

Psychometrics and population norm of the Chinese (HK) SF-36 Health Survey_Version 2

Elegance T P Lam 林定珮, Cindy L K Lam 林露娟, Yvonne Y C Lo 盧宛聰, Barbara Gandek

Summary

Objective: To establish the psychometric properties and norm of the Chinese (HK) SF-36_version 2 Health Survey for the adult population in Hong Kong (HK) to facilitate its application and interpretation.

Design: A cross-sectional random telephone survey of the general adult population.

Subjects: 2410 Chinese adults randomly selected from the general Chinese adult population in Hong Kong. The mean age of the subjects was 42.9 (S.D. 17.3) years, 48% were men and 38% had one or more chronic disease.

Main outcome measures: Responses to the SF-36v2 Health Survey questions were extracted. Item-scale correlations, internal and test-retest reliabilities, and the factor structure of the SF-36v2 Health Survey scores were analysed. The SF-36v2 Health Survey scores were calculated by the standard algorithm to establish the population norm.

Results: All items had 100% scaling success indicating discriminant validity. Internal consistency and test-retest reliabilities of all scales were good (coefficients 0.66 to 0.89). The hypothesized two-factor structure underlying construction of the physical and mental health summary scales was confirmed. The psychometric properties of the SF-36v2 Health Survey were generally better than version 1. There were significant differences in the population norms between versions 1 and 2 of the Chinese (HK) SF-36 Health Survey, especially in the role-physical and role-emotional scales.

Funding: This study was funded by Health Services Research Grant (#711026), the Government of the HKSAR.

Elegance TP Lam, BSc, MMedSc
PhD Candidate,

Cindy LK Lam, MBBS, MD (HK), FRCGP, FHKAM (Family Medicine)
Clinical Professor,

Yvonne YC Lo, MBChB, FRACGP, FHKCFP, FHKAM (Family Medicine)
Clinical Assistant Professor,

Family Medicine Unit, Department of Medicine, the University of Hong Kong.

Barbara Gandek, MS
Scientist,
Health Assessment Lab, Waltham, USA

Correspondence to: Professor Cindy L K Lam, Family Medicine Unit, the University of Hong Kong, 3/F, Ap Lei Chau Clinic, 161 Main Street, Ap Lei Chau, Hong Kong SAR.

Conclusion: The Chinese (HK) SF-36v2 Health Survey is valid and reliable for measuring HRQOL of Chinese adults in Hong Kong, and population norm is now available to support the interpretation of its scores.

Keywords: Quality of life, SF-36, Norm, Chinese, validity, reliability, psychometrics

摘要

目的：確定中國人(香港)SF-36 第二版的健康調查的精神測定特性和標準適用於香港成年人口，以方便其應用和解釋。

設計：橫切面的隨機抽樣電話調查，目標是普羅成年人。

對象：從普羅的香港中國成年人口隨機抽樣抽出2410人。平均年齡為42.9(S.D.17.3)歲，48%為男性，38%有一樣和超個一樣的慢性病。

主要測量內容：SF-36v2 健康調查結果全收集起來。項目/標度相關，內部和重複測試可靠性，和 SF-36v2 成份結構分析。然後用標準的方程式去計算 SF-36v2 健康調查的分樓來算出人口的特性和標準。

結果：所有項目均達到100%定標成功，標示出辨別的可信性。內部協合和重複測試可靠性都非常好(系數為0.66至0.89)。雙份子結構的假設在其基本結構，生理和心理健康，標分受到確定。SF-36v2 的表現比 SF-36v1 為佳。SF-36v2 和 SF-36v1 的人口特性有很明顯分別，尤其是在心理功用和生理功用標度方面。

結論：中文(香港)SF-36v2 健康調查是可靠和可信的，可準確量度香港成年人口的生活質素。現在我們已找出其人口特性來支持其分數的解釋。

主要詞彙：生活質素，SF-36，特性，中文，可信性，可靠性，精神測定

HK Pract 2008;30:185-198

Introduction

The SF-36 Health Survey developed by Ware *et al.* is the most widely used health-related quality of life (HRQOL) measure in Hong Kong (HK) and worldwide.¹⁻³ The survey includes 35 items measurement HRQOL that are summarized into eight multi-item scales, along with

1 item on health change. The first version of the SF-36 Health Survey (version 1) has been adapted and validated in more than 40 populations with norm references available from 14 populations including Hong Kong.^{3,4} Several weaknesses of version 1 were identified: the layout of the questions was inconsistent, colloquial or double negative wordings were used, psychometric performance of the role-physical (RP) and role-emotional (RE) scales (questions 4 and 5, respectively) was suboptimal due to the use of dichotomous (yes/no) responses; and the differentiation between some of the six response options of question 9 may be difficult for some respondents.⁵⁻⁷ Modification to the SF-36 Health Survey was carried out by the original authors to address these problems and version 2.0 (SF-36v2 Health Survey) was produced in 1996.⁴

The changes included reformatting the layout of the questions and answers to a consistent horizontal format, revision of some wordings of questions 3, 4, 5 and 9; replacement of the dichotomous response choices with a 5-point frequency scale for the items of questions 4 and 5; and deleting the response option 'a good bit of the time' from question 9 to change the 6-point to a 5-point scale. Most of the changes in the wordings had already been incorporated into the Chinese (HK) SF-36 Health Survey during the version 1 translation process, and the Chinese translations for all the statements and response options included in the SF-36v2 Health Survey could be extracted from version 1, so repeat translation was not necessary. The differences between the two versions of the Chinese (HK) SF-36 Health Survey are shown in the **Appendix**.

Studies in the United States, United Kingdom, Sweden and Australia have confirmed that SF-36v2 Health Survey was superior to version 1 in that it had fewer missing data, lower floor and ceiling effects, better score precision and a higher sensitivity for the role functioning scales.^{4, 8-11} Version 2 is expected to replace version 1 of the SF-36 in the near future. We need to establish the psychometric properties and norm of the SF-36v2 Health Survey before it can be applied and interpreted properly in our Hong Kong population.

In the population validation and norming study of the Chinese (HK) SF-36 Health Survey in Hong Kong the two role-functioning questions that were modified for version 2 were also included, providing an opportunity for the extraction of population data on the SF-36v2 Health Survey. The aim of this study was to determine the

psychometric properties and population norm of the Chinese (HK) SF-36v2 Health Survey.

Methods

Sample

2410 Chinese adults randomly selected from the general Chinese adult population in HK were interviewed by telephone. The response rate of this survey was 84.4% (2410 out of 2857 sampled). The mean age of the subjects was 42.9 (range 18 to 88, S.D. 17.3) years old, 48% were men and 38% had one or more chronic disease. The sociodemographic characteristics of the subjects were comparable to those found in the population Census. The details of the sampling and survey methods are described in earlier papers.^{3, 12}

The first 240 subjects who agreed to a repeat survey were contacted 2 weeks later to answer the questions again to determine test-retest reliability and 200 (83%) completed the retest.

Survey instruments

The survey instruments consisted of the Chinese (HK) SF-36 Health Survey (version 1), followed by questions 4 (role-physical items) and 5 (role-emotional items) of the Chinese (HK) SF-36v2 Health Survey, and a structured questionnaire on socio demography, morbidity and service utilization. The instruments were administered by trained interviewers in Cantonese.

Data analysis

All data analyses were carried out with the SPSS for Windows 15.0 programme. Statistical significant levels were set at *p* values less than 0.05.

The Chinese (HK) SF-36v2 Health Survey data were extracted from the responses to questions 1, 3, 6, 7, 8, 9, 10 and 11 of version 1, which were the same as those of version 2, and the responses to the SF-36v2 Health Survey questions 4 and 5. The response value 'a good bit of the time' for question 9 items was recoded randomly to the adjoining values of 'most of the time' or 'some of the time' for the calculation of the SF-36v2 vitality and mental health scale scores. Item responses recoding and scale score calculations were carried out according to the standard methods described in the SF-36v2 Health Survey manual.¹⁰

The construct validity of the Chinese (HK) SF-36v2 Health Survey scales was tested by item-scale Pearson correlation to assess whether items:

1. were substantially correlated ($r \geq 0.4$) to the hypothesized scale score.
2. have similar item-scale correlations and equal variance (standard deviation) in the same scale to justify summation without weighting.
3. correlated significantly higher (greater than two standard errors) with their hypothesized scale than other scales. The percentage of this scaling success on item discriminant validity by the total number of item-scale correlations of each scale was calculated.

Factor analysis using the varimax rotation method was done on the scale scores to extract two principal components and test the hypothesized two-dimensional (physical and mental) of the SF-36v2 Health Survey. The two principal components should explain $\geq 60\%$ of the total variance of the SF-36v2 Health Survey scores, and $\geq 70\%$ of the reliable variance of each scale score, as found in the US and other populations.¹³ The pattern of correlations between the eight scales and two rotated components was examined to determine the basis for the components interpretation as physical and mental summary measures.

Internal consistency (reliability) of scale scores was measured by the Cronbach's alpha coefficient and test-retest reliability was assessed by intra-class correlation (ICC). The recommended standard is 0.7 or greater for group comparisons.

Descriptive statistics including mean, standard deviation (SD), ceiling and floor proportions of the scale scores of SF-36v2 Health Survey were calculated for the whole sample and by age-sex groups to be used as the population norm reference.

Results

Validity and reliability of the scales

Table 1 shows the item-scale Pearson correlations between each item and the scales. The correlation between each item and its hypothesized scale (after correction for overlap) was >0.4 except for PF10 (0.38) and GH3 (0.32), supporting item internal consistency. The item-scale correlations and standard deviations of items

of the same scale were similar, supporting equal item weighting. The item-scale correlations for the SF-36v2 Health Survey RP items (0.75 to 0.78) were generally greater than those of version 1 (0.64 to 0.68). The same was also found with the RE items (0.70 to 0.77 for version 2 vs 0.62 to 0.71 for version 1). The item-scale correlations of the version 2 vitality (VT) and mental health (MH) items were similar to those of version 1 despite a change from the 6-point to a 5-point response scale. The item-hypothesized scale correlations were significantly higher than item-other scales correlations for all items, which means scaling success on item discriminant validity was perfect (100%) for all scales.

Cronbach's alpha coefficients of internal consistency (reliability) were above 0.7 for all the Chinese (HK) SF-36v2 Health Survey scales except the general health (GH) scale (0.66), which was the same for both versions 1 and 2 (**Table 2**). The reliabilities of the RP and RE scales of version 2 were better than those of version 1 (0.89 in RP_v2 vs. 0.83 in RP_v1; 0.86 in RE_v2 vs. 0.82 in RE_v1). There was almost no difference in the internal reliability in the VT and MH scales between version 1 and 2 despite a reduction in the number of response options in version 2. Intra-class coefficients (ICC) measuring test-retest reliability were above 0.7 for all scales.

Validity of the two principal component factor structure

The results of the factor analysis with varimax rotation on the Chinese SF-36v2 Health Survey scale scores are shown in **Table 3**. Two principal component factors (physical and mental) with eigenvalues greater than 1.0 (3.61 for factor 1 and 1.01 for factor 2) were extracted from the eight scale scores. The two principal components explained 59% of the total variance of SF-36 scores and 64 to 87% of the reliable variance of each individual scale. The correlations between the scale scores and the two factors were similar to those hypothesized and to those of version 1.¹⁴ The population specific factor coefficients of the Chinese (HK) SF-36v2 Health Survey are compared with those of version 1 with reference to the US standard in **Table 4**.

Population norm of the Chinese (HK) SF-36v2 Health Survey

Table 5 shows the distribution of Chinese (HK) SF-36v2 Health Survey scale scores, compared with corresponding values of version 1 as appropriate. The scores of the PF, BP, GH, SF scales were the same for

Table 1: Item-scale pearson correlations of the Chinese (HK) SF-36v2 Health Survey

Item (Question no.)	Mean (SD)	Item-scale correlation								Scaling
		PF	RP	BP	GH	VT	SF	RE	MH	Success %
PF1 (3a)	2.36 (0.77)	0.53*	0.35	0.34	0.44	0.28	0.12	0.16	0.15	100
PF2 (3b)	2.86 (0.43)	0.65*	0.38	0.29	0.29	0.21	0.14	0.14	0.15	100
PF3 (3c)	2.96 (0.24)	0.51*	0.36	0.25	0.19	0.14	0.17	0.15	0.15	100
PF4 (3d)	2.79 (0.48)	0.61*	0.41	0.28	0.34	0.24	0.18	0.23	0.19	100
PF5 (3e)	2.97 (0.20)	0.52*	0.35	0.21	0.21	0.17	0.16	0.14	0.14	100
PF6 (3f)	2.79 (0.49)	0.56*	0.36	0.32	0.33	0.24	0.11	0.14	0.15	100
PF7 (3g)	2.74 (0.55)	0.63*	0.42	0.29	0.34	0.25	0.15	0.23	0.19	100
PF8 (3h)	2.94 (0.30)	0.65*	0.41	0.28	0.27	0.18	0.18	0.16	0.17	100
PF9 (3i)	2.98 (0.17)	0.53*	0.36	0.21	0.16	0.14	0.17	0.14	0.14	100
PF10 (3j)	2.99 (0.13)	0.38*	0.29	0.20	0.14	0.09	0.20	0.12	0.11	100
RP1 (4a)	4.65 (0.79)	0.44	0.75*	0.42	0.34	0.27	0.39	0.43	0.30	100
RP2 (4b)	4.60 (0.83)	0.45	0.77*	0.42	0.36	0.30	0.38	0.45	0.31	100
RP3 (4c)	4.63 (0.83)	0.49	0.78*	0.41	0.36	0.29	0.36	0.41	0.27	100
RP4 (4d)	4.59 (0.84)	0.51	0.76*	0.45	0.38	0.32	0.39	0.43	0.31	100
BP1 (7)	5.20 (1.19)	0.38	0.41	0.79*	0.39	0.28	0.27	0.23	0.24	100
BP2 (8)	5.20 (1.13)	0.41	0.52	0.79*	0.43	0.31	0.37	0.31	0.28	100
GH1((1)	2.60 (0.97)	0.35	0.30	0.28	0.45*	0.33	0.17	0.19	0.25	100
GH2 (11a)	4.02 (1.26)	0.36	0.35	0.32	0.46*	0.31	0.23	0.23	0.30	100
GH3 (11b)	3.55 (1.23)	0.24	0.24	0.24	0.32*	0.14	0.14	0.16	0.14	100
GH4 (11c)	3.31 (1.36)	0.32	0.26	0.29	0.42*	0.31	0.16	0.20	0.23	100
GH5 (11d)	2.71 (1.33)	0.24	0.22	0.27	0.45*	0.32	0.12	0.18	0.21	100
VT1 (9a)	3.02 (1.07)	0.26	0.23	0.22	0.36	0.50*	0.14	0.23	0.39	100
VT2 (9e)	3.12 (1.00)	0.24	0.22	0.22	0.33	0.53*	0.16	0.24	0.43	100
VT3 (9g)	3.99 (1.00)	0.23	0.28	0.20	0.25	0.46*	0.21	0.25	0.44	100
VT4 (9i)	3.46 (0.97)	0.19	0.26	0.27	0.32	0.46*	0.23	0.27	0.40	100
SF1 (6)	4.61 (0.76)	0.20	0.40	0.31	0.25	0.27	0.60*	0.47	0.39	100
SF2 (10)	4.68 (0.72)	0.19	0.38	0.29	0.20	0.18	0.60*	0.37	0.28	100
RE1 (5a)	4.55 (0.81)	0.24	0.45	0.27	0.26	0.32	0.42	0.75*	0.41	100
RE2 (5b)	4.55 (0.79)	0.21	0.45	0.25	0.27	0.28	0.41	0.77*	0.40	100
RE3 (5c)	4.42 (0.86)	0.21	0.41	0.24	0.25	0.30	0.42	0.70*	0.41	100
MH1 (9b)	4.02 (0.99)	0.14	0.26	0.21	0.25	0.42	0.27	0.32	0.51*	100
MH2 (9c)	4.18 (0.88)	0.19	0.29	0.23	0.26	0.41	0.34	0.42	0.63*	100
MH3 (9d)	3.65 (0.96)	0.16	0.19	0.14	0.19	0.41	0.22	0.26	0.46*	100
MH4 (9f)	3.95 (0.87)	0.18	0.28	0.23	0.26	0.39	0.33	0.40	0.61*	100
MH5 (9h)	3.49 (0.93)	0.17	0.22	0.19	0.27	0.43	0.22	0.26	0.49*	100

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

* Item-scale correlation corrected for overlap (relevant item removed from the calculation of scale score) that is significantly higher than those between the item and other scales by Steiger's t-test (16); standard error= 0.02.

version 1 and 2 because there was no difference in their items or response options. There were significant differences in the mean and SD of the RP and RE scale scores between the two versions. The floor effect (the proportion of respondents scoring at the lowest scores) was markedly reduced from 7.5% in version 1 to 0.6% in

version 2 and from 16.4% in version 1 to 0.3% in version 2 for the RP and RE scales, respectively. There was slight improvement in the ceiling effects (the proportion of respondents scoring at the highest scores) in these two

(Continued on page 193)

Clinically Proven, and Time-tested,

..... **Gold Standard** in Treatment
and Prevention of Eczema,
Irritating, Dry Skin Conditions.

Oilatum[®] Emollient

First-line Bath Emollient for
Controlling Eczematous Skin Conditions



Oilatum[®] Plus

Anti-bacterial Bath Emollient for
Preventing Secondary Bacterial Infection

Oilatum[®] Cream

Uniquely Formulated Emollient Cream for
Long-lasting all Day Protection



CARDURA XL^{*}
4mg, 8mg once daily
WITHOUT TITRATION
(doxazosin GITS)

Rapid Symptomatic Relief in Benign Prostatic Hyperplasia¹



Extended benefits over traditional α -blockers²

- Less postural hypotension²
- Eliminate the complicated titration steps²

References: 1. Roehrborn CG, Praisner A, Kirby R, et al. A double-blind placebo-controlled study evaluating the onset of doxazosin gastrointestinal therapeutic system in the treatment of benign prostatic hyperplasia. *Eur Urol* 2005;48:445-452. 2. Kirby RS, et al. A combined analysis of double-blind trials of the efficacy and tolerability of doxazosin in the gastrointestinal therapeutic system (GITS), doxazosin standard, and placebo in patients with benign prostatic hyperplasia. Detailed prescribing information is available upon request.



Pfizer Corporation Hong Kong Limited

16/F., Stanhope House, 738 King's Road, North Point, Hong Kong

Tel: (852) 2811 9711 Fax: (852) 2579 0599

Website: www.pfizer.com.hk

Rescue the face behind acne — in Asians

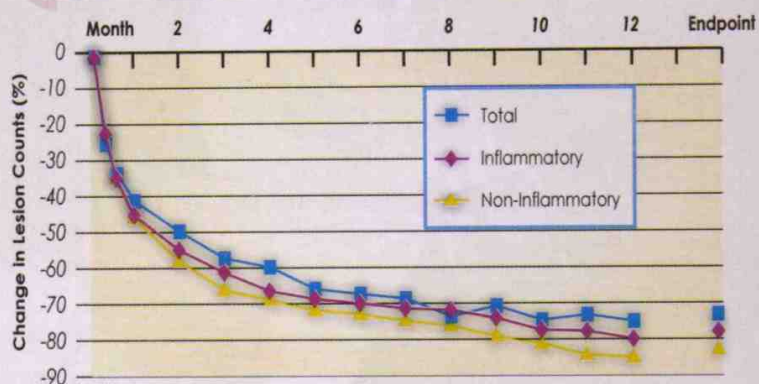


Differin
adapalene



Approval
in Japan

*“The first and only
topical retinoid in Japan”*



Highly effective and well-tolerated up to 12 months for treating Asian (Japanese) acne skin¹

*“Stops acne formation
Right from the start”*

International Consensus²



Mild

Comedonal

Papular/pustular

TOPICAL RETINOID

TOPICAL RETINOID
+ Topical Antimicrobial

Moderate

Papular/pustular

Nodular

TOPICAL RETINOID
+ Oral Antibiotic
+/- BPO

TOPICAL RETINOID
+ Oral Antibiotic
+ BPO

Severe

Nodular/Conglobate

Oral Isotretinoin

Differin targets the root cause and symptoms of acne^{2, 3}

*Differin prevents formation of microcomedones,
the precursor lesions of inflammatory
lesions and comedones.²*

References:

1. Kawashima M et al. Skin Research. 2007;6:504-12
2. Golnick H et al. J Am Acad Dermatol. 2003;49(1):S1-S37.
3. Fitzov E, Kligman A. JEADV. 2001;15(3):13-18.

GALDERMA
Committed to the future
of dermatology



Further information available on request
Galderma Hong Kong Ltd.
Unit 1601, 16/F CRE Building,
303 Hennessy Road, Wanchai, Hong Kong
Tel: (852) 2824 0333

Celebrate the Success
**A DECADE OF
EVIDENCE**

**The World's Most
Prescribed Statin¹ offers**

152 Million Patient-Years' Experience¹

> 400 Clinical Studies in **> 80,000** Patients

10 Landmark Cardiovascular Outcome Trials

Something To Cheer About!

10th Anniversary
LIPITOR
atorvastatin calcium
tablets



Pfizer Corporation Hong Kong Limited

16/F., Stanhope House, 738 King's Road, North Point, Hong Kong

Tel: (852) 2811 9711 Fax: (852) 2579 0599

Website: www.pfizer.com.hk

Table 2: Reliability and inter-scale correlations of the Chinese (HK) SF-36v2 Health Survey

Scale	Reliability		Inter-scale Correlation						
	Cronbach's alpha	Test-retest ICC	PF	RP	BP	GH	VT	SF	RE
PF	0.81	0.88							
RP	0.89	0.70	0.40						
BP	0.88	0.70	0.33	0.40					
GH	0.66	0.86	0.44	0.36	0.39				
VT	0.70	0.79	0.32	0.33	0.28	0.41			
SF	0.75	0.77	0.14	0.39	0.26	0.22	0.27		
RE	0.86	0.75	0.18	0.46	0.22	0.26	0.33	0.47	
MH	0.77	0.77	0.18	0.32	0.21	0.29	0.55	0.37	0.44

PF = Physical Functioning; RP = Role Physical; BP = Bodily Pain; GH = General Health; VT = Vitality; SF = Social Functioning; RE = Role Emotional; MH = Mental Health. ICC = Intra-class correlation.

Table 3: Correlations (r) between the Chinese (HK) SF-36v2 Health Survey Scales and two principal components

Scale	Hypothesized association		Chinese (HK) SF-36v2 Health Survey			Chinese (HK) SF-36 Health Survey		
	Physical	Mental	Physical ^a	Mental ^a	Reliable variance ^b	Physical ^a	Mental ^a	Reliable variance ^b
PF	+	-	0.83	0.07	0.86	0.82	0.04	0.83
RP	+	-	0.68	0.41	0.71	0.66	0.36	0.69
BP	+	-	0.71	0.22	0.64	0.72	0.21	0.65
GH	*	*	0.69	0.26	0.83	0.70	0.27	0.86
VT	*	*	0.32	0.61	0.69	0.38	0.60	0.71
SF	*	+	0.20	0.67	0.65	0.21	0.65	0.63
RE	-	+	0.19	0.75	0.71	0.08	0.78	0.74
MH	-	+	0.14	0.81	0.87	0.20	0.78	0.83

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

^aCorrelation between each scale and rotated principle component.

^bProportion of the reliable variance (total variance explained divided by the Cronbach's alpha) of each scale explained by the two components.

+Strong association ($r \geq 0.70$)

*Moderate association ($0.30 < r < 0.70$)

-Weak association ($r \leq 0.30$)

role functioning scales in version 2, but they were still very large. Table 6 shows the population mean Chinese (HK) SF-36v2 Health Survey scores of all subjects and by age and sex groups.

Discussion

The results confirmed that the Chinese (HK) SF-36v2 Health Survey satisfied all the scaling assumptions with a scaling success rate of 100%. The correlation between

item GH3 and the GH score was relatively low (0.32), which was the same with both versions. Feedback from the interviewers revealed that some respondents said that they were not sure of the answer to this item because they did not know the health status of others. This might also be the reason why the internal reliability of the GH scale did not reach the standard of 0.7

The psychometric properties of version 2, especially for the two role functioning (RP and RE) scales, were much better than those of version 1 in terms of internal

Table 4: Principal component factor coefficients of the Chinese (HK) SF-36v2 Health Survey Scales

Scale	Principal Component Factor Coefficients					
	US Standard SF-36 Health Survey *		Chinese (HK) SF-36v2 Health Survey		Chinese (HK) SF-36 Health Survey	
	Physical	Mental	Physical	Mental	Physical	Mental
PF	0.424	-0.230	0.479	-0.229	0.461	-0.227
RP	0.351	-0.123	0.276	0.025	0.275	0.013
BP	0.318	-0.097	0.357	-0.098	0.355	-0.095
GH	0.250	-0.016	0.332	-0.067	0.325	-0.051
VT	0.029	0.235	-0.004	0.265	0.033	0.251
SF	-0.008	0.269	-0.099	0.341	-0.078	0.331
RE	-0.192	0.434	-0.130	0.393	-0.194	0.448
MH	-0.221	0.486	-0.180	0.442	-0.122	0.412

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

* The Standard factor coefficients are the same for SF-36v2 Health Survey and SF-36 Health Survey, derived from the 1990 US general population study (10.15).

Table 5: Comparison between versions 1 and 2 of the Chinese (HK) SF-36v2 Health Survey Scale Scores

Scale	Mean (N=2410)	SD	% Floor	% Ceiling	95% CI
PF*	91.83	12.89	0.20	46.00	(91.31, 92.34)
RPv2	90.44	17.93	0.60	64.40	(89.72, 91.15)
RPv1	82.43	30.97	7.50	69.10	(81.19, 83.66)
BP*	83.98	21.89	0.50	54.70	(83.10, 84.85)
GH*	55.98	20.18	1.00	0.50	(55.17, 56.78)
VTv2	59.92	18.36	0.20	1.70	(59.18, 60.65)
VTv1	60.27	18.65	0.20	1.70	(59.53, 61.02)
SF*	91.19	16.57	0.10	70.80	(90.53, 91.85)
REv2	87.67	18.16	0.30	55.40	(86.94, 88.39)
REv1	71.66	38.36	16.40	58.40	(70.13, 73.19)
MHv2	71.46	16.67	0.00	4.50	(70.79, 72.12)
MHv1	72.79	16.57	0.00	4.50	(72.12, 73.45)

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

* No difference in the values between version 1 and version 2 because the items & response options are identical.

reliability, floor effects and smaller standard deviations, suggesting that it would be more sensitive and responsive. This illustrated the advantage of a 5-point response scale over that of a dichotomous scale. The floor effects were almost eliminated by the change from version 1 to 2, implying that the measure would be more able to detect deterioration in HRQOL. The ceiling effects of the SF-36v2 Health Survey were not much improved from those of version

1, as found in population studies in the US and other countries.^{4, 8-10} High ceiling effects in the pain and role-functioning scales are intrinsic to general population studies because most subjects are healthy. It is more important for a HRQOL measure to have low floor effect in the normal population so that it can detect deterioration from 'normal', as in the case of the SF-36v2 Health Survey. On the other hand, the ceiling effect should be low in patient populations if

Table 6: Population mean Chinese (HK) SF-36v2 Health Survey Scores by Age-sex

Scale	All ages					
	Male (n=1152)		Female (n=1258)		Total (n=2410)	
	Mean	SD	Mean	SD	Mean	SD
PF	94.02	10.88	89.82	14.19	91.83	12.89
RP	91.91	16.41	89.09	19.12	90.44	17.93
BP	87.07	18.98	81.14	23.91	83.98	21.89
GH	59.32	19.43	52.92	20.38	55.98	20.18
VT	61.30	17.24	58.65	19.25	59.92	18.36
SF	91.45	16.00	90.96	17.08	91.19	16.57
RE	88.35	17.33	87.04	18.88	87.67	18.16
MH	71.64	16.13	71.29	17.16	71.46	16.67
PCS (HK norm)	51.79	8.06	48.35	10.21	50.00	9.40
MCS (HK norm)	49.78	9.18	50.20	10.07	50.00	9.65
PCS (US norm)	54.30	6.05	51.73	7.48	52.96	6.95
MCS (US norm)	48.42	8.12	48.52	8.89	48.47	8.53

Scale	41-64 years old					
	Male (n=305)		Female (n=390)		Total (n=695)	
	Mean	SD	Mean	SD	Mean	SD
PF	93.03	10.80	88.73	13.16	90.62	12.36
RP	92.68	16.30	88.21	20.29	90.17	18.76
BP	86.50	20.38	79.57	24.30	82.61	22.90
GH	56.43	19.72	50.74	20.08	53.24	20.11
VT	62.52	16.92	58.38	19.85	60.20	18.73
SF	93.69	14.27	91.44	18.06	92.43	16.53
RE	91.23	16.16	86.39	20.34	88.51	18.76
MH	72.52	16.83	71.59	18.06	72.00	17.52
PCS (HK norm)	50.77	8.11	47.35	9.75	48.85	9.22
MCS (HK norm)	51.42	8.38	50.52	10.70	50.91	9.75
PCS (US norm)	53.45	6.09	50.96	7.16	52.05	6.82
MCS (US norm)	49.75	7.63	48.73	9.50	49.18	8.74

Scale	65 or above years old					
	Male (n=174)		Female (n=195)		Total (n=369)	
	Mean	SD	Mean	SD	Mean	SD
PF	84.34	17.19	74.59	20.63	79.19	19.67
RP	88.58	20.03	81.25	26.09	84.71	23.68
BP	84.82	22.79	70.77	28.22	77.39	26.70
GH	54.41	19.69	44.47	21.46	49.16	21.21
VT	62.93	20.59	56.47	19.28	59.52	20.14
SF	92.74	16.69	91.47	17.87	92.07	17.31
RE	90.13	17.44	87.61	22.34	88.80	20.20
MH	77.41	16.93	72.44	19.65	74.78	18.56
PCS (HK norm)	46.09	10.82	39.01	13.23	42.35	12.64
MCS (HK norm)	54.40	8.77	54.26	9.82	54.33	9.33
PCS (US norm)	50.37	7.70	45.30	9.24	47.69	8.90
MCS (US norm)	51.93	8.00	51.10	9.00	51.49	8.54

PF = Physical Functioning; RP = Role Physical; BP = Bodily Pain; GH = General Health;
 VT = Vitality; SF = Social Functioning; RE = Role Emotional; MH = Mental Health;
 PCS (HK norm) = Physical summary score normed on HK population mean and standard deviation;
 MCS (HK norm) = Mental summary score normed on HK population mean and standard deviation;
 PCS (US norm) = Physical summary score normed on US population mean and standard deviation;
 MCS (US norm) = Mental summary score normed on US population mean and standard deviation.

Key messages

1. Version 2 of the SF-36 Health Survey (SF-36v2) has improvements in the clarity of wording, questionnaire format and number of response options over the first version.
2. The Chinese (HK) SF-36v2 is valid with 100% scaling success in convergent and discriminant validity, and reliable.
3. The two principal component factor structure and coefficients of the Chinese (HK) SF-36v2 is equivalent to the US original, so the standard scoring algorithm for the calculation of the two summary scores is applicable to the Chinese population.
4. The SF-36v2 is likely to be more sensitive than version 1 because it has less floor effect and a better internal reliability.
5. The appropriate population norms should be used for the interpretation of the data of version 1 or 2 of the SF-36 Health Survey because there were significant differences in their population mean scores.

the measure is to be used to assess the effectiveness of treatment. Further studies are required to determine the ceiling effect of the SF-36v2 Health Survey in patient populations.

The hypothesized two principal component factor structure that is the conceptual base of the SF-36 physical and mental summary (PCS & MCS) scores was also confirmed. The two components explained 59% of the total variance of the Chinese (HK) SF-36v2 Health Survey scores, which was slightly better than the 58% found in version 1,¹⁴ and approaching the expected standard of 60%. The two components explained 70% or more of the reliable variance of each scale score except for the BP (64%), VT (69%) and SF (65%) scales, similar to those found with version 1.¹⁴ The replication of the two principal factor structure means that summation of the Chinese (HK) SF-36v2 Health Survey scale scores into the physical and mental summary scores is valid. The physical and mental factor coefficients (weightings for the calculation of the SF-36 physical and mental summary scores) were almost the same

between versions 1 and 2, and comparable to the standard derived from the US population. Equivalence between the population specific and standard (US) summary scale scoring algorithms of Chinese (HK) SF-36 Health Survey was confirmed in a previous study.¹⁴ Thus the Chinese (HK) SF-36v2 Health Survey physical and mental summary (PCS and MCS) scales should be scored by the standard algorithm for better international comparability.

There was significant difference in the population mean RP and RE scale scores between version 1 and version 2 indicating that normative values of version 1 cannot be used for the interpretation of version 2 data. The total population mean scores and standard deviations shown in **Table 6** should be used for norm-based scoring of the Chinese (HK) SF-36v2 Health Survey.^{14,15} We believe the normative scores are still applicable although the data were collected nearly 10 years ago, based on the findings by studies in the US showing that population mean SF-36 Health Survey scores remained very stable with a change of less than 3% (3 points in a scale range of 100) over 10 years.¹⁵

Limitation

The Chinese (HK) SF-36v2 Health Survey data presented in this paper were extracted from answers to relevant version 1 questions and the two SF-36v2 Health Survey role-functioning questions that were administered after version 1 questions, which could have an order effect on the responses. A general population survey with a stand-alone SF-36v2 Health Survey should be carried out to confirm the psychometric properties and update the population norm if resources are available.

Conclusion

The validity and reliability of the Chinese (HK) SF-36_Version 2 have been confirmed for the adult population in Hong Kong. Population norm (mean and standard deviation) of the Chinese (HK) SF-36v2 Health Survey is now available to facilitate the interpretation of scores. There was significant difference in the population means between versions

1 and 2, the appropriate norm reference should be used for comparison. Version 2 of the Chinese (HK) SF-36 Health Survey should be preferred to version 1 in future applications because it has better psychometric properties. The Chinese (HK) SF-36v2 Health Survey is expected to be more sensitive and responsive than the original version, which will need to be confirmed by further studies.

Acknowledgements

The SF-36® and SF-36v2® are registered trademarks of Medical Outcomes Trust. A copy of the Chinese (HK) SF-36v2 Health Survey and licence to use can be obtained from QualityMetric <http://www.qualitymetric.com>. ■

References

- Ware JE Jr., Snow KK, Kosinski M, *et al.* SF-36 Health Survey - Manual and Interpretation Guide. Boston: The Health Institute, 1993.
- Lam CLK, Gandek B, Ren XS, *et al.* 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.
- Lam CLK, Lauder IJ, Lam TP, *et al.* Population based norming of the Chinese (HK) version of the SF-36 health survey. *HK Pract* 1999;21:460-470.
- Ware JE Jr. SF-36 Health Survey Update. *SPINE* 2000;25:3130-3139.
- Bullinger M, Alonso J, Apolone G, *et al.* Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. *J Clin Epidemiol* 1998;51:913-923.
- Keller SD, Ware JE, Gandek B, *et al.* Testing the equivalence of translations of widely used response choice labels: Results from the IQOLA Project. *J Clin Epidemiol* 1998;51:933-948.
- Wagner AK, Gandek B, Aaronson NK, *et al.* Cross-cultural comparisons of the content of SF-36 translations across 10 countries: Results from the IQOLA Project. *J Clin Epidemiol* 1998;51:925-932.
- Jenkinson C, Stewart-Brown S, Petersen S, *et al.* Assessment of the SF-36 version 2 in the United Kingdom. *J Epidemiol Community Health* 1999;53:46-50.
- Taft C, Karlsson J, Sullivan M. Performance of the Swedish SF-36 version 2.0. *Qual Life Res* 2004;13:251-256.
- Ware JE Jr., Kosinski MA, Dewey JE. How to score version 2 of the SF-36 health survey. Lincoln: Quality Metric Inc., 2000.
- Hawthorne G, Osborne RH, Taylor A, *et al.* The SF36 Version 2: critical analyses of population weights, scoring algorithms and population norms. *Qual Life Res* 2007;16:661-673.
- Lam CLK, Fong DY, Lauder IJ, *et al.* The effect of health-related quality of life (HRQOL) on health service utilisation of a Chinese population. *Soc Sci Med* 2002;55:1635-1646.
- Ware JE Jr., Kosinski M, Gandek B, *et al.* The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project. *J Clin Epidemiol* 1998;51:1159-1165.
- Lam CLK, Tse EYY, Gandek B, *et al.* The SF-36 summary scales were valid, reliable, and equivalent in a Chinese population. *J Clin Epidemiol* 2005;58:815-822.
- Ware JE Jr., Kosinski M. SF-36 Physical & Mental Health Summary Scales: A Manual for Users of Version 1. Lincoln, Rhode Island: QualityMetric, 2001.
- Steiger JH. Tests for comparing elements of a correlation matrix. *Psychol Bull* 1980; 87: 245-251.

Appendix: Summary of Difference between Versions 1 and 2 of the Chinese (HK) SF-36 Health Survey

Question Number

Version 1

3, introduction
The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

下列各項是您日常生活中可能進行的活動。以您目前的健康狀況，您在進行這些活動時，有沒有受到限制？如果有的話，程度如何？

4, introduction

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

在過去四個星期裏，您在工作或其它日常活動中，會不會因為身體健康的原因而遇到下列的問題？

4, response choices

Yes/ No
會 / 不會

5, introduction

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

在過去的四個星期裏，您在工作或其它日常活動中，會不會由於情緒方面的原因(比如感到沮喪或焦慮)遇到下列的問題？

5, response choices

Yes/ No
會 / 不會

Version 2

The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

下列問題是關於您日常生活中可能進行的活動。以您目前的健康狀況，您在進行這些活動時，有沒有受到限制？如果有的話，程度如何？

During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

在過去四個星期裏，您在工作或其它日常活動中，有多少時間會因為身體健康的原因而遇到下列的問題？

All of the time/ Most of the time/ Some of the time/ A little of the time/
None of the time
常常如此 / 大部分時間 / 有時 / 偶爾 / 從來沒有

During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

在過去四個星期裏，您在工作或其它日常活動中，有多少時間會由於情緒方面的原因(比如感到沮喪或焦慮)遇到下列的問題？

All of the time/ Most of the time/ Some of the time/ A little of the time/
None of the time
常常如此 / 大部分時間 / 有時 / 偶爾 / 從來沒有

All of the time/ Most of the time/ Some of the time/ A little of the time/
None of the time

常常如此 / 大部分時間 / 有時 / 偶爾 / 從來沒有
Have you been happy?
您感到快樂？

9, response choices
All of the Time/ Most of the Time/ A good bit of the time/Some of the Time/ A Little of the Time/ None of the Time

常常如此 / 大部分時間 / 相當多時間 / 有時 / 偶爾 / 從來沒有
Have you been a happy person?
您是個快樂的人？