of SF-36 in Taiwan. They are as follows:

1. Research for health policy and health behavior:

In Taiwan, National Health Insurance (NHI) program was officially implemented in March 1, 1995. It provides universal coverage and reduces the price of medical care and deletes the barriers of health care utilization. Hence, people may increase the use of medical care after deletion of the economic barriers. Therefore, it is necessary to observe the effects of NHI on health care utilization and expenditure in order provide information for health policy makers. In order to evaluate the health care utilization given the same state of health, measuring the health status in general population would be an important issue.

2. Cross-cultural adaptation of medical outcome measures:

To compare health status outcomes across countries and conduct multinational trials of drug therapies and other treatments (Anderson, 1994), it is necessary to have standardized questionnaires and scoring methods as well as proof that the same health attributes are being measured in each country. A recent comparison of international health statistics underscored the consequences of a lack of standardized health status information; the report concluded that "there are virtually no population-based data available with which to make meaningful international comparisons on the prevalence of disease and disability" (Aaronson, 1988).

Translation of Chinese-version SF-36

The translation of Chinese-version SF-36 follows the guidelines for cross-cultural adaptation proposed by Guillemin et al (1993). Preliminary Chinese versions were obtained through a 6-month period of translation procedure. A group of bilingual collaborators who are researchers in studies of health behavior and health policy participated in the work of translation during this period. Because of the

cultural differences, the Chinese version and the English version differ in the wording of several items to make it acceptable for Chinese subjects. These differences have been discussed with the researchers who developed the original version. Back-translations were used during this translation process. And then a committee review the different versions of translation in order to produce a final version of the modified measure based on the various translations and back-translations. The four committe members are expert in health behavior and in the intent of the measure and the concepts to be explored. The final version of the SF-36 has been used among colleagues to examine its acceptability as well as content validity.

The Need of Validation of Chinese Version SF-36

Previous studies indicate that inadequate language translation may lead to reduction of the content validity of measurement (Berkanovic, 1980; Deyo, 1984), so it is necessary to validate the Chinese version SF-36 before it is widely used.

To make sure the data we collect do provide correct information, two crucial aspects of correctness should be considered: reliability, the extent to which measures give consistent or accurate results, and valid, the extent to which the results pertain directly to the desired attribute or characteristics being measured.

Validity studies help us to understand what a difference or a change in a score means. When enough evidence has been accumulated to show that a scale measures the intended health concept and does not measure other concepts, the scale is said to be validated. As long as the process of validation continues, new information is produced about the interpretation and meaning of scores.

Reliability

The evaluation of the reliability of any measurement procedure consists of estimating how much of the variation in a score is real or truth as opposed to chance or random errors (Selltiz et al., 1976). A reliability of 0.70 indicates that 70% of the measured variance is reliable; reliability coefficients are therefore proportions. Reliability examines the consistency of results from different measures designed to evaluate the same variable. Acceptable reliability differs depending on what is being analyzed: comparisons among individuals or across administrations to the same individual require high reliability (values >0.90); group comparisons, needed to compare average health status scores between diagnostic or treatment groups, do not require as high a reliability (values of 0.5 or 0.70 or higher are acceptable)

(Helmstadter, 1964; Nunnally, 1978).

Validity

a. Content Validity

The validity of questionnaires in the health field has most often been evaluated by means of content, construct, or criterion validation. Content validity (whether the test offers an adequate sample of the construct) is a challenge in the health field because of the breadth of health variables. Content validation requires the existence of a defining standard against which one can compare the content of a measure. Standards can be based on well-accepted theoretical definition, on published standards, or on interviews with those who are experiencing the types of health problems under study. When construction of the SF-36 began, Ware published a set of standards for evaluating the content validity of general health measures intended to be comprehensive (1987). These standard were applied in constructing the SF-36.

b. Construct Validity

When construct validation is used, both the test and the underlying theory must be evaluated. There are three steps to accumulate evidence of validity related to theoretical constructs: (1) specify the domain of variables, that is, prepare a blueprint for the constructs; (2) establish the internal structure of the observed variables; and (3) verify theoretical relationships between scale scores and external criteria (Ware, 1992). One method of testing the underlying theory is to test the differences between two patient groups known to differ in some way. For example, patients with a relatively minor and uncomplicated medical condition should score better in mental health (theorized construct) than patients with a psychiatric illness, and the average mental health scores of these patient groups should differ significantly. The mean difference between such groups in the Medical outcomes Study (MOS) was very large: 30.78 points on a 100 point scale (McHorney et al., 1993). The comparison demonstrates validity for the Mental Health (MH) scale because the mental health scores were much lower for patients with psychiatric disease (known to have poor mental health by definition of their disease).

c. Internal Validity: Convergent and Discriminant Validity

Convergent and discriminant validity are at the foundation of construct

validation. Convergent validity is supported when different methods of measuring the same construct provide similar results. Discriminant validity examines whether a measure of one underlying construct can be differentiated from another construct. For example, in the MOS, measures of physical functioning, mobility, and satisfaction with physical abilities were expected to yield results that "converage" at least moderately with one another because they are all hypothesized to assess physical health. In tests of discriminant validity, different measures are expected to yield different results. For example, one would not expect a measure of physical functioning to be highly related to a measure of depression or of loneliness (Ware, 1992).

d. Criterion validity

Criterion validity demonstrates that test scores are systematically related to one or more outcome criteria. This technique can be used when external evidence is available for use as a criterion against which the results of the test can be compared. The "criteria" were selected because they (1) are important (clinically, socially); (2) represent plausible outcomes of the variations in functioning and well-being measured by scales; and (3) were measured independently of the scale in question. Examples of correlations with external evidence occur when (1) health status and resource use are negatively correlated, (2) age and physical health are negatively correlated (according to the theory that physics function declines with increasing age), or (3) physical and mental health each have a positive correlation with general health (Ware, 1992).

e. Factorial Validity

Factor analysis provides an empirical test of the construct validity of the SF-36 in relation to its hypothesized structure. In the absence of agreed upon "criteria" for validating a scale, the validity of each scale can be tested using factor analytic methods.

The SF-36 was constructed to represent two major dimensions of health-physical and mental - that have been confirmed empirically in previous studies (Hays & Stewart, 1990; Ware, Davies-Avery, & Brook, 1980). Thus, two principal components can be constructed from the correlations among SF-36 health scales and rotated them to orthogonal simple structure. The orthogonal solution has the advantage of permitting interpretation of correlations across components to estimate the factor content of each scale.

Then two components were interpreted on the basis of their correlations with the SF-36 scales. If the pattern of results across scales was very consistent with expectation for physical and mental "dimensions" of health as summarized in Table 2, they were labeled "physical" and "mental" accordingly. If the two-dimensional structure had not been confirmed or the interpretation of the factors turned out to be ambiguous, these components could not have been used as "criteria" in testing the validity of each scale.

f. Clinical tests of Validity

Clinical tests of validity were based on criteria used to form mutually exclusive patient groups (McHorney et al., 1992; 1993). These groups differed in the severity of their conditions as defined by clinical measures of physical and mental (psychological) morbidity. In McHorney studies (1992,1993), the least severe comparison group was limited to patients with only a minor medical condition such as uncomplicated hypertension while the group with serious physical morbidity included patients with congestive heart failure (CHF) and complications (e.g., edema, orthopnea); myocardial infarction survivors with substantial morbidity (e.g., noteworthy and recurring angina and /or severe CHF symptomatology); hypertension patient with a history of a stroke; and diabetic patients with noteworthy complications (e.g., severe autonomic neuropathy). The group used to test validity in relation to clinical criteria of mental health was limited to patients with severe mental morbidity such as current unipolar affective disorder (major depression or dysthymia) or serious depressive symptoms.

1. Mean differences in SF-36 scale scores are for comparisons between the least severe group and the groups either severe physical or mental morbidity. Standardized effect size (ES) is the group difference divided by the general population standard deviation (SD). The relative validity (RV) is the ratio of pair-wise F-statistics, specifically the F for the comparison scale divided by the F for the most valid scale based on the same two-group comparison. The F-ratios analyzed in estimating RV are those for the difference between group means relative to the within-group means is larger and the error term is small. Thus, a larger F reflects greater discriminant validity and/or greater precision in estimating group means. RV estimates indicate how valid each scale is in discriminating between clinical groups, relative to the most valid SF-36 scale. RV is useful in addressing the issue of "conceptual relevance" and answers the question: how sensitive is each SF-36 health concept to differences in the levels of physical and mental morbidity defined by these clinical groups, relative to the

best scale?

Specific Aims

There are two aims in this study. The first aim of this study is to test the validity and reliability of a Chinese-Language version of the MOS 36-item short form health survey (SF-36) for measuring the health status among general population and primary care attenders. The second aim of this study is to compare the health status between these two populations. Therefore, the specific aims of this study are:

1. To provide estimates of reliability and validity of Chinese version SF-36 among two samples, one for general population and the other for primary care attenders.
2. To provide health profiles of Chinese version SF-36 among general population and primary care attenders and to make comparisons between these two groups.

Methods

Study Design

Cross-sectional study design with two samples will be used in the present study. SF-36, Chinese Health Questionnaire (CHQ), clinical diagnosis, sociodemographic factors are measured at the same time point. The structure of the study design is shown in Figure 1.

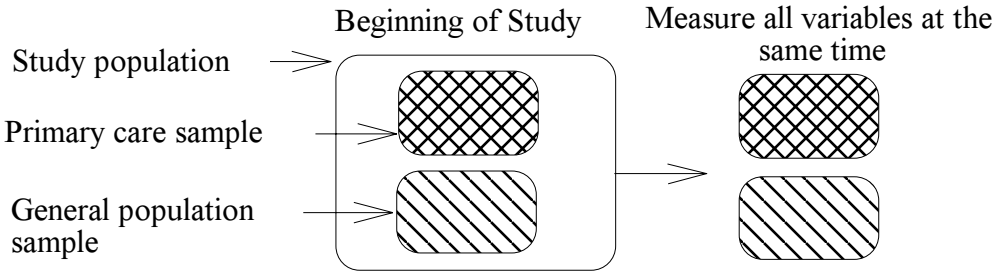

Figure 1: Study design of the present study.

Subjects

The Chinese version of the SF-36 was tested in two samples: (1) one from primary care attenders in a large teaching hospital; (2) the other from a general population in Taichung city.

Primary Care Sample

Six hundreds consecutive patients who attend general practice in China Medical College Hospital in Taichung were recruited. All study subjects were administered questionnaires to collect information about SF-36, CHQ, sociodemographic factors, medical history, and so on. The detailed steps for the administration of questionnaires are described in the section of Administration of the Questionnaire. Selection criteria for this study are those who could and would like to complete the self-rating questionnaires while exclusion criteria are those who have cognitive problems.

General Population Sample

The target population was all individuals who resided in Taichung City at the beginning of the study. The sampling frame of this study uses the set of all family records from Bureau of Household. Since Taiwan has good registration of household, we believe that this sampling would provide good reliability. Multistage sample design was used in this study, which consists of 4 strata and 2 sampling method. The four strata are the district, Li, household, and resident. There was a total of 600 residents selected in this study, and 425 agreed to participate. Thus the overall response rate is 70.83%. The sample size and sampling probabilities adopted in this study are determined by several considerations: power, available resources of the study, and heterogeneity of population. Within the first three strata, systematic sampling was applied with a probability that proportionate to the number of households in each sampling unit, whereas in the last strata, kish procedure was applied to select a residence from a household.

Evaluation of Non-response Bias

Data were also collected from non-response subjects to evaluate the possibility of non-response bias. To evaluate differences between response and non-response groups, face-to-face interview for primary care sample and telephone calls for general population sample were made to collect key variables in the questionnaire. These questions were formulated to approximate the same question in the written survey. These variables were calculated for responders and non-responders in order to evaluate the potential effects of non-response bias.

Administration of the Questionnaire

The questionnaire will be self-administered for primary care sample while it will be mailed out/mailed back for general population. The steps for administration of questionnaire for primary care attenders will follow the guidelines suggested by Ware et al. (1992).

The purpose of these guidelines is to establish rapport with the respondent and encourage completion of the questionnaire. The administrator can emphasize to respondents the importance of their answers to the completion of the study or to the addition to their medical records. The administrator can also answer questions and address concerns about the SF-36, and ensure the questionnaire is filled out correctly and completely.

Measurement

SF-36

The SF-36 is a short questionnaire with 36 items which measure eight multi-item variables: physical functioning (10 items), social functioning (2 items), role limitations due to physical problems (4 items), role limitations due to emotional problems (3 items), mental health(5 items), energy and vitality (4 items), pain (2 items), and general perception of health (5 items). There is a further unscaled single item on changes in respondents‘ health over the past year. For each variable item scores are coded, summed, and transformed to a scale from 0 (worst possible health state measured by the questionnaire) to 100 (best possible health state). For the SF-36, a high score indicates better perceived health state.

Clinical Criteria

Using clinical criteria, two mutually exclusive groups were formed: Group 1, no minor or uncomplicated chronic medical conditions; and Group 2, minor or serious (uncomplicated or complicated) chronic medical conditions. The classification of groups will be used to evaluate the clinical test of validity.

Patients classified as having a chronic medical condition included hypertension, diabetes mellitus, heart disease, anemia, incontinence of urine, duodenal ulcer, chronic hepatitis B, hepatitis C, tuberculosis, and so on.

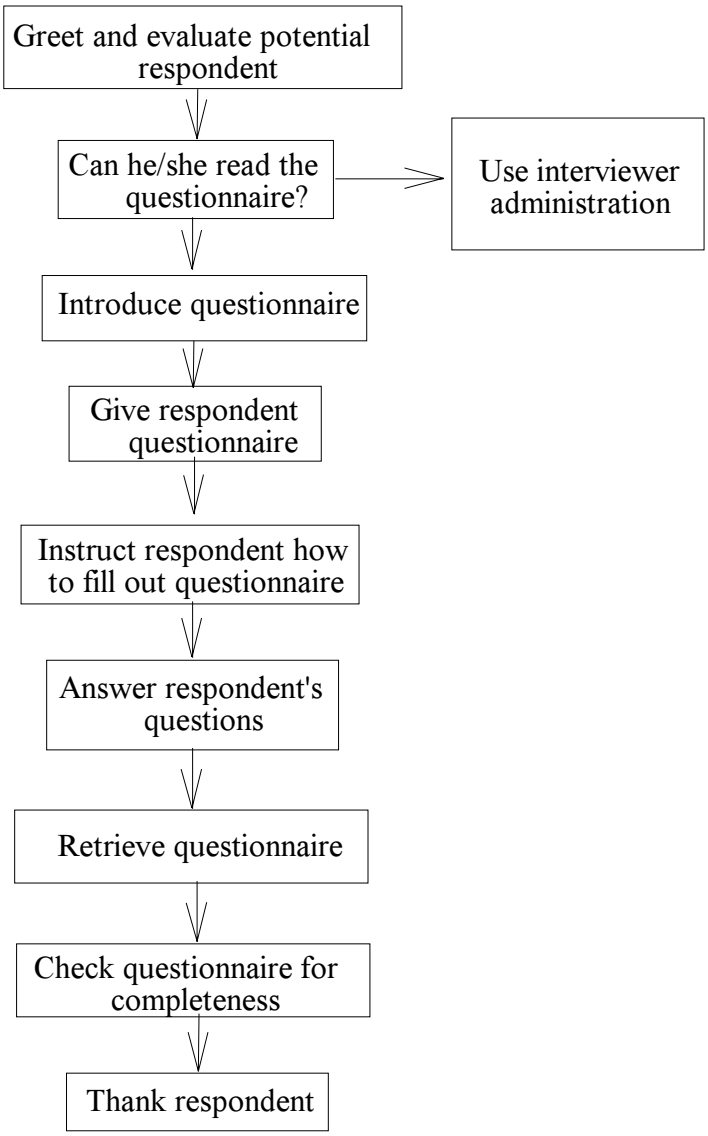

Figure 1: SF-36 administration flow chart for primary care attenders.

Sociodemographic factors

Age, gender, level of education will be collected in the questionnaire.

Life event

This variable was measured by a self-report questionnaire that consisted of 60 items grouped into 10 problem domains covering housing, work, financial status, legal matters, social and leisure activities, family status, child-parent interaction and marital relationship. For each of the 10 domains the presence of social problems was determined and the total score was then computed by adding up the number of domains for which social problems were identified.

Statistical Analysis

Reliability

The internal consistency form of reliability was assessed in this study. Internal consistency is the extent to which items within a dimension are correlated with each other. It will be examined by three methods: item-scale correlation (Streineer, 1990), and Cronbach alpha (Cronbach, 1951). Item-scale correlations, which assess the extent to which an item is related to the remainder of its scale, should exceed 0.4 (Kline, 1986) whereas Cronbach alpha, which measures the overall correlation between items within a scale, should exceed 0.7 (Nunnally, 1994) or 0.8 (Ware, 1993) to be considered acceptable.

Validity

Six aspects of validity will be evaluated: Internal validity (convergent and discriminant), criterion validity, construct validity, clinical test of validity, relative validity, and factorial validity.

Internal validity

The convergent and discriminant validity of SF-36 was examined by the multitrait multimethod matrix (Campbell & Fiske, 1959).. For convergent validity, the correlation between comparable dimensions on SF-36 and Chinese Health Questionnaire (CHQ) - for example, between mental health and depression and poor family relation - should be higher than the correlations between less comparable dimensions - For example, physical functioning and social dsyfunction. We'll test discriminant validity by comparing item to own scale correlation with item to other scale correlation. The item to own scale correlation should be higher if the categories within the SF-36 questionnaire are valid.

Construct, Criterion, and Clinical tests of validity

Construct validity assesses the extent to which a measure is related to criteria derived from an established clinical or social theory or "construct". One method is to examine construct validity, where hypotheses or constructs concerning the expected distribution of health between groups are examined by the measure being validated

(Streiner, 1989; McDowell, 1987). Therefore, the scales will be compared to assessments of physical and mental health based on information independent of SF-36. The study population was stratified along three variables corresponding to mental and physical health. The mental health variable has 3 levels:

Group 1: number of life events $\leqq 1$; Group 2: number of life events 2-5
Group 3: number of life events ≥ 6

Two variables for the physical health variable, the first one has 2 levels:

Group 1: no chronic condition; Group 2: any chronic conditions

The second one has 3 levels:
Group 1: 18-34 years old; Group 2: 35-49 years old; Group 3: ≥ 50 years old

As to clinical test validity, one-way ANOVA will be applied to make comparisons between these groups.

Relative Validity

The relative validity of each scale in measuring each dimension of health was assessed by the ratio of variance explained by the scale of interest (i.e., the scale coefficient squared) to the variance explained by the "best" scale (McHorney, 1993; Liang, 1985). Relative validity was assessed by the ratio of F-statistics, derived from one-way ANOVA models[12] for comparisons between mental health groups and among physical health groups. Again, the scale with the highest F-statistic was the reference with a relative validity of 1.

Exploratory Factory Analysis

In addition to evaluation of five aspects of validity, tests of validity will be applied in this study. Exploratory factory analysis (Child, 1990), a technique of psychometric validation, assesses the agreement between hypothetical factors that go to make up the measure and the scales designed to assess those factors. If the Chinese version of SF-36 is a valid measure for use, the scales defined by this authors should merge from a factor analysis of these two samples from general population, and items relating to a particular scale should be grouped together within a single factor. 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 1.1 (Jolliffe, 1986).

Results

Random Sample of General Population

Table 1 provides information on the distributions of sociodemographic characteristics, number of chronic diseases, and having illness during the past 6 months. Of 426 respondents, 155 (36.4%) were 18-34 years old, 231 (54.2%) were male, 220 (55.8%) had more than 12 years of education, 283 (69.4%) had income more than 3,5000 NT dollars, 53 (12.4%) didn't have any chronic disease, and 32 (7.5%) had been ill during the past 6 months.

The abbreviated English and Chinese content for each SF-36 item and scale assignment are shown in Table 2. These scales were constructed to be multidimensional. The SF-36 survey includes a single-item measure of health transition, which is not used to score any of the eight multi-item scales.

The number and percent of participants missing each of the 36 items is presented in Table 3 for the random sample of Taichung city population. Missing-value rates for the 36 items were consistently low, ranging from 0.2 to 1.2 and averaging 0.66.

Table 4 presents the percentage of items within each scale that were computable. For the total sample, these percentages were very high across scales, ranging from a low of 96.0% (RE) to a high of 99.5 (PF). Data completeness was not significantly different across scales among different subgroups. Older subgroups (>50 year old) and subgroup with chronic disease has slightly higher rate of complete items in all eight scales and subgroup with illness also has slightly higher rate of complete items in all scales except for PF and GH.

Average scores and quartiles of score distributions (Table 5) indicated that the population was generally in good health. Substantial ceiling effects were observed for 5 of the 8 scales while no substantial floor effect was observed for all 8 scales. The scales with substantial ceiling effects were physical functioning, role-physical, social functioning, and role-emotional.

Table 6 presents the item means and standard deviations and results of item-scale correlation coefficients. Standard deviations of items belonging to a given scale were fairly homogeneous. A possible exception was the physical functioning scale, where standard deviations varied from 0.27 to 0.68. This was due to higher proportion of respondents answering "limited a little" for "vigorous activities" than

other items. Three phenomena were observed from the correlation coefficients. The first one was that we observed fairly homogeneous correlation coefficients between an item and its hypothesized scale. The second is that almost all correlation coefficients between an item and its hypothesized scale had strong associations (≥0.7). The last was that the correlation coefficients between an item and other scales were much smaller than coefficients between an item and its hypothesized scale.

Results of scale tests, item-discriminant validity and item-convergent validity based on the matrix in Table 6 are presented summarized in Table 7. Perfect scaleing success rates for item-discriminant and item-convergent validity were achieved across all eight SF-36 scales. In 277 comparisons out of 280, the correlation between an item and its hypothesize scale exceeded correlations with all others scales by more than 2 standard errors. In addition, all items satisfied the criterion set a priori for convergent validity, i.e. a correlation with own scale ≥0.4. Thus, the success rate for discriminant validity was 98.9%, and for convergent validity, 100.0%.

Table 8 presents Cronbach's α across scales for overall group and 15 subgroups. These subgroups differed in terms of sociodemographic characteristics and chronic conditions. Overall, Cronbach's α ranged from 0.63 to 0.97. Minimum standards of reliability for purposes of group comparisons (≥0.5 or ≥0.7) were satisfied for overall group for all SF-36 scales in this population while 4 Cronbach's α for 15 subgroups did not satisfied with this minimum standards (scales of vitality and mental health for 9-12 years of education and social functioning scale for age >65 years old and for male). Among different scales, the social functioning scale had the lowest values of Cronbach's α; possibly because this scale contains only two items. It also had more variation across different subgroups relative to other scales, particularly for gender. The scale of role-physical and physical functioning had the highest internal consistency relative to the other scales for overall and all subgroups, and had more homogeneous coefficients across different subgroups. Role-physical was also the only scale that consistently exceeded the minimum standard of 0.90 for comparisons of scores for individual patients while physical function exceeded this standard except for subgroups of 35-49 years old and >12 year of education. In general, all Cronbach's α values of all scales were consistent across different subgroups.

Validation by Factor Analysis

Factor analysis identified seven relevant factors, with eigenvalues ranging from 1.01 to 14.02 and with proportions of total variance ranging from 2.87% to 40.05% (Table 9). The proportion of total variance of these seven factors explained by these items ranged from 59.0% (MH2) to 87.4% (for RP3) (not shown in the table). The mental health and vitality scales were combined together and then separated into two

factors (factors 2 and 5). Factor 1 was formed by 8 items of physical functioning and 1 item of social functioning. The other one item of social functioning (SF2) did not have any coefficient higher than 0.4, indicating little contribution to any factors. The highest coefficient of SF2 was 0.36 for factor 1. This might imply factor 1 corresponded to the combination of physical functioning and social functioning. The remaining 2 items of physical functioning combined with bodily pain and then formed factor 7. The other 3 factors corresponded to 3 scales of the SF-36: role-physical, general health perception, and role-emotional.

Validation by the Hypothesized Dimensionality of the SF-36 scales

We used principal component analysis to test the hypothesized dimensionality of the SF-36 scales. Because we hypothesized two dimensions to underline the structure of the eight scales, we extracted two principal components. To facilitate interpretation, we further rotated the components to orthogonal structure using the varimax method. The proportion of variability in one of the principal components explained by each scale was obtained by squaring the corresponding correlation coefficient. To evaluate the factorial validity of each scale as a measure of each component, we first squared each factor loading (scale-component correlation) to estimate the proportion of variance shared with that component (common-factor variance). We defined the scale sharing the most variance with each component as the most valid measure of that component. For each component, we then estimated relative validity (RV) for each scale by dividing the variance shared with the component by that estimate for the most valid scale. These ratios indicate in proportional terms how much less valid each scale is relative to the most valid scale. The higher the RV of a scale, the more precisely or efficiently it measures the underlying construct of interest as defined by the most valid scale.

Factor analysis of eight health scales produced 2 principal components. The first ("physical health") explained 56.5% of total variance, while the second ("physical health") explained 13.1%, for a total of 69.6%. The proportion of total variance explained by these 6 scales varied between 45 and 86%. Only 6 out of the 16 observed correlations between individual scales and principal components followed the pattern that was hypothesized by McHorney et al (Table 10). We found that scales of general health, social functioning, role-emotional, and mental health correlated more strongly with "physical" component than was predicted. Scales of physical functioning, role-physical, bodily pain, and general health correlated more strongly with "mental" component than was expected while role-emotional correlated slightly less strongly with "mental" component than was expected. Even though the

concordance rates with hypothesized correlations was low, the order of correlation within each component was generally consistent with a priori hypothesized by McHorney et al. The relative validity of a scale was given by the ratio of explained variance to that of the best scale: physical functioning for the "physical" component, and mental health for the "mental" component. In general, the patterns of relative validity were consistent with prediction.

Validation by Norm-based Interpretation

Lower scores on the SF-36 reflect poorer health state. Table 11 shows normative data in the form of means and standard deviations, broken down by age, gender, education, income, chronic disease and having illness. Overall, older subjects reported significantly poorer health on all scales of the SF-36 except for mental health than did younger subjects (all significant scales $p<0.001$, except for role-emotional $p=0.0346$). Women only reported poorer health on vitality scale than did men ($p=0.0146$). There were significant differences in scores among subjects with different levels of education on all scales of the SF-36 except for role-emotional and mental health ($p<0.001$ on vitality and physical functioning scales, $p<0.01$ on bodily pain, general perception of health, and social functioning; and $p<0.05$ on role-physical). Subjects with lower income reported poorer health on physical functioning, role-physical, general perception of health, and vitality ($p<0.001$ on general perception of health; $p<0.01$ on physical functioning and vitality; and $p<0.05$ on role-physical). Subjects with chronic disease had significantly lower scores on all scales than those without ($p<0.001$ on all scales except for role-emotional $p<0.01$). Subjects reporting an illness during previous 6 months had significantly lower scores on all scales than those without ($p<0.001$ on all scales except for physical functioning, social functioning, role-emotional, $p<0.01$, and mental health $p<0.05$).

Construction Validation

Table 12 shows the means score in the group with no chronic disease, mean difference between groups with and without chronic disease, F-statistics, and estimates of RV. Patients with any chronic diseases scored significantly lower on all eight scales compared to patients with no chronic disease. General health scale was the most valid in detecting differences between patients with and without chronic disease. Vitality scale was the second most valid scale, followed by the role-physical, social functioning, physical functioning, and bodily pain. As hypothesized, the best mental health scales (mental health and role-emotional) performed most poorly in this test.

Primary Care Sample

Table 13 provides information on the distributions of sociodemographic characteristics, number of life event, taking medicine, and having chronic disease among outpatients. Of 284 outpatients, 140 (49.3%) were 18-34 years old, 133 (47.0%) were male, 170 (73.9%) had more than 12 years of education, 228 (80.6%) had more than one life event during the past month, 138 (49.1%) were taking medicine, and 117 (41.8%) had any chronic disease.

The number and percent of outpatients missing each of the 36 items is presented in Table 14 for the outpatient sample of primary care setting. Missing-value rates for the 36 items were consistently low, ranging from 0.0 to 2.4 and averaging 1.44.

Table 15 presents the percentage of items within each scale that were computable for the outpatient sample of primary care setting. For the total sample, these percentages were very high across scales, ranging from a low of 97.6% (RE) to a high of 99.3 (MH). Data completeness was not significantly different across scales among different subgroups. In general, age group of 35 to 49 years old, education group less than 9 years, and subgroup with more than 5 life events had slightly lower rate of complete items in all eight scales.

Average scores and quartiles of score distributions (Table 16) indicated that the population was not in good health. Substantial ceiling effects were observed for 4 of the 8 scales and they are physical functioning, role-physical, bodily pain, role-emotional. Moderate floor effects were observed in scales of role-physical and role-emotional.

Table 17 presents the item means and standard deviations and results of item-scale correlation coefficients. Standard deviations of items belonging to a given scale were fairly homogeneous. A possible exception was the physical functioning scale, where standard deviations varied from 0.27 to 0.68. This was due to higher proportion of respondents answering “limited a little” for “vigorous activities” than other items. We also observed three phenomena from the item-scale correlation coefficients. The first one was that we observed fairly homogeneous correlation coefficients between an item and its hypothesized scale. The second is that almost all correlation coefficients between an item and its hypothesized scale had strong to moderate associations (0.7-0.3). The last was that the correlation coefficients between an item and other scales were much smaller than coefficients between an item and its hypothesized scale.

Results of scale tests, item-discriminant validity and item-convergent validity

based on the matrix in Table 17 are presented summarized in Table 18. Perfect scaling success rates for item-discriminant and item-convergent validity were achieved across 6 and 5 of eight SF-36 scales, respectively. In 270 comparisons out of 280, the correlation between an item and its hypothesize scale exceeded correlations with all others scales by more than 2 standard errors. In addition, all items except for 2 items, one for physical functioning and the other for mental health, satisfied the criterion set a priori for convergent validity, i.e. a correlation with own scale ≥0.4. Thus, the success rate for discriminant validity was 96.4%, and for convergent validity, 94.3%.

Table 19 presents Cronbach's α across scales for overall group and 15 subgroups. These subgroups differed in terms of sociodemographic characteristics, life events, and chronic conditions. Overall, Cronbach's α ranged from 0.61 to 0.89. Minimum standards of reliability for purposes of group comparisons (≥0.5) were satisfied for overall group for all SF-36 scales in this outpatient sample while 4 Cronbach's α for 15 subgroups were not satisfied with this minimum standards (scales of bodily pain for life events ≤ 1 and those without taking any medicine and mental health scale for education ≤ 9 years and life events ≤ 1.). These Cronbach's α below minimum standards also more varied across subgroups. Among different scales, the social functioning scale had the highest values of Cronbach's α, and next were role-physical, physical functioning, and role-emotional. In general, all Cronbach's α values of all scales were consistent across different subgroups.

Validation by Factor Analysis

Factor analysis identified 8 relevant factors, with eigenvalues ranging from 1.12 to 10.61 and with proportions of total variance ranging from 3.21% to 30.30% (Table 20). The proportion of total variance of these 8 factors explained by these items ranged from 41.6% (PF6) to 86.3% (for BP1) (not shown in the table). Physical functioning scale separated into 2 factors (factors 1 and 3). Mental health and vitality scales were combined together and then separated into two factors (factors 2 and 4). Factor 5 was formed by 3 items of role-emotional and 1 item of social functioning. Although the coefficient of the other item of social functioning (SF2) is not greater than 0.4 in factor 5, the coefficient of this social functioning item was highest in factor 5. The other 3 factors corresponded to 3 scales of the SF-36: role-physical, general health perception, and role-emotional.

Validation by the Hypothesized Dimensionality of the SF-36 scales

We used principal component analysis to test the hypothesized dimensionality of the SF-36 scales in this outpatient sample. Factor analysis of eight health scales

produced 2 principal components. The first ("physical health") explained 50.5% of total variance, while the second ("physical health") explained 13.0%, for a total of 63.5%. The proportion of total variance explained by these 6 scales varied between 52.0% and 77.9%. Only 8 out of the 16 observed correlations between individual scales and principal components followed the pattern that was hypothesized by McHorney et al (Table 21). We found that scales of physcial functioning and bodily pain did not correlate strongly enough with "physical" component than was predicted while role-emotional and mental health correlated slightly more strongly with "physical" component than was predicted. Scales of physical functioning, role-physical, bodily pain, and vitality correlated more strongly with "mental" component than was expected. Even though the concordance rates with hypothesized correlations was not high, the order of correlation within each component was generally consistent with a priori hypothesized by McHorney et al. The relative validity of a scale was given by the ratio of explained variance to that of the best scale: physical functioning for the "physical" component, and mental health for the "mental" component. In general, the patterns of relative validity were consistent with prediction.

Validation by Distinguishing Subgroups

Lower scores on the SF-36 reflect poorer health state. Table 22 shows means and standard deviations, broken down by age, gender, education, life event, taking medicine, and chronic disease. Overall, older subjects reported significantly poorer health on physical functioning and role-physical than did younger subjects ($p<0.001$ for physical functioning and $p<0.01$ for role-physical). Women only reported poor health on physical functioning, role-physical, general health, and role-emotional scales than did men (all $p<0.01$). There were significant differences in scores among subjects with different levels of education on physical functioning and role-physical scales of the SF-36 ($p<0.001$ on physical functioning scale and $p<0.01$ on role-physical). Subjects with higher number of life events reported poorer health on vitality, social functioning, role-emotional, and mental health (all $p<0.001$ except for mental health $p<0.01$). Outpatients who were taking medicine had significantly lower scores on all scales except for role-emotional ($p<0.001$ on role-physical, general health, and vitality; $p<0.01$ on bodily pain and mental health; and $p<0.05$ on social functioning). Subjects with chronic disease had significantly lower scores on all scales except for social functioning, role-emotional, and mental health than those without ($p<0.001$ on general health, $p<0.01$ on role-physical, and $p<0.05$ on physical functioning, bodily pain, and vitality).

Construction Validation

Table 23 shows the means score in the age group 18-34 years old, mean difference between age group of 18-34 years old and age groups of 35-49 and ≥ 50 years old, F-statistics, and estimates of RV. Older patients scored significantly lower on physical functioning, role-physical, bodily pain, and general health compared to younger patients. Physical functioning scale was the most valid in detecting differences between different age groups. Role-physical scale was the second most valid scale. The other scales performed pretty poorly in this test.

Table 24 shows the means score in the group with ≤ 1 life events, mean difference between group of ≤ life events and groups of 2-5 and ≥ 6 life events, F-statistics, and estimates of RV. Outpatients with more life events scored significantly lower on social functioning, role-emotional, and mental health compared to younger patients. Role-emotional scale was the most valid in detecting differences between different life event groups. Social functioning scale was the second most valid scale, followed by vitality and mental health. Physical functioning, role-physical, bodily pain, and general health scales performed pretty poorly in this test.

Regression Model of 8 Scales of SF-36

We then further examined the differences of 8 scales of SF-36 between random sample of general population and outpatient sample adjusting for the effects of age, gender, education, and chronic conditions (Table 25). We can see that the differences of 8 scales were all statistically significant after controlling for the other variables in the model, ranging from –2.9 (physical functioning) to –17.5 (role-emotional). Those who had chronic conditions also reported significantly lower scores of all scales after considering the other variables in the model, ranging from 3.9 (physical functioning) and 17.3 (role-physical). Scales of vitality and mental health were significantly higher among those having 9-12 years of education than among those having less than 9 years of education. Those having more than 12 years of education reported significantly higher scores than those having less than 9 years of education in scales of physical functioning, role-physical, vitality, social functioning, and mental health. Gender did not exert any significant effect on any scales of SF-36. Those who were greater than 50 years old reported significantly poorer health status than those who were 18-34 years old in scales of physical functioning, role-physical, bodily pain, and general health. The percentages of 8 scales of SF-36 explained by age, gender, education, chronic condition, and

outpatients ranged from 8.78% (bodily pain) to 17.31% (physical functioning).

Next we examine the impacts (from the standardized beta coefficients, not shown in the table) of age, gender, education, chronic condition, and outpatients on each scale of SF-36. Age group >50 years old had the greatest impact on physical functioning, followed by chronic condition and outpatient. For scales of role-physical and general health, chronic conditions had the greatest impact on them, followed by outpatient. Chronic conditions also had the greatest impact on bodily pain, and then followed by >50 age group. For scales of vitality, social functioning, role-emotional, and mental health, outpatient had the greatest impact on them, followed by chronic condition.

Discussion

The SF-36 is a brief and easy to use questionnaire. It has been shown to be suitable for self-administration or face-to-face interview among clinical patients and managed care organization members in several languages. The reliability and validity of the Chinese version of the SF-36 administered through face-to-face interview in a random sample of the general population and through self-administration in primary care settings have never been reported. Our study showed that the Chinese version SF-36 was favorable for face-to-face interviews and self-administration. For face-to-face interviews, it took just five minutes to complete and achieving a high response rate and remarkable low missing rate (0.2-1.4%) while for self-administration, it took about 10 minutes to complete and remarkable low missing rate (0.7-2.4%). Therefore, the Chinese version SF-36 questionnaire appears to be an acceptable measure of the health status of a Chinese general population.

Our findings supported the claims of internal consistency of the domains of the SF-36 across diverse groups 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. Two main differences from the hypothetical construct were observed in our population. First, the items of vitality were closely correlated with those of mental health scale, which is similar to the results of Garratt, *et al*. The items of these two scales consisted of two factors in our study, but only one factor in Garratt *et al*'s study. Second, the items of bodily pain clustered together with the items of physical functioning in the random sample of

general population while the items of bodily pain formed an independent factor in the primary care sample. In Garratt *et al*'s study, the items of role limitations due to physical problems cluster together in addition to those of bodily pain and social functioning. The items of the other four factors precisely corresponded to the hypothetical scales. Such precise correspondence between factors and scales is rare in factor analysis and thus confirms the validity of the SF-36 in a Chinese general population.

Estimates of internal consistency for the SF-36 scales across different cultural populations have been reported in 9 studies, as shown in Table 26. All estimates exceeded accepted standards for measures used for group comparisons. For each scale, the median of the reliability coefficients across studies exceeds 0.80, with the exception of the social functioning scale (the median for this two-item scale is 0.77). These results support the use of the SF-36 scales in studies of health status that are based on group-level analyses. No scale consistently exceeded the 0.90 standard of reliability, which consider a minimum standard for comparisons of scores for individual patients. All of the published coefficients exceed the minimum standard of 0.50 suggested by Helmstadter (1964) for group comparisons; all but three exceed the 0.70 standard for individual comparison suggested by Nunnally (1978).

Subjects with these previously identified chronic conditions reported significantly compromised health status compared to similar subjects without any chronic conditions after considering the effect of age, gender, education, and type of sample. Most of the effects were both statistically and clinically significant. For example, the subjects with chronic conditions had noticeably negative effects on the 8 SF-36 scales, ranging from 3.9 to 17.3 points below the scores for the subjects with no chronic conditions. The reduction in health status associated with chronic conditions was similar in magnitude than those reported for chronic physical illnesses such as low back pain, arthritis, and diabetes (Wells, K.B., et al., 1989), which implied the severe impact of chronic conditions. For instance, the negative effect of chronic conditions on general perception was 13.6 points, which was the same as the impact of diabetes and congestive heart failure (about 13 points).

Subjects from primary care settings reported significantly compromised health status compared to subjects of general population after considering the effect of age, gender, education, and chronic conditions. Most of the effects were both statistically and clinically significant. For example, the subjects from primary care settings had noticeably negative effects on the 8 SF-36 scales, ranging from 2.9 to 17.5 points below the scores for the subjects of general population. The reduction in health status

associated with chronic conditions was somewhat lower in magnitude than those reported for chronic physical illnesses such as low back pain, arthritis, and diabetes (Wells, K.B., et al., 1989). For instance, the negative effect of chronic conditions on general perception was 11.5 points, which was a little lower than the impact of diabetes and congestive heart failure (about 13 points). Age exerts significant effect on physical-related scales such as physical functioning, role-physical, bodily pain, and general health while gender does not exert any effect on 8 scales. Education exerts moderate positive effects on physical functioning, role-physical, vitality, social functioning, and mental health, ranging from 4.1 to 7.9 points higher the scores for the subjects with lower level of education. The reduction in health status associated with lower education was a little lower in magnitude than those reported for chronic physical illnesses such as low back pain, arthritis, and diabetes (Wells, K.B., et al., 1989). For instance, the negative effect of lower education on physical functioning in this study was –5.8, which was somewhat worse than the impact of arthritis (-7.5) and diabetes (-6.6) and low back pain (9.1).

Reference

Aaronson, N.K., Acquadro, C., Alonso, J., Apolone, G., Bucquet, D., Bullinger, M., Bungay, K., Fukuhara, S., Gandek, B., Keller, S., Razavi, R., Sanson-Fisher, M., Sullivan, S., Wood-Dauphinee, S., Wagner, A., & Ware, J.E. (1992). International quality of life assessment (IQOLA) project. Quality of Life Research, 1,349-351.

American College of Physicians. (1988). Comprehensive functional assessment for elderly patients. Annals of Internal Medicine, 109, 70-72.

Anderson, R. T., Aaronson, N. K., & Wilkin, D. Critical review of the international assessments of health-related quality of life. Quality of life research, 1993, 2, 369-95.

Bergner M, bobbitt RA, Carter WB, Gilson BS. The sickness impact profile: development and final revision of a health status measure. Med Care 1981; 19: 787-805.

Berkanovic E. The effect of inadequate language translation on Hispanics' responses to health surveys. Am J Publ Health 1980; 70: 1273-1281.

Berzon, R., Hays, R. D., & Shumarker, S. A. International use, application and performance of health-related quality of life instruments. Quality of life Research, 1993,2,367-68.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;;i:307-10.

Bombardier, C., Ware, J.E., Russell, I.J., Larson, M., Chalmers, A., & Read, J. L. (1986). Auranofin therapy and quality of life in patients with rheumatoid arthritis: Results of a multicenter trial. American Journal of Medicine, 81, 565-578.

Brook, R.H., Ware, J.E., Rogers, W.H., Keeler, E.B., Davies, A.R., Donald, C.A., Goldberg, G.A., Lohr, K.N., Masthay, P.C., &Newhouse, J.P. (1983). Does free care improve adults' health: Results from a randomized controlled trial. New England Journal of Medicine. 309, 1426-1434.

Brazier JE, Harper R, Jones NBM, et al. Validating the SF-36 health survey questionnaire: a new outcome measure for primary care. Br Med J 1992;305_160-4.

Bullinger, M., Anderson, R., Cella, D., & Aaronson, N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. Quality of Life Research, 1993,2,451-59.

Campbell DT, Fishe DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull 1959;56:81-105.

Cheng TA, Wu JT, Chong MY, & Williams P. Internal consistency and factor structure of the Chinese Health Questionnaire. Acta Psychiatr Scand. 82: 304-308, 1990.

Child D. The essentials of factor analysis. 2nd ed. London: Cassell, 1990.

Chong MY, Wilkinson G. Validation of 30- and 12-item versions of the Chinese Health Health (CHQ) in patients admitted for general health screening. Pshchol Med 1989:19; 495-505.

Cluff, L. E. (1981). Chronic disease, function and the quality of care. Journal of Chronic Diseases. 34, 299-304.

Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334.

Croog, S.H., Levine, S., Testa, M.A., Brown, B., Bulpitt, C.J., Jenkins, D., Klerman, G.L., &Williams, G.H. (1986) The effects of antihypertensive therapy on the quality of life. New England Journal of Medicine, 314, 1657-1664.

Elinson, J., & Mattson, M.E. (1984). Assessing the quality of life in clinical trials of cardiovascular therapies: Introduction to the panel presentations. In N.K. Wenger, M.E. Mattson, C.D. Furberg & J Elinson (Eds.), Assessment of quality of life in clinical trials of cardiovascular therapies (pp. 143-145). New York, NY: Le Jacq Publishing Company.

Ellwood, P.M. (1988). Outcomes management: A technology of patient experience [Shattuck Lecture]. New England Journal of Medicine. 318, 1549-1556.

Fowler, F.J., Wennberg, J.E., Timothy, R.P., Barry, M.J., Mulley, A.G., & Henley, D. (1988). Symptom status and quality of life following prostatectomy. Journal of the American Medical Association, 259, 3018-3022.

Garratt A, Ruta D, Abdalla MI, Buckingham JK, Russell IT. The SF36 health survey questionnaire: an outcome measure suitable for routine use in the NHS? Br Med J 1993; 306:1440-4.

Geigle, R., & Jones, S.B. (1990). Outcomes measurement: A report from the front. Inquiry, 27, 7-13.

Hays, R.D., & Stewart, A.L. (1990). The structure of self-reported health in chronic disease patients. Journal of Consulting and Clinical Psychology, 2,22-30.

Howard, K.I., & Forehand, G.G. (1962). A method for correcting item-total correlations for the effect of relevant item inclusion. Educational and Psychological Measurement, 22, 731-735.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.

Jenkinson C, Wright L, Coulter A. Quality of life measurement in health care: a review of measures and population norms for the UK SF-36. Oxford, Health Services Research Unit, Oxford University, 1993.

Jolliffe IT. Principal component analysis. New York: Springer Verlag, 1986.

Kerlinger FN. Foundations of behavioral research. New York: Holt, Rinehart, and Winston, 1973.

Kline P. A handbook of test construction. London: Methuen, 1986.

Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum 1985; 28:542-547.

Lyons RA, Perry HM, Littlepate BNC. Evidence of the validity of the short-form 36 questionnaire (SF-36) in an elderly population. Age Ageing 1994; 23:182-4.

McDermott, W. (1981). Absence of indicators of the influence of its physicians on a society's health: Impact of physician care on society. American Journal of Medicine. 70, 8333-843.

McDowell I, Newell C. Measuring health: A guide to rating scales and questionnaires. New Yourk, NY: Oxford University Press, 1987.

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: results from the medical outcomes study. Med Care 1992;30(suppl.): MS 253-65.

McHorney Ca, Ware JE, 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;31:247-263.

McHorney CA, Ware JE, Lu R, et al. The MOS 36-Item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994;32:40.

Nelson, D.C., & Berwick, D.M. (1989). The measurement of health status in clinical practice. Medical Care, 27, S77-S90.

Nunnally JC, Bernstein IH. Psychometric Theory, 3rd ed. New York: McGraw-Hill; 1994.

Streineer GL, Norman DR. Health measurement values: a practical guide to their development and use. Oxford: Oxford University Press, 1990.

Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press, 1989.

Schroeder, S.A. (1987). Outcome assessment 70 years later: Are we ready? New England Journal of Medicine, 316, 160-162.

Stewart, A.L., Greenfield, S., Hays, R.D., Wells, K.B., Rogers, W.H., Berry, S.D., McGlynn, E.A., & Ware, J.E. (1989). Functional status and well-being of patients with chronic conditions: Results from the Medical Outcomes Study. Journal of the American Medical Association, 262, 907-913.

Stewart, A.L., Hays, R.D., & Ware, J.E. (1988). The MOS Short-Form General Health survey: Reliability and validity in a patient population. Medical Care, 26, 724-735.

Tarlov, A. R. (1983). The increasing supply of physicians, the changing structure of the health-services system, and the future practice of medicine (shattuck Lecture). New England Journal of Medicine. 398, 1235-1244.

Ware JE. SF-36 Health Survey Manual & Interpretation Guide. Boston, MA.: The Health Institute; 1993.Ware JE, Brook RH, Williams KN, Stewart AL, Davies-Avery A. Conceptualisation and measurement of health for adults in the health insurance study. Vol I. Model of health and methodology. Santa Monica, CA, Rand Corporation, 1980. Publication no R-1987/1-HEW.

Ware, J.E. (1986). The assessment of health status. In L.H. Aiken & D. Mechanic (Eds.), Applications of social sciences to clinical medicine and health policy (pp. 204-228). New Brunswick, NJ: Rutgers University Press.

Ware, J.E. (1990a). Measuring patient function and well-being: Some lessons from the Medical Outcomes Study. In K.A. Heitgoff & Heitgoff & K.N. Lohr (Eds), Effectiveness and outcomes in health care: Proceedings of an invitational conference by the Institute of Medicine, Division of Health Care Services (pp. 107-119). Washington, DC: National Academy Press.

Ware JE, Sherbourne CD. The MOS 36-item short form health survey (SF-36):

conceptual framework and item selection. Med Care 1992;30:473-83.

Ware JE, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: A user‘s manual. Boston, MA: The Health Institute, New England Medical Center, 1994.

Ware JE, Keller SD, Gandek B, et al. Evaluating translations of health status surveys: Lessons from the IQOLA project, International Journal of Technolog Assessment in Health Care, 1995;11:525.

Wells, K.B., Stewart, A., Hays, R.D., Burnam, M.A., Rogers, W., Daniels, M., Berry, S., Greenfield, S., and Ware, J. (1989) The functioning and well-being of depressed patients: results from the Medical outcomes Study. J. Am. Med. Assoc. 262, 914-919.

Table 1. Sample size for subgroup analysis: A Random Sample of General Population From Taichung City.

	n	%
Age		
18-34	155	36.4
35-49	144	33.8
50-65	67	15.7
>65	60	14.1
Gender		
Male	231	54.2
Female	195	45.8
Education[a]		
<9	103	26.1
9-12	71	18.0
>12	220	55.8
Income[a]		
<NT$3,5000	125	30.6
≥NT$3,5000	283	69.4
Having Chronic Disease[a]		
No	373	87.6
Yes	53	12.4
Having Illness or Bed Days Past 6 months[a]		
No	393	92.5
Yes	32	7.5

[a]Numbers do not equal to 426 due to missing data.

Table 2. Item groupings and abbreviated item content for the SF-36 survey: A Random Sample of General Population From Taichung City

Health Scale		Abbreviated Item Content (in English)	Abbreviated Item Content (in Chinese)
Physical functioning (PF)	PF1	Vigorous activities	費力活動
	PF2	Moderate activities	中等程度活動
	PF3	Lifting or carrying groceries	手提雜貨
	PF4	Climbing several flights of stairs	爬數層樓梯
	PF5	Climbing one flight of stairs	爬一層樓梯
	PF6	Bending, kneeling, or stooping	彎腰、跪下或蹲下
	PF7	Walking more than a mile	走路超過 1 公里
	PF8	Walking several blocks	走過數個街口
	PF9	Walking one block	走過一個街口
	PF10	Bathing or dressing yourself	自行洗澡或穿衣
Role physical (RP)	RP1	Limited in the kind of work or other activities	可以做的工作或其他活動的種類受到限制
	RP2	Cut down the amount of time spent on work or other activities	做工作或其它活動的時間減少
	RP3	Accomplished less than would like	完成的工作量比您想要完成的較少
	RP4	Difficulty performing the work or other activities	做工作或其他活動有困難
Bodily pain (BP)	BP1	Intensity of bodily pain	身體疼痛程度有多嚴重
	BP2	Extent pain interfered with normal work	身體疼痛對日常工作妨礙程度如何
General health perceptions (GH)	GH1	Is your health: excellent, very good, fair, poor	目前的健康狀況
	GH2	My health is excellent	我的健康狀況好得很
	GH3	I am as healthy as anybody I know	和任何一個我認識的人來比，我和他們一樣健康
	GH4	I seem to get sick a little easier than other people	好像比別人較容易生病
	GH5	I expect my health to get worse	我想我的健康會越來越壞
Vitality (VT)	VT1	Feel full of pep	充滿活力
	VT2	Have a lot of energy	精力充沛
	VT3	Feel worn out	筋疲力竭
	VT4	Feel tired	覺得疲倦
Social functioning (SF)	SF1	Frequency health problems interfered with social activities	身體健康或情緒問題有多少時候會妨礙社交活動
	SF2	Extent health problems interfered with normal social activities	健康或情緒問題，對平常活動的妨礙程度如何
Role-emotional (RE)	RE1	Cut down the amount of times spent on work or other activities	做工作或其它活動的時間減少
	RE2	Accomplished less than would like	完成的工作量比您想要完成的較少
	RE3	Didn't do work or other activities as carefully as usual	做工作或其它活動時不如以往小心

Table 2. Item groupings and abbreviated item content for the SF-36 survey: A Random Sample of General Population From Taichung City

Health Scale		Abbreviated Item Content (in English)	Abbreviated Item Content (in Chinese)
Mental health (MH)	MH1	Been a very nervous person	非常緊張
	MH2	Felt downhearted and blue	覺得悶悶不樂和憂鬱
	MH3	Felt so down in the dumps nothing could cheer you up	非常沮喪，沒有任何事情可以讓您高興起來
	MH4	Been a happy person	覺得快樂
	MH5	Felt calm and peaceful	心情平靜
Reported change	TRAN	Rating of health now compared to one year ago	和一年前比較目前的健康狀況

Table 3. Item Frequency Distributions and Number and Percent Missing: A Random Sample of General Population From Taichung City.

	Item Frequency Distribution						Missing	
Item	1	2	3	4	5	6	f	%
PF1	45	111	266	-[b]	-	-	4	0.9
PF2	26	67	330	-	-	-	3	0.7
PF3	15	53	356	-	-	-	2	0.5
PF4	23	64	337	-	-	-	2	0.5
PF5	12	24	388	-	-	-	2	0.5
PF6	13	53	358	-	-	-	2	0.5
PF7	17	53	354	-	-	-	2	0.5
PF8	10	34	380	-	-	-	2	0.5
PF9	7	22	395	-	-	-	2	0.5
PF10	4	15	403	-	-	-	4	0.9
RP1	62	361	-	-	-	-	3	0.7
RP2	56	367	-	-	-	-	3	0.7
RP3	50	373	-	-	-	-	3	0.7
RP4	56	367	-	-	-	-	3	0.7
BP1	3	6	31	69	56	257	4	0.9
BP2	5	6	27	72	310	0	6	1.4
GH1	42	117	101	150	15	-	1	0.2
GH2	12	36	82	122	170	-	4	0.9
GH3	8	35	72	204	100	-	7	1.6
GH4	13	72	119	78	138	-	6	1.4
GH5	13	42	69	197	99	-	6	1.4
VT1	6	37	71	124	120	67	1	0.2
VT2	5	43	75	116	135	51	1	0.2
VT3	1	6	34	154	155	74	2	0.5
VT4	3	6	44	187	143	41	2	0.5
SF1	6	6	23	45	342	-	4	0.9
SF2	9	7	36	130	240	-	4	0.9
RE1	44	379	-	-	-	-	3	0.7
RE2	55	368	-	-	-	-	3	0.7
RE3	47	376	-	-	-	-	3	0.7
MH1	3	12	24	140	162	84	1	0.2
MH2	2	6	9	103	187	118	1	0.2
MH3	1	13	61	114	160	76	1	0.2
MH4	2	4	17	116	185	101	1	0.2
MH5	1	29	91	128	125	50	2	0.5
TRAN	12	117	242	43	10	-	2	0.5

[a]N=428.

[b]Responses not possible

Table 4. Percent of patients for whom scale scores were computable.

	PF	RP	BP	GH	VT	SF	RE	MH
Total	99.5	99.3	99.4	97.3	97.2	99.6	96.0	99.1
Age (years)								
18-34	99.4	99.4	98.7	97.4	98.7	97.4	99.4	99.4
35-49	99.3	98.6	97.9	98.6	100.0	98.6	98.6	100.0
50-65	100.0	100.0	98.5	100.0	100.0	100.0	100.0	100.0
>65	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Gender								
Male	99.1	99.1	98.3	98.3	99.6	98.7	99.1	99.6
Female	99.5	99.5	99.0	99.0	99.5	98.5	99.5	99.5
Education								
<9	100.0	100.0	99.0	100.0	99.0	99.0	100.0	100.0
9-12	98.6	98.6	97.2	95.8	100.0	98.6	98.6	100.0
>12	99.5	99.1	98.6	98.6	99.1	98.2	99.1	99.5
Income								
<NT$3,5000	99.2	99.2	99.2	99.2	99.2	99.2	99.2	99.2
≥NT$3,5000	99.6	99.3	98.6	98.2	99.6	98.6	99.3	100.0
Having Chronic Disease								
No	99.5	99.2	98.4	98.4	99.5	98.4	99.2	99.7
Yes	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Having Illness or Bed Days Past 6 months								
No	99.5	99.2	98.5	98.7	98.5	98.5	99.2	99.7
Yes	96.9	100.0	100.0	96.9	100.0	100.0	100.0	100.0

Table 5. Description of 8 scales of the MOS 36-item Short Form Health Survey (Chinese version) submitted to a random sample of 426 residents in Taichung city, Taiwan.

	Physical functioning	Role-physical	Bodily pain	General health	Vitality	Social functioning	Role-emotional	Mental health
Number of items	10	4	2	5	4	2	3	5
Percentage non-respondents	0.5	0.7	1.4	1.4	0.5	1.4	0.7	0.2
Mean score	89.9	86.8	79.3	66.8	66.7	88.4	88.5	72.4
Percentiles; 25^{th}	90	100	72	52	55	87.5	100	64
Percentiles; 50^{th}	100	100	90	67	70	100	100	72
Percentiles; 75^{th}	100	100	90	82	80	100	100	84
Percentage at ceiling	59.7	81.3	0	2.9	2.8	55.5	84.4	3.3
Percentage at floor	0.2	9.0	0.7	0.5	0	1.0	6.6	0

Table 6. Item means and standard deviations and correlations between SF-36 items and hypothesized scales in a random sample of 426 residents in Taichung city, Taiwan.

Item	Mean	SD	PF	RP	BP	GH	VT	SF	RE	MH	TRAN
PF1	1.52	0.68	.68*	.45	.50	.51	.40	.51	.30	.21	.30
PF2	1.72	0.57	.79*	.41	.55	.47	.40	.54	.32	.24	.29
PF3	1.80	0.48	.82*	.41	.54	.42	.35	.49	.34	.20	.28
PF4	1.74	0.55	.85*	.45	.52	.45	.37	.54	.30	.20	.24
PF5	1.89	0.40	.77*	.48	.49	.37	.33	.51	.36	.16	.27
PF6	1.81	0.46	.76*	.42	.51	.38	.33	.51	.28	.20	.24
PF7	1.79	0.49	.80*	.51	.50	.44	.40	.58	.33	.22	.30
PF8	1.87	0.40	.78*	.47	.45	.40	.38	.54	.36	.21	.27
PF9	1.92	0.33	.71*	.40	.46	.41	.40	.51	.31	.21	.27
PF10	1.95	0.27	.64*	.34	.41	.33	.28	.56	.29	.18	.26
RP1	0.85	0.35	.39	.80*	.45	.44	.40	.47	.52	.28	.26
RP2	0.87	0.34	.41	.83*	.46	.42	.39	.50	.61	.31	.24
RP3	0.88	0.32	.58	.86*	.50	.44	.40	.55	.47	.25	.28
RP4	0.87	0.34	.54	.82*	.48	.47	.44	.54	.50	.31	.32
BP1	4.23	1.12	.57	.48	.69*	.53	.44	.56	.35	.35	.34
BP2	3.61	0.77	.58	.51	.76*	.49	.46	.68	.46	.36	.32
GH1	1.95	1.08	.46	.45	.47	.53*	.59	.43	.32	.41	.47
GH2	2.95	1.09	.38	.39	.46	.72*	.58	.45	.28	.48	.33
GH3	2.84	0.95	.44	.37	.39	.69*	.50	.47	.31	.42	.29
GH4	2.61	1.19	.36	.32	.40	.64*	.49	.44	.22	.45	.35
GH5	2.78	1.02	.47	.44	.48	.79*	.62	.51	.37	.50	.33
VT1	3.21	1.23	.43	.40	.43	.61	.62*	.48	.36	.56	.39
VT2	3.14	1.21	.43	.38	.44	.67	.68*	.50	.35	.64	.40
VT3	3.60	0.93	.22	.29	.28	.38	.52*	.37	.26	.59	.18
VT4	3.38	0.90	.28	.31	.31	.48	.62*	.43	.31	.57	.28
SF1	3.68	0.76	.62	.52	.63	.49	.46	.58*	.45	.32	.30
SF2	3.39	0.87	.53	.50	.54	.53	.56	.58*	.43	.51	.28
RE1	0.90	0.31	.35	.55	.40	.35	.36	.47	.78*	.35	.22
RE2	0.87	0.34	.38	.56	.38	.33	.37	.44	.83*	.37	.22
RE3	0.89	0.31	.30	.45	.36	.32	.38	.41	.69*	.34	.21
MH1	3.64	1.00	.02	.13	.19	.26	.43	.19	.20	.47*	.09
MH2	3.93	0.90	.28	.30	.33	.44	.54	.43	.36	.64*	.19
MH3	3.52	1.05	.15	.24	.28	.43	.59	.36	.31	.65*	.23
MH4	3.84	0.90	.27	.25	.34	.49	.61	.42	.34	.70*	.22
MH5	3.17	1.12	.23	.27	.27	.49	.66	.36	.27	.56*	.28
TRAN	1.82	0.74	.33	.30	.36	.45	.41	.32	.24	.27	-

Table 7. Scaling properties of the 8 scales of the SF-36 in a Chinese general population, Taichung, Taiwan.

Scale	Items' standard deviation (range)	Correlations of items with own scale (range)	Correlations of items with other scales (range)	Convergent validity[a]	Discriminant validity[b]
Physical Functioning	0.27~0.68	0.64~0.85	0.16~0.58	10/10 (100%)	80/80 (100%)
Role-Physical	0.32~0.35	0.80~0.86	0.24~0.61	4/4 (100%)	32/32 (100%)
Bodily Pain	0.77~1.06	0.69~0.76	0.32~0.68	2/2 (100%)	16/16 (100%)
General Perception of Health	0.95~1.23	0.53~0.79	0.22~0.62	5/5 (100%)	40/40 (100%)
Vitality	0.90~1.23	0.50~0.68	0.18~0.67	4/4 (100%)	32/32 (100%)
Social Functioning	0.87~0.76	0.58	0.28~0.62	2/2 (100%)	14/16 (87.5%)
Role-Emotional	0.31~0.34	0.69~0.83	0.21~0.56	3/3 (100%)	24/24 (100%)
Mental Health	0.90~1.12	0.47~0.70	0.02~0.66	5/5 (100%)	39/40 (97.5%)

a. Item correlations with own scale ≥0.40.

b. Item correlations with own scale significantly greater than those with other scales.

Table 8. Estimates of internal consistency for the SF-36 questionnaire in a Chinese general population.

Scale	Physical Functioning	Role-Physical	Bodily Pain	General Health Perception	Vitality	Social Functioning	Role-Emotional	Mental Health
Overall	0.93	0.93	0.84	0.86	0.79	0.76	0.88	0.81
Age								
18-34	0.90	0.84	0.77	0.86	0.77	0.76	0.87	0.80
35-49	0.88	0.93	0.78	0.82	0.78	0.79	0.85	0.81
50-65	0.91	0.97	0.84	0.84	0.83	0.75	0.88	0.87
>65	0.94	0.94	0.86	0.88	0.80	0.68	0.91	0.80
Gender								
female	0.94	0.90	0.84	0.84	0.78	0.83	0.88	0.82
male	0.93	0.95	0.81	0.89	0.80	0.67	0.87	0.80
Education								
≤9	0.93	0.96	0.82	0.87	0.79	0.78	0.86	0.84
9-12	0.92	0.91	0.80	0.81	0.69	0.78	0.83	0.63
>12	0.87	0.90	0.77	0.85	0.79	0.73	0.88	0.84
Income								
<NT$3,5000	0.95	0.95	0.85	0.87	0.82	0.81	0.89	0.82
≥NT$3,5000	0.92	0.92	0.81	0.86	0.77	0.74	0.88	0.81
Chronic disease								
0	0.92	0.89	0.77	0.84	0.76	0.74	0.85	0.80
>0	0.95	0.96	0.90	0.87	0.82	0.74	0.94	0.83
Having illness								
0	0.92	0.92	0.76	0.85	0.76	0.71	0.88	0.80
>0	0.95	0.90	0.93	0.92	0.91	0.84	0.81	0.86

Table 9. The factorial structure and factor loadings of the SF-36 questionnaire in a Chinese general population, Taichung, Taiwan.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Name of Factor	Physical Functioning	Mental Health and Vitality I	Role-Physical	General Health Perception	Mental Health and Vitality II	Role-Emotional	Pain and Physical Functioning
	PF2 (0.59)	MH1 (0.76)	RP1 (0.80)	GH1 (0.46)	VT1 (0.70)	RE1 (0.81)	PF1 (0.67)
	PF3 (0.71)	MH2 (0.63)	RP2 (0.77)	GH2 (0.77)	VT2 (0.72)	RE2 (0.82)	PF2 (0.59)
	PF4 (0.71)	MH4 (0.71)	RP3 (0.81)	GH3 (0.75)	MH3 (0.78)	RE3 (0.74)	BP1 (0.59)
	PF5 (0.83)	VT3 (0.82)	RP4 (0.77)	GH4 (0.67)	MH5 (0.83)		BP2 (0.47)
	PF6 (0.69)	VT4 (0.69)		GH5 (0.73)			
	PF7 (0.77)						
	PF8 (0.87)						
	PF9 (0.86)						
	PF10 (0.74)						
	SF1 (0.54)						
Eigenvalue	14.02	3.90	2.33	1.63	1.24	1.11	1.01
Proportion of variance explained	40.05%	11.14%	6.65%	4.67%	3.53%	3.16%	2.87%

a. Items in bold belong to scale of vitality.

Table 10. Hypothesized associations between scales of the MOS 36-Item Short Form Health Survey (Chinese version) and postulated physical and mental components of health, compared to results of a factorial analysis in a Chinese general population, Taichung, Taiwan.

	Hypothesized Assoication ∞		Factorial Analysis: Rotated principal components			Relative validity *f*	
	Physical	Mental	Correlations With Physical	Mental	Variance explained	Physical	Mental
Physical functioning	+	_	0.76	0.59	0.70	1.00	0.02
Role-physical	+	_	0.85	0.65	0.64	0.89	0.05
Bodily pain	+	_	0.76	0.65	0.66	0.87	0.08
General health	*	*	0.73	0.84	0.68	0.36	0.52
Vitality	*	*	0.68	0.89	0.85	0.18	0.87
Social functioning	*	+	0.75	0.70	0.71	0.83	0.18
Role-emotional	_	+	0.73	0.63	0.45	0.54	0.10
Mental	_	+	0.52	0.80	0.86	0.03	1.00

∞ Same as in McHorney *et al*. [?]: + strong association ($r \geq 0.7$), * moderate association ($0.3<r<0.7$), - week association ($r \leq 0.3$)
f Ratio of variance in principal component explained by a given scale to variance explained by best scale.

Table 11. Means (standard deviations) of the SF-36 scores for specific subgroups in a Chinese general population.

Scale	Physical Functioning	Role-Physical	Bodily Pain	General Health Perception	Vitality	Social Functioning	Role-Emotional	Mental Health
Age (in years)								
18-34 (n=154)	95.8 (11.7)*	92.0 (22.3)*	83.8 (13.9)*	72.8 (20.4)*	69.8 (15.3)*	91.6 (15.9)*	89.0 (28.0)†	72.4 (14.6)
35-49 (n=144)	95.5 (11.3)	90.7 (26.4)	81.6 (15.4)	70.1 (19.9)	68.6 (16.0)	90.8 (15.8)	90.1 (26.3)	73.6 (14.9)
50-65 (n=67)	85.8 (18.2)	81.0 (37.7)	74.6 (18.4)	63.0 (19.9)	66.3 (17.4)	87.3 (16.5)	92.5 (23.8)	72.1 (16.2)
>65 (n=60)	66.2 (29.0)	70.4 (42.1)	67.6 (21.3)	47.9 (22.7)	54.3 (17.3)	75.6 (26.0)	78.9 (37.8)	69.7 (15.1)
Gender								
female (n = 195)	89.2 (20.1)	84.5 (31.9)	77.8 (18.3)	64.5 (21.2)	64.5 (16.0)	87.5 (21.0)	86.1 (31.2)	71.1 (14.9)
male (n = 231)	90.6 (18.3)	88.6 (29.6)	80.5 (16.2)	68.7 (22.7)	68.5 (17.6)	89.1 (16.1)	90.5 (26.2)	73.5 (15.1)
Education								
≤9 (n = 103)	84.8 (21.6)*	81.1 (37.1)†	75.5 (19.3)‡	61.9 (22.1) ‡	61.7 (17.1)*	84.6 (20.1) ‡	86.7 (30.0)	70.1 (15.3)
9-12 (n = 71)	92.9 (15.4)	89.6 (27.1)	81.2 (14.5)	68.8 (18.9)	71.1 (13.8)	91.6 (15.2)	94.8 (19.4)	73.5 (11.3)
>12 (n = 220)	95.0 (11.4)	90.4 (25.8)	82.2 (14.6)	71.1 (20.6)	69.1 (16.1)	91.4 (14.7)	87.8 (29.6)	73.5 (15.9)
Income								
<NT$3,5000 (n=125)	84.7 (24.2) ‡	80.8 (36.6) †	77.1 (18.8)	60.3 (23.6)*	63.1 (18.2) ‡	85.4 (21.0)	87.4 (30.2)	70.1 (15.4)
≥NT$3,5000 (n=283)	92.1 (16.1)	88.9 (28.2)	79.9 (16.7)	69.4 (20.9)	67.8 (16.2)	89.5 (17.4)	88.8 (28.2)	73.1 (15.0)
Chronic disease								
No (n = 373)	91.4 (17.2)*	89.4 (27.7)*	81.0 (14.9)*	68.3 (21.1)*	67.8 (16.1)*	89.9 (16.1)*	90.6 (26.2) ‡	73.0 (14.8)*
Yes (n = 53)	74.1 (30.4)	57.0 (44.1)	61.3 (26.2)	48.7 (26.1)	53.3 (20.2)	73.0 (29.1)	65.6 (41.0)	65.3 (16.2)
Having illness								
No (n = 393)	92.1 (16.0) ‡	90.7 (25.3)*	81.1 (14.6)*	70.0 (19.8)*	69.0 (15.2)*	90.8 (15.1) ‡	90.6 (25.5) ‡	73.5 (14.4) †
Yes (n = 32)	74.7 (30.1)	59.4 (47.1)	66.8 (26.5)	44.1 (24.1)	50.2 (19.4)	71.9 (28.7)	73.6 (42.0)	64.8 (17.3)

* $p<0.001$; ‡ $p<0.01$; † $p<0.05$

Table 12. The Chinese-version short form 36 scores and their relative validity in 2 groups* of general population defined by whether having a chronic disease.

Scale	Mean score in group 1	Difference between group 1 and group 2	F-statistic	Relative Validity
Physical functioning	92.1	17.4	41.8	0.56
Role-physical	90.7	31.3	54.1	0.72
Bodily pain	81.1	14.3	34.4	0.46
General health	70.0	25.9	75.0	1.00
Vitality	69.0	18.8	65.9	0.88
Social functioning	90.8	18.9	54.2	0.72
Role-emotional	90.6	17.0	17.0	0.23
Mental health	73.5	8.7	16.1	0.21

* Group 1: no chronic disease; Group 2: having one or more chronic diseases

Table 13. Sample size for each subgroup among outpatients

	n	%
Age		
18-34	140	49.3
35-49	98	34.5
>50	46	16.2
Gender		
Male	133	47.0
Female	150	53.0
Education[a]		
<9	33	14.3
9-12	27	11.7
>12	170	73.9
Number of Life Event[a]		
≤ 1	55	19.4
2-5	125	44.2
≥6	103	36.4
Taking Medicine[a]		
No	143	50.9
Yes	138	49.1
Having Chronic Diseases[a]		
No	163	58.2
Yes	117	41.8

[a]Numbers do not equal to 426 due to missing data.

Table 14. Item Frequency Distributions and Number and Percent Missing: A Outpatient Sample of Primary Care Setting.

	Item Frequency Distribution						Missing	
Item	1	2	3	4	5	6	f	%
PF1	27	97	161	-[b]	-	-	3	1.0
PF2	13	40	229	-	-	-	6	2.1
PF3	4	41	240	-	-	-	3	1.0
PF4	7	62	213	-	-	-	6	2.1
PF5	3	20	261	-	-	-	4	1.4
PF6	8	54	223	-	-	-	3	1.0
PF7	5	44	237	-	-	-	2	0.7
PF8	3	27	254	-	-	-	4	1.4
PF9	1	15	266	-	-	-	6	2.1
PF10	3	5	277	-	-	-	3	1.0
RP1	79	205	-	-	-	-	4	1.4
RP2	87	197	-	-	-	-	4	1.4
RP3	68	214	-	-	-	-	6	2.1
RP4	74	209	-	-	-	-	5	1.7
BP1	99	57	85	32	11	2	2	0.7
BP2	136	104	31	15	0	0	2	0.7
GH1	2	37	70	121	56	-	2	0.7
GH2	30	49	62	87	56	-	4	1.4
GH3	47	115	74	41	5	-	6	2.1
GH4	25	49	106	60	42	-	6	2.1
GH5	22	95	98	54	13	-	6	2.1
VT1	23	70	54	94	37	3	7	2.4
VT2	23	50	72	86	50	3	4	1.4
VT3	9	13	40	119	86	16	5	1.7
VT4	16	21	56	144	44	4	3	1.0
SF1	165	98	14	7	2	-	2	0.7
SF2	5	19	101	110	51	-	2	0.7
RE1	89	194	-	-	-	-	5	1.7
RE2	94	187	-	-	-	-	7	2.4
RE3	93	190	-	-	-	-	5	1.7
MH1	35	28	47	113	53	7	5	1.7
MH2	8	11	26	93	118	26	6	2.1
MH3	26	79	66	89	24	1	3	1.0
MH4	7	10	33	114	101	20	3	1.0
MH5	30	87	64	78	16	7	6	2.1
TRAN	8	23	107	117	33	-	0	0.0

[a]N=288.

[b]Responses not possible

Table 15. Percent of patients for whom scale scores were computable.

	PF	RP	BP	GH	VT	SF	RE	MH
Total	99.0	98.3	98.6	97.9	99.0	98.6	97.6	99.3
Age (years)								
18-34	100.0	99.3	99.3	100.0	99.3	100.0	97.9	99.3
35-49	96.9	95.9	96.9	93.9	98.0	95.9	95.9	99.0
>50	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Gender								
Male	99.2	97.7	99.2	99.2	98.5	98.5	97.7	99.2
Female	98.7	98.7	98.0	96.7	99.3	98.7	97.4	99.3
Education								
<9	97.0	97.0	97.0	93.9	97.0	93.9	97.0	97.0
9-12	96.3	100.0	100.0	96.3	100.0	100.0	100.0	100.0
>12	99.4	97.7	98.8	98.2	99.4	99.4	97.1	100
Number of Life Events								
≤ 1	100.0	100.0	98.2	98.2	100.0	98.2	98.2	100.0
2-5	97.2	100.0	99.2	98.4	100.0	99.2	100.0	100.0
≥6	98.2	95.1	98.1	97.1	97.1	98.1	94.2	98.1
Taking Medicine								
No	99.3	99.3	98.6	97.9	98.6	98.6	97.9	99.3
Yes	98.6	97.1	98.6	97.8	99.3	98.6	97.8	99.3
Having Chronic Disease								
No	98.8	98.2	98.2	96.9	98.2	98.2	97.5	98.8
Yes	99.1	98.3	99.1	99.1	98.3	99.1	97.4	99.1

Table 16. Description of 8 scales of the MOS 36-item Short Form Health Survey (Chinese version) submitted to a sample of 288 outpatients in primary care setting.

	Physical functioning	Role-physical	Bodily pain	General health	Vitality	Social functioning	Role-emotional	Mental health
Number of items	10	4	2	5	4	2	3	5
Percentage non-respondents	1.0	1.7	1.4	2.1	1.0	1.4	2.4	0.7
Mean score	89.4	73.1	74.1	53.3	55.5	76.3	67.9	59.7
Percentiles; 25th	85	50	62	40	45	62.5	33.3	48
Percentiles; 50th	95	100	74	55	55	75	100	60
Percentiles; 75th	100	100	100	67	70	87.5	100	72
Percentage at ceiling	49.8	59.5	31.8	0.4	0.7	15.7	57.0	0.4
Percentage at floor	0	13.3	0	0	0.7	0.4	20.9	0.4

Table 17. Item means and standard deviations and correlations[a] between SF-36 items and hypothesized scales in a sample of 288 outpatients in primary care setting.

Item	Mean	SD	PF	RP	BP	GH	VT	SF	RE	MH	TRAN
PF1	2.47	0.66	.63*	.51	.33	.37	.29	.37	.31	.21	.20
PF2	2.76	0.52	.64*	.37	.30	.23	.21	.21	.21	.19	.17
PF3	2.83	0.41	.61*	.35	.30	.25	.19	.18	.23	.16	.18
PF4	2.72	0.50	.67*	.51	.33	.39	.25	.23	.29	.17	.11
PF5	2.91	0.32	.57*	.29	.13	.13	.10	.19	.17	.13	.08
PF6	2.75	0.49	.51*	.34	.32	.21	.22	.24	.22	.18	.07
PF7	2.81	0.43	.63*	.44	.21	.24	.28	.27	.36	.21	.09
PF8	2.89	0.35	.64*	.41	.21	.23	.23	.28	.30	.17	.08
PF9	2.94	0.25	.59*	.31	.14	.11	.12	.14	.22	.06	.01
PF10	2.96	0.24	.39*	.26	.04	.07	.12	.16	.22	.07	.01
RP1	1.73	0.45	.38	.70*	.34	.42	.33	.35	.41	.24	.29
RP2	1.69	0.46	.42	.77*	.36	.43	.39	.43	.51	.31	.20
RP3	1.76	0.43	.46	.73*	.35	.38	.31	.34	.40	.28	.17
RP4	1.74	0.44	.56	.66*	.41	.45	.44	.41	.45	.28	.28
BP1	4.84	1.17	.30	.38	.79*	.41	.31	.33	.24	.20	.29
BP2	4.58	1.19	.41	.49	.79*	.43	.41	.44	.36	.32	.32
GH1	2.48	1.13	.25	.44	.34	.51*	.45	.30	.27	.31	.28
GH2	3.31	1.26	.23	.42	.42	.56*	.40	.33	.30	.29	.15
GH3	3.56	0.99	.21	.24	.21	.45*	.33	.30	.24	.34	.12
GH4	3.16	1.15	.27	.36	.27	.47*	.37	.23	.24	.20	.44
GH5	3.22	0.99	.31	.37	.37	.67*	.49	.33	.31	.30	.28
VT1	3.78	1.22	.21	.37	.32	.51	.67*	.44	.40	.55	.28
VT2	3.65	1.22	.22	.35	.28	.48	.60*	.37	.41	.52	.22
VT3	4.08	1.07	.34	.32	.23	.32	.55*	.39	.36	.58	.30
VT4	3.66	1.05	.30	.35	.29	.43	.61*	.51	.44	.56	.34
SF1	4.46	0.76	.27	.39	.33	.32	.43	.44*	.52	.43	.27
SF2	3.64	0.91	.29	.38	.34	.37	.45	.44*	.45	.41	.19
RE1	1.69	0.46	.29	.50	.25	.33	.43	.48	.78*	.40	.23
RE2	1.67	0.47	.28	.46	.23	.34	.44	.48	.75*	.39	.18
RE3	1.67	0.47	.36	.41	.33	.31	.45	.53	.68*	.42	.24
MH1	3.49	1.31	.19	.19	.15	.22	.25	.20	.18	.27*	.13
MH2	4.35	1.08	.20	.25	.06	.20	.49	.39	.37	.60*	.22
MH3	3.97	1.15	.13	.29	.24	.31	.55	.43	.43	.44*	.13
MH4	4.23	1.03	.18	.22	.16	.26	.54	.43	.34	.60*	.30
MH5	4.06	1.21	.11	.23	.20	.29	.59	.37	.32	.50*	.11
TRAN	3.50	0.9	.17	.28	.40	.35	.34	.27	.24	.25	-

a: corrected for overlap

Table 18. Scaling properties of the 8 scales of the SF-36 in a an outpatient sample of a primary care setting.

Scale	Items' standard deviation (range)	Correlations of items with own scale (range)	Correlations of items with other scales (range)	Convergent validity[a]	Discriminant validity[b]
Physical Functioning	0.24~0.66	0.39~0.67	0.11~0.56	9/10 (90%)	80/80 (100%)
Role-Physical	0.43~0.46	0.66~0.77	0.19~0.51	4/4 (100%)	32/32 (100%)
Bodily Pain	1.17~1.19	0.79	0.04~0.42	2/2 (100%)	16/16 (100%)
General Perception of Health	0.99~1.26	0.45~0.67	0.07~0.51	5/5 (100%)	40/40 (100%)
Vitality	1.05~1.22	0.55~0.67	0.10~0.59	4/4 (100%)	31/32 (96.9%)
Social Functioning	0.76~0.91	0.44	0.14~0.48	2/2 (100%)	11/16 (68.8%)
Role-Emotional	0.46~0.47	0.68~0.78	0.17~0.52	3/3 (100%)	24/24 (100%)
Mental Health	1.03~1.31	0.27~0.60	0.06~0.58	4/5 (80%)	36/40 (90.0%)

a. Item correlations with own scale ≥0.40.

b. Item correlations with own scale significantly greater than those with other scales.

Table 19. Estimates of internal consistency for the SF-36 questionnaire in an outpatient sample of a primary care setting.

Scale	Physical Functioning	Role-Physical	Bodily Pain	General Health Perception	Vitality	Social Functioning	Role-Emotional	Mental Health
Overall	0.86	0.87	0.61	0.76	0.80	0.89	0.85	0.72
Age								
18-34	0.88	0.81	0.57	0.77	0.83	0.88	0.80	0.74
35-49	0.80	0.89	0.61	0.74	0.78	0.86	0.88	0.71
≥50	0.85	0.91	0.69	0.77	0.76	0.92	0.94	0.69
Gender								
female	0.85	0.83	0.62	0.78	0.82	0.89	0.84	0.78
male	0.86	0.88	0.59	0.73	0.78	0.88	0.85	0.65
Education								
≤9	0.85	0.86	0.67	0.67	0.75	0.94	0.87	0.40
9-12	0.83	0.93	0.60	0.71	0.77	0.83	0.89	0.71
>12	0.85	0.83	0.58	0.77	0.84	0.89	0.84	0.79
Number of Life Events								
≤ 1	0.89	0.86	0.32	0.78	0.67	0.85	0.85	0.47
2-5	0.85	0.90	0.68	0.74	0.78	0.88	0.84	0.74
≥6	0.84	0.83	0.57	0.78	0.86	0.90	0.85	0.76
Taking Medicine								
No	0.88	0.86	0.45	0.72	0.78	0.89	0.84	0.73
Yes	0.84	0.86	0.68	0.72	0.81	0.88	0.87	0.69
Having Chronic Disease								
No	0.88	0.86	0.54	0.75	0.78	0.88	0.85	0.73
Yes	0.84	0.86	0.71	0.73	0.83	0.89	0.86	0.72

Table 20. The factorial structure and factor loadings of the SF-36 questionnaire in an outpatient sample of a primary care setting, Taichung, Taiwan.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Name of Factor	Physical Functioning I	Mental Health and Vitality I	Physical Functioning II	Mental Health and Vitality II	Role-Emotional And Social Functioning	Role-Physical	General Perception of Health	Bodily Pain
	PF1 (0.74)	MH1 (0.58)	PF5 (0.73)	MH3 (0.73)	RE1 (0.77)	RP1 (0.75)	GH1 (0.55)	BP1 (0.87)
	PF2 (0.78)	MH2 (0.75)	PF8 (0.64)	MH5 (0.75)	RE2 (0.75)	RP2 (0.74)	GH3 (0.67)	BP2 (0.83)
	PF3 (0.63)	MH4 (0.78)	PF9 (0.88)	VT1 (0.71)	RE3 (0.69)	RP3 (0.70)	GH4 (0.53)	
	PF4 (0.61)	VT3 (0.76)	PF10 (0.82)	VT2 (0.78)	SF1 (0.56)	RP4 (0.54)	GH5 (0.78)	
	PF7 (0.61)	VT4 (0.69)						
Eigenvalue	10.61	3.78	2.29	1.69	1.45	1.38	1.16	1.12
Proportion of variance explained	30.30%	10.80%	6.55%	4.82%	4.14%	3.95%	3.32%	3.21%

Table 21. Hypothesized associations between scales of the MOS 36-Item Short Form Health Survey (Chinese version) and postulated physical and mental components of health, compared to results of a factorial analysis in an outpatient sample of a primary care setting.

	Hypothesized Assoication ∞		Factorial Analysis: Rotated principal components			Relative validity *f*	
	Physical	Mental	Correlations With Physical	Mental	Variance explained	Physical	Mental
Physical functioning	+	_	0.68	0.49	0.62	1.00	0.02
Role-physical	+	_	0.88	0.71	0.68	0.95	0.14
Bodily pain	+	_	0.67	0.48	0.53	0.83	0.04
General health	*	*	0.69	0.64	0.52	0.56	0.24
Vitality	*	*	0.63	0.82	0.76	0.10	0.89
Social functioning	*	+	0.63	0.70	0.63	0.26	0.61
Role-emotional	_	+	0.69	0.84	0.56	0.21	0.57
Mental	_	+	0.49	0.74	0.78	0.01	1.00

∞ Same as in McHorney *et al*. [?]: + strong association (r ≥0.7), * moderate association (0.3<r<0.7), - week association (r≤ 0.3)
f Ratio of variance in principal component explained by a given scale to variance explained by best scale.

Table 22. Means (standard deviations) of the SF-36 scores for specific subgroups in an outpatient sample of a primary care setting.

Scale	Physical Functioning	Role-Physical	Bodily Pain	General Health Perception	Vitality	Social Functioning	Role-Emotional	Mental Health
Age (in years)								
18-34 (n=140)	92.8 (14.3)*	79.3 (32.3) ‡	76.8 (21.7)	55.2 (18.8)	55.2 (18.5)	75.1 (16.6)	65.2 (40.2)	59.7 (16.1)
35-49 (n=98)	89.9 (13.5)	71.8 (39.0)	73.2 (21.1)	52.6 (19.7)	56.8 (17.1)	77.5 (17.6)	73.8 (40.0)	60.5 (15.8)
≥50 (n=46)	79.1 (19.8)	58.2 (44.4)	69.6 (24.7)	50.3 (22.8)	54.0 (19.9)	76.9 (21.7)	67.4 (44.7)	58.8 (19.2)
Gender								
male (n = 152)	92.7 (13.6) ‡	79.6 (33.2) ‡	74.5 (22.5)	57.5 (19.6) ‡	57.8 (19.0)	77.2 (18.9)	76.7 (37.0) ‡	61.4 (17.5)
female (n = 133)	86.3 (17.5)	67.2 (40.3)	73.9 (22.2)	49.7 (19.6)	53.3 (17.5)	75.4 (17.0)	60.8 (43.1)	58.2 (15.5)
Education								
≤9 (n = 33)	78.9 (18.8) *	53.1 (41.5) ‡	72.3 (25.4)	47.3 (20.4)	54.2 (21.5)	76.2 (20.5)	71.9 (40.7)	56.9 (15.8)
9-12 (n = 27)	86.3 (18.3)	70.4 (42.2)	71.6 (22.3)	52.8 (19.5)	55.7 (16.5)	72.7 (23.8)	66.7 (43.4)	62.2 (14.8)
>12 (n = 171)	92.7 (13.0)	78.0 (34.4)	75.7 (21.7)	55.2 (19.2)	55.6 (18.5)	76.0 (16.3)	67.3 (40.9)	60.2 (17.1)
Number of Life Events								
≤ 1 (n=55)	85.9 (19.0)	75.0 (37.6)	76.3 (21.5)	52.7 (21.0)	60.3 (16.1)*	82.9 (32.9)*	83.3 (32.9)*	63.2 (14.2) ‡
2-5 (n=125)	90.8 (14.8)	74.4 (38.0)	76.5 (20.8)	55.8 (19.1)	57.8 (17.8)	78.3 (17.7)	73.3 (38.6)	61.6 (16.0)
≥6 (n=103)	90.0 (14.9)	70.7 (37.0)	70.6 (23.5)	51.0 (20.0)	50.2 (18.9)	70.2 (17.8)	53.6 (43.7)	55.8 (17.5)
Taking Medicine								
No (n = 143)	91.9 (15.3) ‡	82.9 (32.3)*	77.9 (21.4) ‡	61.3 (16.9)*	59.8 (16.0)*	78.7 (14.6) †	73.1 (38.5)	62.8 (15.5) ‡
Yes (n = 138)	86.9 (16.0)	63.1 (40.2)	70.8 (22.4)	45.5 (19.6)	51.4 (19.6)	73.6 (20.5)	63.7 (43.0)	57.0 (17.1)
Having Chronic Disease								
No (n = 163)	91.1 (15.7) †	78.9 (34.4) ‡	76.8 (21.7) †	57.5 (19.3)*	57.4 (17.4) †	76.5 (17.5)	69.8 (40.4)	60.8 (16.2)
Yes (n = 117)	87.2 (15.8)	64.6 (40.3)	70.9 (22.1)	47.5 (18.9)	52.8 (19.3)	76.0 (18.4)	66.1 (42.1)	58.6 (16.9)

* $p<0.001$; ‡ $p<0.01$; † $p<0.05$

Table 23. The Chinese-version short form 36 scores and their relative validity in 3 groups* of an outpatient sample of a primary care setting defined by their age.

Scale	Mean score in group 1	Difference between group 1 and group 2	Difference between group 1 and group 3	F-statistic	Relative Validity
Physical functioning	92.8	2.9	13.7	14.18	1.00
Role-physical	79.3	7.5	21.1	5.83	0.17
Bodily pain	76.8	3.6	7.2	2.07	0.02
General health	55.2	2.6	4.9	1.24	0.01
Vitality	55.2	-1.6	1.2	0.42	0.00
Social functioning	75.1	-2.4	-1.8	0.56	0.00
Role-emotional	65.2	-8.6	-2.2	1.24	0.01
Mental health	59.7	-0.8	0.9	0.17	0.00

Group 1: 18-34 years old; Group 2: 35-49 years old; Group 3: ≥ 50 years old.

Table 24. The Chinese-version short form 36 scores and their relative validity in 3 groups* of an outpatient sample of a primary care setting defined by their number of life events.

Scale	Mean score in group 1	Difference between group 1 and group 2	Difference between group 1 and group 3	F-statistic	Relative Validity
Physical functioning	85.90	-4.88	4.09	1.88	0.03
Role-physical	75.00	-0.6	-4.34	0.35	0.00
Bodily pain	76.33	-0.12	5.78	2.30	0.04
General health	52.67	-3.17	-1.63	1.68	0.02
Vitality	60.27	-2.43	10.07	7.43	0.41
Social functioning	82.87	4.54	12.7	11.16	0.92
Role-emotional	83.33	10.00	29.72	11.66	1.00
Mental health	63.20	1.54	7.4	5.09	0.19

Group 1: ≤ 1 life events; Group 2: 2-5 life events; Group 3: ≥6 life events

Table 25. The estimated parameters of sodiodemographic factors, chronic conditions and inpatient sample among a combined sample of outpatients and residents of Taichung, Taiwan.

	Estimated Parameters (Standatd Error)							
Variable	PF	RP	BP	GH	VT	SF	RE	MH
Intercept	90.9 (2.0)*	86.0 (4.3) *	80.2 (2.5)*	67.6 (2.7) *	62.9 (2.3) *	86.7 (2.4) *	88.8 (4.6) *	68.1 (2.1) *
Age								
Age1	0.8 (1.4)	0.1 (3.0)	-1.5 (1.7)	0.1 (1.9)	1.6 (1.6)	1.7 (1.6)	2.6 (3.2)	2.3 (1.5)
Age2	-11.2 (1.8) *	-8.2 (4.0) †	-6.3 (2.3) ‡	-7.0 (2.5) ‡	-0.7 (2.1)	0.5 (2.2)	2.1 (4.3)	2.5 (2.0)
Gender								
Female	0.3 (1.2)	0.6 (2.6)	2.9 (1.5)	2.0 (1.6)	2.5 (1.4)	-0.4 (1.4)	-2.2 (2.7)	1.0 (1.3)
Education								
Edu1	3.34 (2.0)	5.7 (4.4)	1.1 (2.5)	2.8 (2.7)	6.4 (2.3) ‡	3.9 (2.4)	3.9 (4.7)	4.4 (2.2) †
Edu2	5.8 (1.7) *	7.9 (3.7) †	2.4 (2.1)	4.1 (2.3)	4.8 (2.0) †	4.5 (2.0) †	-0.2 (4.0)	4.1 (1.8) †
Having chronic condition	-3.9 (1.5) †	-17.3 (3.4) *	-8.4 (1.9) *	-13.6 (2.1) *	-9.2 (1.8) *	-6.8 (1.8) *	-8.1 (3.6) †	-5.5 (1.7) *
Outpatient	-2.9 (1.3) †	-10.8 (2.9) *	-3.7 (1.7) †	-11.5 (1.8) *	-9.6 (1.5) *	-12.4 (1.6) *	-17.5 (3.1) *	-10.8 (1.4) *
R^2	17.31%	12.28%	8.78%	21.16%	16.59%	16.80%	9.24%	15.28%

* $p<0.001$; ‡ $p<0.01$; † $p<0.05$

Table 26. Estimates of internal consistency for the SF-36 questionnaire across different populations

Author	PF	RP	BP	GH	VT	SF	RE	MH	Sample
Li *et al*.	0.93	0.93	0.84	0.86	0.79	0.76	0.88	0.81	Random Sample of Chinese Population, n=426
Li *et al*.	0.86	0.87	0.61	0.76	0.80	0.89	0.85	0.72	Primary care sample, n=288
McHorney, *et al*.[3]	0.94	0.89	0.88	0.83	0.87	0.63	0.81	0.82	Random Sample of U.S. Population, n=1,692
McHorney, *et al*.[9]	0.93	0.84	0.82	0.78	0.87	0.85	0.83	0.90	Patients with Chronic Conditions, n=3,445
Perneger, *et al*.[10]	0.92	0.84	0.78	0.77	0.81	0.77	0.76	0.86	Random Sample of Managed Care Organization members, n=1,007
Garratt, *et al.*[11]	0.92	0.89	0.86	0.83	0.86	0.80	0.86	0.86	Patients with Chronic Conditions and General Population Sample, n=1,310 and 542, respectively
Sullivan, *et al.*[12]	0.91	0.88	0.93	0.84	0.85	0.83	0.79	0.87	A non-random sample of Sweden general population, n=8,930
Brazier, *et al.*[13]	0.93	0.96	0.85	0.95	0.96	0.73	0.96	0.95	General Practice Patients, n=1,582
Jenkinson, *et al*.[14]	0.90	0.88	0.82	0.80	0.85	0.76	0.80	0.83	Random Sample of U.K. Population, n=13,042

PF: Physical Functioning; RP: Role Physical; BP: Bodily Pain; GH: General Health; VT: Vitality; SF: Social Functioning; Re: Role Emotional; MH: Mental Health.

Chinese population: A random sample of 426 subjects of the general population with their age ranged from 18 to 65.

U.S.A population[9]: A sample of 3445 members of health maintenance organization with their age range from 18 to 75.

Switzerland population[10]: A sample of 1007 members of a managed care organization with their age ranged from 18 to 44.

Scotland population[11]: A sample of 542 subjects of the general population with their age ranged from 18 to 91.

Sweden population[12]: A sample of 8930 subjects of the general population with their age ranged from 15 to 93.