Title	Is the standard SF-12 Health Survey valid and equivalent for a Chinese population?
Author(s)	Lam, CLK; Tse, EYY; Gandek, B
Citation	Quality Of Life Research, 2005, v. 14 n. 2, p. 539-547
Issued Date	2005
URL	http://hdl.handle.net/10722/48616
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Running Head: The SF-12 for the Chinese

Is the Standard SF-12 Health Survey Valid and Equivalent for a Chinese Population?

Cindy L.K. Lam, MBBS, MD, FRCGP, FHKAM (Family Medicine)

Associate Professor, Family Medicine Unit, the University of Hong Kong

Eileen Y. Y. Tse, MBChB, FRCGP, FHKAM (Family Medicine)

Assistant Professor, Family Medicine Unit, the University of Hong Kong

Barbara Gandek, MS

Senior Project Director, IQOLA Project, Health Assessment Lab, Boston, USA

Correspondence: Family Medicine Unit, The University of Hong Kong,
3rd Floor, Ap Lei Chau Clinic, 161 Main Street,
Ap Lei Chau, Hong Kong SAR.
Fax: 852 28147475; E mail: clklam@hku.hk

Is the *Standard* SF-12 Health Survey Valid and Equivalent *for a Chinese Population*?

Abstract

Introduction: Chinese is the world's largest ethnic group but few health-related quality of life (HRQOL) measures have been tested on them. The aim of this study was to *determine if the standard* SF-12 was valid and equivalent for a Chinese population. **Methods**: The SF-36 data of 2410 Chinese adults randomly selected from the general population of Hong Kong (HK) were analysed. The Chinese (HK) specific SF-12 items and scoring algorithm were derived from the Hong Kong Chinese population data by multiple regressions. The SF-36 PCS and MCS scores were used as criteria to assess the content and criterion validity of the SF-12. The standard and Chinese (HK) specific SF-12 PCS and MCS scores were compared for equivalence. **Results**: The *standard SF-12 explained 82% and 89% of the variance of the SF-36 PCS and MCS scores*, respectively, and the effect size differences between the SF-36 and SF-12 scores were less than 0.3. Six of the Chinese (HK) specific SF-12 items were different from those of the standard SF-12, but the effect size differences between the Chinese (HK) specific and standard SF-12 scores were generally less than 0.3. **Conclusions**: The standard SF-12 was valid and equivalent for the Chinese, which would enable more Chinese to be included in clinical trials that measure HRQOL.

Keyword: Health-related quality of life, SF-12, Chinese, Validity, Equivalence

Introduction

Chinese make up nearly a quarter of the world's population. They should be included in global and international clinical trials but this is often not possible in studies that measure health-related quality of life (HRQOL) because of language and cultural barriers. Most HRQOL measures are in English and originate from the Western culture, so they need to be translated and validated before they can be applied to the Chinese. The Chinese (Hong Kong) translation of the MOS 36-item Short Form Health Survey (SF-36) and its Physical and Mental Health Summary (PCS and MCS) Scales have been shown to be valid and equivalent for the Chinese [1-5], but the length of the SF-36 limits its acceptability in some clinical trials that need to measure a number of outcomes. The *standard* SF-12 Health Survey (SF-12), an abbreviated form of the SF-36 that *yeilds* the PCS and MCS scores, is becoming a popular HRQOL measure in clinical trials because it can be completed in a few minutes [6, 7].

The items and scoring algorithm of the standard SF-12 were derived from data of a US general population survey in 1990 [6-8]. The standard SF-12 PCS and MCS scores are norm-based on the US general population whose mean is 50 and standard deviation (SD) is 10 [6, 9]. The 12 items include two from each of the Physical Functioning, Role-Physical, Role-Emotional and Mental Health scales and one item from each of the Bodily Pain, General Health, Vitality and Social Functioning scales of the SF-36. The items were selected by multiple regressions in order to explain the largest proportion of the total variance in the SF-36 PCS and MCS scores. The response to each item is weighted separately by the PCS and MCS regression

coefficient and then summated to give the standard SF-12 PCS and MCS scores, respectively.

A small number and weighting of items may make a HRQOL measure more culture-sensitive [10, 11]. All previous studies on the validity and equivalence of the standard SF-12 were carried out in Caucacian populations [8, 12]. There was very little data from any Chinese or Asian population whose cultures are quite different from those of the West. *The rank orders by item mean of three (PF9, GH3 and RE3) SF-36 items were found to be different between the HK Chinese and US populations [1]. Although the differential item functioning (DIF) of a few items did not affect the validity of the SF-36 scales that summated all the items without weighting [13], they may have an effect on the validity and equivalence of the much shorter standard SF-12.*

The aim of this study was to *determine if* the *standard* SF-12 was valid and equivalent for the Chinese population of Hong Kong, *or whether a Chinese (HK) specific SF-12 was needed.* The *standard* SF-12 is valid if it really measures the SF-36 PCS and MCS scores, which are what it purports to measure. The selected items should be representative and adequate in explaining the SF-36 PCS and MCS scores (content validity), and the SF-12 should give *similar* PCS and MCS scores as the SF-36 (criterion validity). *The standard SF-12 is equivalent if no more than three of the 12 items selected specifically from the Chinese (HK) population were different from those of the standard SF-12, as that found in other countries (item equivalence)[8];* and if there is no important difference between the results of the Chinese (HK) specific and standard algorithms (measurement equivalence)[8, 14, 15].

Methods

Data of 2410 Chinese adults randomly selected from the general population of Hong Kong that were collected in the Chinese (Hong Kong) SF-36 norming survey in 1998 were used for analysis in this study. The detailed sampling and data collection methods have been described in previous papers [16, 17]. *All subjects answered the Chinese (Hong Kong) translation of the SF-36 and a structured questionnaire on sociodemographic data. Each subject was also asked to indicate whether he/she had ever been diagnosed by a doctor to have hypertension, diabetes mellitus, heart disease, stroke, chronic pulmonary disease, joint disease, psychological illness or any other chronic disease. A subject was classified as not having any chronic disease if the responses to the chronic disease questions were all negative.* Table 1 shows that the sociodemographic characteristics of the subjects were similar to those of the general adult population in Hong Kong [18]. *The sample was comparable to the US population sample [19] that was used to derive the standard SF-12 in mean age (42.9 Vs 43.6 years) and sex distribution (47.8% Vs 48% males).*

The Chinese (HK) specific SF-12 items were selected by multiple regressions of the Chinese (HK) specific SF-36 PCS and MCS scores derived from the HK Chinese adult population [3], based on the criteria of the International Quality of Life Assessment (IQOLA) Project for the cross-cultural adaptation of the SF-12 [8]. The Chinese (HK) specific PCS and MCS regression constants and coefficients for each item response were obtained by regressing the Chinese (HK) specific SF-36 PCS and MCS scores on the Chinese (HK) specific item scores. The standard SF-12 PCS and MCS scores were calculated by the standard algorithm described in the SF-12 Manual [6]. The Chinese (HK) specific and standard mean SF-12 PCS and MCS scores were determined for all subjects and by self-reported chronic disease groups.

Content validity was assessed by the proportion of total variance of the SF-36 PCS and MCS scores explained by the SF-12 PCS and MCS, and ≥90% was the expected standard [6, 8]. It was further assessed by Pearson correlations between the SF-12 and SF-36 PCS and MCS scores and the expected standard was ≥0.9 [6, 8]. Effect size difference between corresponding SF-12 and SF-36 PCS and MCS scores was used to determine if the SF-12 gave similar or different results from those of the SF-36 (Criterion validity). Effect size difference between the SF-36 and SF-12 scores was calculated by dividing their difference by the standard deviation (SD) of the SF-36 summary score.

Measurement equivalence between the standard and Chinese (HK) specific SF-12 was first assessed by Pearson correlations (expected standard ≥0.9) and then the effect size differences between the standard and Chinese (HK) specific scores. The effect size difference was calculated by dividing the difference between the corresponding SF-12 scores by the SD of the Chinese (HK) specific SF-12 score. *The standard and Chinese (HK) SF-12 scores were also compared by chronic disease groups in order to determine if they performed differently in different groups. Heart disease, chronic pulmonary disease, psychological problem and chronic joint disease were used as tracer conditions because they represent a spectrum of chronic diseases that are known to affect HRQOL [20].*

There is no consensus on what the minimally important difference (MID) should be, Cohen's [21] moderate effect size difference of 0.3 to 0.5 was adopted as the standard in this study, based on the findings from previous studies on other HRQOL measures [22-24].

The SAS programme was used for the multiple regressions analyses. The SPSS Programme for Windows 10.0 (SPSS Inc. Chicago, IL, USA) was used for all other data analyses.

Results

The Chinese (HK) Specific SF-12 PCS and MCS Scales

First forward stepwise regressions of the Chinese (HK) specific SF-36 PCS and MCS scores on the SF-36 items selected two items each from the Physical Functioning (PF1, PF8) and Mental Health (MH3, MH4) Scales, and one item each from the Role-Physical (RP3), Bodily Pain (BP1), Social Functioning (SF1) and Role-Emotional (RE3) Scales. The second forward stepwise regressions, with the General Health item (GH1) and the above items forced into the model, selected the remaining items (RP2, VT4 and RE1) that explained the greatest variance of the HK Chinese specific SF-36 PCS and MCS scores. *It is an IQOLA criterion to select GH1 as the General Health item because it is an item common to many HRQOL measures [8].* Table 2 shows the Chinese (HK) specific SF-12 items, in comparison with the standard SF-12 items. The items that were different are shown in bold. The numbers in brackets correspond to the question numbers in the SF-36 Health Survey.

Table 3 shows the regression coefficients of the Chinese (HK) specific SF-12 items and those of the standard SF-12 items, derived from the HK general Chinese population sample. *The regression coefficient of the best response choice of each item is not shown because it is the indicator variable.* The Chinese (HK) specific PCS and MCS regression coefficients of each item response were used separately to weight each item response for the calculation of the PCS and MCS scores. *The weight for the best response choice of each item is zero.* Summation of the relevant regression constant and item response PCS and MCS regression coefficients would give the Chinese (HK) specific SF-12 PCS and MCS scores, respectively.

Content and Criterion Validity of the SF-12 PCS and MCS

The R square in Table 3 indicates the proportion of total variance in the SF-36 PCS or MCS scores that was explained by the corresponding SF-12 summary score. The standard SF-12 PCS and MCS explained 82% and 89% of the total variances of the standard SF-36 PCS and MCS, respectively. The Chinese (HK) specific SF-12 PCS and MCS explained 88% and 90% of the total variances of the Chinese (HK) specific SF-36 PCS and MCS, respectively.

Table 4 shows the correlations between the SF-12 and SF-36 PCS and MCS scores. The correlations between the corresponding SF-36 and SF-12 summary scores all reached the expected standard of 0.9.

The mean and standard deviation (SD) of the Chinese (HK) specific and standard SF12 and SF-36 PCS and MCS scores of the whole sample and by self-reported chronic disease groups are shown in Table 5. The effect size differences (effect size 1) between corresponding SF-36 and SF-12 scores were all less than 0.3.

Measurement Equivalence between the Chinese (HK) specific and Standard SF-12

As shown in Table 4, the correlations between the corresponding standard and Chinese (HK) specific SF-12 PCS and MCS scores were just short of 0.9. The standard and Chinese (HK) specific SF-12 scores are compared in Table 5. The mean standard SF-12 PCS and MCS for the overall HK Chinese population were 50.2 and 48.4, respectively, which were similar to the US general population means of 50.

The Chinese (HK) specific and standard SF-12 detected similar significant differences between each chronic disease group and the ‘no chronic disease’ group. The largest difference between the Chinese (HK) specific and standard SF-12 scoring algorithms was the PCS score of people reporting heart diseases, with an effect size of 0.36.

Discussion

The standard SF-12 did not satisfy the criterion on item equivalence for the Chinese population in Hong Kong. Six items of the Chinese (HK) specific SF-12 were different from those of the standard SF-12, suggesting some cultural differences in differential item functioning (DIF) of the SF-36 between the Chinese and US populations [13, 25]. The number of items that were different between the standard and Chinese (HK) SF-12 was larger than those found in nine European countries, probably because there are more differences between the Chinese than European cultures and the US culture [8]. No other country has selected PF1, BP1 and RE1 as the best SF-12 items, which may reflect a Chinese cultural uniqueness. *However, the findings from the Chinese population in Hong Kong may not be generalizable to other Chinese populations whose social and economic developments are different. Studies comparing the population specific SF-12 items between Chinese populations in Mainland China, Taiwan, Singapore and Western countries could provide interesting information on whether DIF is ethnic or population specific.*

The standard SF-12 PCS explained only 82% of the total variance of the SF-36 PCS score because three items (two from the Physical Functioning and one from the Bodily Pain Scales) that contributed strongly to the *standard* SF-12 PCS score were not the best items for the HK Chinese population. Despite this deficiency, there were very strong correlations (≥0.9) between the standard SF-12 and SF-36 PCS and MCS scores, and there were no important differences (effect size <0.3) between corresponding SF-36 and SF-12 scores in different groups of subjects. The findings supported the content and criterion validity of the standard SF-12 for the Chinese population in Hong Kong.

The mean *standard* SF-12 PCS and MCS scores of the HK subjects were only 0.2 and 1.6 points different from the US population mean of 50, suggesting that the standard SF-12 was equivalent for this Chinese population. Therefore, pooling of the SF-12 data between the US and HK Chinese populations may be possible.

As expected, the Chinese (HK) specific SF-12 had better psychometric properties than the standard SF-12, which could imply better sensitivity and responsiveness for the Chinese. However, the Chinese (HK) specific SF-12 did not seem to differentiate between '*chronic disease' and 'no chronic disease'* groups better than the standard SF-12. The differences in the SF-12 scores obtained by the two scoring algorithms were all smaller than the minimally important difference (MID). Small improvements in cultural specificity and psychometric properties may not necessarily translate to real advantages in practice, and they have to be balanced against a decrease in international comparability. This point has also been highlighted by Skevington et al [26, 27], who found that country specific items did not significantly improve the performance of the standard WHO Quality of Life (WHOQOL) Assessment Form .

A major limitation of this cross-sectional study was that it could not assess the responsiveness of the standard and Chinese (HK) specific SF-12. The information on the sensitivity of the SF-12 in discriminating between chronic disease groups was also limited by possible errors in subjects' self-reporting, and the small number of conditions studied. Further studies are required to determine the responsiveness and sensitivity of the Chinese (HK) specific SF-12 and standard SF-12 as outcome measures in clinical trials.

It must also be pointed out that subjects in this study answered the full SF-36 survey from which the data of the standard and Chinese (HK) specific SF-12 were

extracted. Context effects could lead to different results when people answer the 12 items as a stand-alone survey instead of embedded items of a longer survey. The measurement equivalence between the standard and Chinese (HK) specific SF-12 should be assessed again by studies that administer the two forms as independent stand-alone surveys.

Conclusions

The standard SF-12 Health Survey has been shown to be valid and equivalent for the Chinese in Hong Kong. It can substitute the SF-36 for the measurement of the Physical and Mental Health Summary (PCS and MCS) scores for the Chinese. The standard SF-12 scoring algorithm is recommended for the Chinese so that cross-cultural comparison and pooling of data are possible. The Chinese (HK) specific SF-12 showed better psychometric properties than the standard SF-12. It may have a place in small local studies that require a more sensitive HRQOL measure instead of international comparability

This was the first study to show that the standard SF-12 Health Survey was valid and equivalent for a Chinese population. We hope our findings will encourage more studies of the standard SF-12 in Chinese populations in China, Singapore and Western countries, so that enough data can be accumulated to support the use of this popular HRQOL measure on the world's largest ethnic group and Chinese can be included in more clinical trials. Further studies on the population specific SF-12 items in other Chinese populations could provide interesting information on whether differential item functioning is ethnic or population specific.

Acknowledgement

The general population norming survey of the Chinese (Hong Kong) SF-36 was approved by the Ethics Committee of the University of Hong Kong (EC 842-96). It was funded by the Health Services Research Grant, the Government of Hong Kong SAR (HSRC #711026).

I would like to thank Alex Chan, Willis Ho, Joanna Shing, Ka-Lai Chan, Wai-Hung Yu, June Chan, Chi-Kwan Wong, Wing-Yee Lai, Yick-Lok Chan and Hing-Wai Tsang, for their help in data collection and analysis.

Parts of this paper have been submitted to the University of Hong Kong for the award of the Doctor of Medicine degree.

Conflict of Interest: Nil

References

1. *Lam C L K, Gandek B, Ren X S, Chan M S. Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey. J Clin Epidemiol 1998; 51: 1139-1147.*
2. *Lam C L K. Reliability and construct validity of the Chinese (Hong Kong) SF-36 for patients in primary care. H K Pract 2003; 25: 468-475.*
3. *Lam C L K, Tse E Y Y, Gandek B. Construct validity, reliability and equivalence of the SF-36 Physical and Mental Health Summary Scales in a Chinese population. Under review.*
4. *Thumboo J, Fong K Y, Chan S P, et al. The equivalence of English and Chinese SF-36 versions in bilingual Singapore Chinese. Quality of life Research 2002; 11: 495-503.*
5. *Thumboo J, Fong K Y, Machin D, et al. A community-based study of scaling assumptions and construct validity of the English (UK) and Chinese (HK) SF-36 in Singapore. Quality of Life Research 2001; 10: 175-188.*
6. *Ware J E, Kosinski M, Keller S D, How to Score the SF-12 Physical and Mental Health Summary Scales. Second ed. Boston: The Health Institute, New England Medical Center; 1995. 96.*
7. *Ware J E, Kosinski M, Keller S D. A 12-Item Short-Form Health Survey. construction of scales and preliminary tests of reliability and validity. Medical Care 1996; 34: 220-233.*
8. *Gandek B, Ware J E, Aaronson N K. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. J Clin Epidemiol 1998; 51: 1171-1178.*
9. *Ware J E ,Kosinski M, Norm-Based scoring for Physical and Mental Summary Measures. In: Ware JE and Kosinski M, eds. SF-36 Physical & Mental Health Summary Scales: A Manual for Users of Version 1, Lincoln, Rhode Island:QualityMetric, Inc.; 2001: 28-35.*
10. *Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health - related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993; 46: 1417-32.*
11. *McColl E, Christiansen T, Konig-Zahn C, Making the right choice of outcome measure. In: Hutchinson A, Bentzen N, and Konig-Zahn C, eds. Cross Cultural Health Outcome Assessment: A User's Guide, Ruinen, the Netherlands:European Research Group on Health Outcomes; 1997: 12-26.*
12. *Lim L L Y ,Fisher J D. Use of the 12-item Short-Form (SF-12) Health Survey in an Australian heart and stroke population. Quality of Life Research 1999; 8: 1-8.*
13. *Bjorner J B, Kreiner S, Ware J E, Damsgaard M T, Berch P. Differential item functioning in the Danish translation of the SF-36. J Clin Epidem 1998; 51: 1189-1202.*
14. *Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. Quality of Life Research 1998; 7: 323-335.*
15. *Bullinger M, Ensuring international equivalence of quality of life measures: problems and approaches to solutions. In: Orley J and Kuyken W, eds. Quality of Life Assessment: International Perspectives, Berlin:Springler-Verlag; 1994: 33-40.*

16. *Lam C L K, Lauder I J, Lam T P, Gandek B. Population based norming of the Chinese (HK) version of the SF-36 Health Survey. H K Pract 1999; 21: 460-470.*
17. *Lam C L K, Fong D Y T, Lauder I J, Lam T P. The effect of health-related quality of life (HRQOL) on health service utilisation of a Chinese population. Soc Sc Med 2002; 55: 1635-1646.*
18. *Census & Statistics Department Hong Kong, Main Tables of the 2001 Population Census. Hong Kong: Census & Statistics Department, HKSAR; 2002. 1-346*
19. *Gandek B ,Ware J E. Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach. J Clin Epidemiol 1998; 51: 953-959.*
20. *Lam C L K, Lauder I J, Lam T P D. The impact of chronic diseases on health services and quality of life of a Chinese population. Asia Pacific Fam Med 2003; 2: 98-106.*
21. *Cohen J, The t test for measures. In: Cohen J, ed. Statistical Power Analysis for the Behavioral Sciences, Hillsdale, New Jersey:Lawrence Erlbaum Associates; 1988: 19-74.*
22. *Kazis L E, Anderson J J, Meenan R F. Effect sizes for interpreting changes in health status. Medical Care 1989; 27: S178-S189.*
23. *Wyrwich K W, Nienaber N A, Tierney W M, Wolinsky F D. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Medical Care 1999; 37: 469-478.*
24. *Norman G R, Sridhar F G, Walter S D, Guyatt G H. The relation of distribution- and anchor-based approaches in interpretation of changes in health related quality of life. Med Care 2001; 39: 1039-1047.*
25. *Holman R, Lindeboom R, Vermeulen M, Glas C A W, de Haan R J. The Amsterdam Linear Disability Score (ALDS) Project, differential item functioning with regard to gender. QOL Newsletter 2002; 29: 13-14.*
26. *Skevington S M. Advancing cross-cultural research on quality of life: observations drawn from the WHOQOL development. Quality of Life Research 2002; 11: 135-144.*
27. *Skevington S M, Bradshaw J, Saxena S. Selecting national items for the WHOQOL: conceptual and psychometric considerations. Social Science & Medicine 1999; 48: 473-487.*

Table 1: Sociodemographic Characteristics of Study Sample Compared with the Hong Kong General Population

	Sample	Hong Kong General Adult Population [a]
	N=2,410	**N=5,333,610**
Mean Age (years)	42.9	42.3
Age Group (years)		
18-44	56.7%	58.6%
45-64	23.7%	27.4%
65 or above	15.3%	14.0%
Refused to answer	4.2%	0%
Male	47.8%	48.3%
Female	52.2%	51.7%
Marital Status		
Now Married	58.0%	59.4%
Never Married	33.8%	31.9%
Widow/Widower	5.8%	6.0%
Divorced/Separated	1.3%	2.7%
Refused to Answer	1.1%	0%
Educational Level		
No Schooling	6.9%	8.4%
Primary	22.3%	20.5%
Secondary	52.2%	54.6%
Tertiary	17.8%	16.4%
Refused to Answer	0.9%	0%
Social Class by Occupation		
Managers and administrators	N.A.	10.7% [b]
Professional	3.1%	5.5%
Associate Professional	14.7%	15.0%
Skilled Worker	35.4%	33.5% [c]
Semi-skilled Worker	24.6%	15.0% [d]
Non-skilled Worker	14.4%	19.8% [e]
Refused to Answer	7.7%	0%

a. Data from the Hong Kong 2001 Population Census.
b. This occupation category is not applicable to the social class by occupation classification
c. Craft workers, plant and machine operators and assemblers.
d. Service and shop sales workers.
e. Workers in elementary occupation, agriculture and fishery, and unclassified.

Table 2: The Chinese (HK) Specific SF-12 Items Compared with the Standard SF-12 Items

SF-36 Scales	Chinese (HK) specific SF-12 Items	Standard SF-12 Items
Physical Functioning (PF)		
	PF1 (3a) Vigorous activities	PF2 (3b) Moderate activities
	PF8 (3h) Walking several blocks	PF4 (3d) Climbing several flights
Role-Physical (RP)		
	RP2 (4b) Accomplished less	RP 2 (4b) Accomplished less
	RP3 (4c) Limited in kind of work	RP3 (4c) Limited in kind of work
Bodily Pain (BP)		
	BP1 (7) How much bodily pain have you had	BP2 (8) how much did pain interfered with work
General Health (GH)		
	GH1 (1)Your health is……	GH1 (1) Your health is……
Vitality (VT)		
	VT4 (9i) Did you feel tired	VT2 (9e) Did you have a lot of energy
Social Functioning (SF)		
	SF1 (6) Extent social activities was interfered	SF2 (10) How much time social activities was interfered
Role-Emotional (RE)		
	RE1 (5a) Cut down time on work	RE2 (5b) Accomplish less
	RE3 (5c) Didn't do work as carefully	RE3 (5c) Didn't do work as carefully
Mental Health (MH)		
	MH3 (9d) Felt calm & peaceful	MH3 (9d) Felt calm & peaceful
	MH4 (9f) Felt downhearted & blue	MH4 (9f) Felt downhearted & blue

Table 3: Forward Stepwise Regressions of SF-36 PCS and MCS Scores on the SF-12 Item Responses

Item_ Response Scores	PCS Regression Coefficients		MCS Regression Coefficients	
	Chinese (HK) specific	Standard	Chinese (HK) specific	Standard
PF1_1	-8.042639	-----	2.795780	-----
PF1_2	-3.641426	-----	1.121187	-----
PF2_1	-----	-6.609693	-----	3.461042
PF2_2	-----	-2.782074	-----	1.314947
PF4_1	-----	-6.269240	-----	2.586866
PF4_2	-----	-2.427698	-----	0.752688
PF8_1	-16.203705	-----	7.818665	-----
PF8_2	-7.963922	-----	3.164988	-----
RP2_1	-4.343623	-4.390177	-0.705448	1.022170
RP3_1	-5.044296	-5.047476	0.256528	1.278842
BP1_1	-17.012005	-----	3.635025	-----
BP1_2.2	-12.695771	-----	2.349628	-----
BP1_3.1	-9.002881	-----	1.931547	-----
BP1_4.2	-6.377284	-----	1.507313	-----
BP1_5.4	-3.772960	-----	0.652800	-----
BP2_1	-----	-12.257268	-----	2.208989
BP2_2	-----	-10.594807	-----	2.500285
BP2_3	-----	-7.912197	-----	1.500170
BP2_4	-----	-4.970550	-----	1.033358
GH1_1	-8.704344	-8.042873	-0.841167	0.184282
GH1_2	-5.382641	-4.663071	-1.133139	-0.389631
GH1_3.4	-3.230279	-2.706827	-0.660725	-0.349572
GH1_4.4	-1.936141	-1.671905	-0.795015	0.330309
VT2_1	-----	-1.704222	-----	-7.001461
VT2_2	-----	-1.355533	-----	-5.031671
VT2_3	-----	-0.262164	-----	-4.012001
VT2_4	-----	-0.150904	-----	-2.677302
VT2_5	-----	0.150005	-----	-1.396547
VT4_1	-2.301203	-----	-6.694192	-----
VT4_2	-1.673615	-----	-6.555417	-----
VT4_3	-1.217702	-----	-4.965228	-----
VT4_4	-0.849186	-----	-2.403254	-----
VT4_5	-0.495087	-----	-1.041427	-----
SF1_1	2.955278	-----	-14.617923	-----
SF1_2	1.116653	-----	-12.142296	-----
SF1_3	1.433979	-----	-7.841254	-----
SF1_4	0.861761	-----	-4.676580	-----
SF2_1	-----	0.286656	-----	-8.236227
SF2_2	-----	-0.189464	-----	-6.857423
SF2_3	-----	0.193895	-----	-5.284785
SF2_4	-----	0.482796	-----	-3.301877
RE1_1	2.468990	-----	-6.099051	-----
RE2_1	-----	2.747609	-----	-6.981024
RE3_1	1.642657	2.143392	-5.120612	-5.946570
MH3_1	0.486081	2.865890	-8.496928	-8.255860
MH3_2	1.644377	3.500893	-8.257450	-6.883770
MH3_3	0.696675	2.694178	-6.255882	-5.404594
MH3_4	0.864621	2.333822	-4.238056	-3.439909
MH3_5	0.774435	1.609226	-2.544268	-1.943186
MH4_1	0.851938	4.534201	-12.868018	-15.794343
MH4_2	-0.119061	2.494064	-9.187208	-12.925241
MH4_3	1.319095	2.212045	-7.247869	-9.157472
MH4_4	0.987409	1.627192	-4.368062	-5.395771
MH4_5	0.717032	0.870407	-2.320460	-2.871620
Constant	60.175534	55.551534	62.742378	61.557734
R Square	0.8766	0.8232	0.9017	0.8897

Table 4: Correlations between the SF-36 and SF-12 PCS and MCS Scores

	Std36PCS	Std12MCS	HK36MCS	HK12PCS
Std36MCS	-0.126	0.938	0.985	0.022
Std12PCS	0.897	-0.073	-0.021	0.847
HK36PCS	0.975	0.050	0.000	0.936
HK12MCS	-0.049	0.894	0.950	0.040

Notes

Std36PCS = SF-36 PCS calculated by the standard (US) scoring algorithm
Std36MCS = SF-36 MCS calculated by the standard (US) scoring algorithm
Std12PCS = SF-12 PCS calculated by the standard (US) scoring algorithm
Std12MCS = SF-12 MCS calculated by the standard (US) scoring algorithm
HK36PCS = SF-36 PCS calculated by the Chinese (HK) specific scoring algorithm
HK36MCS = SF-36 MCS calculated by the Chinese (HK) specific scoring algorithm;
HK12PCS = SF-12 PCS calculated by the Chinese (HK) specific scoring algorithm
HK12MCS = SF-12 MCS calculated by the Chinese (HK) specific scoring algorithm

Table 5: Chinese (HK) Specific and Standard PCS and MCS Scores by Groups

	Mean (SD)			
	Std PCS	**HK PCS**	**Std MCS**	**HK MCS**
All Subjects (N=2410)				
SF-36	51.4 (7.7)	50.0 (10.0)	48.0 (9.4)	50.0 (10.0)
SF-12	50.2 (7.0)	50.0 (9.4)	48.4 (8.8)	50.0 (9.5)
Effect size 1	0.16	0	0.04	0
Effect size 2	0.02		0.17	
No Chronic Disease (n=1493**)**				
SF-36	53.8 (5.5)	53.5 (6.7)	48.6 (8.7)	50.5 (9.2)
SF-12	52.3 (4.8)	53.2 (6.3)	49.1 (8.1)	50.6 (8.8)
Effect size 1	0.27	0.04	0.06	0.01
Effect size 2	0.14		0.17	
Any Chronic Disease (n=917)				
SF-36	47.4 (8.9)	44.3 (11.8)	47.0 (10.4)	49.2 (11.1)
SF-12	46.9 (8.5)*	44.7 (11.0)*	47.4 (9.8)*	49.1(10.5)*
Effect size 1	0.06	0.03	0.04	0.01
Effect size 2	0.20		0.16	
Heart Disease (n=94)				
SF-36	41.7 (11.0)	36.0 (14.7)	46.9 (10.2)	49.2 (10.9)
SF-12	41.8 (10.6)*	37.0 (13.4)*	46.7 (9.4)	48.7 (10.9**)**
Effect size 1	0.01	0.07	0.02	0.05
Effect size 2	0.36		0.18	
Psychological Diseases (n=94)				
SF-36	45.8 (9.6)	41.6 (12.9)	40.2 (10.8)	41.6 (11.8)
SF-12	45.4 (9.0)*	42.0 (11.7)*	41.4 (10.1)*	42.7 (11.4)*
Effect size 1	0.04	0.03	0.11	0.09
Effect size 2	0.29		0.11	
Pulmonary Diseases (n=128)				
SF-36	46.6 (10.9)	43.6 (14.2)	44.6 (10.6)	46.0 (11.1)
SF-12	46.3 (10.6)*	43.9 (13.4)*	44.4 (10.4)*	46.3 (10.6)*
Effect size 1	0.03	0.02	0.02	0.03
Effect size 2	0.18		0.18	
Joint Diseases (n=473)				
SF-36	45.7 (9.3)	41.8 (12.3)	47.1 (10.3)	49.3 (11.1)
SF-12	45.4 (9.1)*	42.6 (11.4)*	47.4 (9.6)*	49.0 (10.7)*
Effect size 1	0.03	0.07	0.03	0.03
Effect size 2	0.25		0.15	

Notes

Std PCS= PCS score calculated by the standard (US) scoring algorithm; HK PCS = PCS calculated by the Chinese (HK) specific scoring algorithm; Std MCS= MCS calculated by the standard (US) scoring algorithm; HK MCS= MCS calculated by the Chinese (HK) specific scoring algorithm

Effect Size 1 *difference between SF-12 and SF-36 summary score / SD of SF-36 summary score*

Effect Size 2 *difference between standard and Chinese (HK) specific summary score/ SD of the Chinese (HK) specific summary score*

* *Difference between' no chronic disease' and disease group is significant by the two-sample t test, with $p<0.05$*