Original Article

Validity of a Revised Short Form-12 Health Survey Version 2 in Different Ethnic Populations

Maudrene LS Tan, [1]*MSc*, Hwee Lin Wee,* [2,3]*PhD*, Agus Salim, [4]*PhD*, Jeannette Lee, [1]*PhD*, Stefan Ma, [5]*PhD*, Derrick Heng, [5]*PhD*, E-Shyong Tai, [1,6]*PhD*, Julian Thumboo,* [3,6]*FAMS (Rheumatology)*

Abstract

Introduction: The Short Form-12 version 2 (SF-12v2) is a shorter version of the Short Form-36 version 2 (SF-36v2) for assessing health-related quality of life. As the SF-12v2 could not be resolved into the physical- and mental-component summary score (PCS and MCS, respectively) in the general population of Singapore, this study aims to determine and validate the Singapore SF-12 version 2 (SG-12v2). Materials and Methods: The SG-12v2 was generated using the same methodology as the SF-12v2. Bootstrap analysis was used to determine if the SG-12v2 were significantly different from the SF-12v2. Content validity was assessed using percentage of variance (R^2) of the Singapore version of SF-36v2 PCS and MCS explained by the SG-12v2 items. Agreement between the SF-36v2 and the SG-12v2 was assessed using Bland-Altman diagrams. Criterion validity was demonstrated if effect size differences between SF-36v2 and SG-12v2 were small (Cohen's criteria). Known-group validity of SG-12v2 was reported for participants with and without chronic diseases. Results: Five items differed between the SG-12v2 and SF-12v2. Bootstrap analysis confirmed that SG-12v2 and SF-12v2 were significantly different. The SG12v2 explained 94% and 79% of the R^2 of the SF-36v2 PCS and MCS, respectively. Agreement was good and effect size differences were small (<0.3). Participants with chronic diseases reported lower SG-12v2 scores compared to participants without chronic diseases. Conclusion: The SG-12v2 offers advantage over the SF-12v2 for use in the general population of Singapore. The SG-12v2 is a valid measure and will be particularly useful for large population health surveys in Singapore.

Ann Acad Med Singapore 2016;45:228-36

Key words: Health-related quality of life, Singapore, Bland-Altman, Bootstrap

Introduction

The Short Form-36 version 2 (SF-36v2) health-related quality of life (HRQoL) survey is a well established 36-item instrument that had been used in many clinical and epidemiologic studies to assess the quality of life of subjects all around the world. Even though it consists of only 36 items which can be completed within 5 to 10 minutes, in certain studies, it might be considered too lengthy.[1] The SF-12 version 2 (SF-12v2) health survey was developed in the United States as a shorter alternative to the SF-36v2.[1] It represents a subset of items from the SF-36v2 that adequately captures the variance in the HRQoL as measured by the SF-36v2 and can be used in large scale studies where researchers are interested in health states that may have different effects on overall physical and mental health outcomes.

With longer life expectancy, HRQoL increasingly becomes an important topic of study among researchers worldwide. Between 2006 and 2015, 8 studies in Singapore were published using the SF-12v2 to measure HRQoL in specific populations[2-9] and we believe that the number of studies using the SF-12v2 will continue to increase as the

[1]Saw Swee Hock School of Public Health, National University of Singapore, Singapore
[2]Department of Pharmacy, National University of Singapore, Singapore
[3]Department of Rheumatology & Immunology, Singapore General Hospital, Singapore
[4]Department of Mathematics and Statistics, La Trobe University, Australia
[5]Epidemiology and Disease Control Division, Ministry of Health, Singapore
[6]Department of Medicine, National University Health System, Singapore
Address for Correspondence: Prof Julian Thumboo, Department of Rheumatology and Immunology, Singapore General Hospital, Outram Road, Singapore 169608.
Email: julian.thumboo@sgh.com.sg

SF-12v2 reduces the burden of completion compared to the SF-36v2. However, when we analysed the SF-12v2 in the Singapore general population, it was observed that the SF-12v2 could not be resolved into the expected physical- and mental-component summary score (PCS and MCS respectively) unlike the SF-36v2. Thus, the aim of this study was to identify the most suitable items to contribute to the Singapore SF-12v2 (SG-12v2) and to determine its validity and reliability for the multi-ethnic population of Singapore. The SG-12v2 should: 1) be representative and adequate in explaining the SF-36v2 PCS and MCS scores (content validity); 2) give similar PCS and MCS scores as the SF-36v2 instrument (criterion validity); and 3) detect differences in PCS and MCS between individuals with and without chronic medical conditions (i.e. known group validity).

Materials and Methods

Study Design and Participants

A total of 10,747 participants from 4 previous cross-sectional surveys (from 1982 to 1998) were contacted between 2004 and 2007. Of these, 7188 participants were included in the analysis. The description of the demographics can be seen in Table 1. Detailed sample selection methods for the original studies have been described elsewhere.[10-13] Briefly, all studies were a random sample of participants from the Singapore population, with disproportionate sampling stratified by ethnicity to increase the numbers from the minority ethnic groups (Malays and Asian Indians). Questionnaires were administered at the subjects' homes and all interviewed participants were invited to attend a clinical examination for additional tests and collection of biological specimens, shortly after the home visit. Ethics approval was obtained from 2 institutional review boards before study commencement. Informed consent was obtained before conducting the study.

Data Collection

Data on demographic factors and medical history were collected using interviewer-administered questionnaires. Ethnic group was self-reported and participants were classified as being Chinese, Malay or Asian Indian. Marital status was classified as 'never married', 'currently married' or 'separated/divorced/widowed'. Education level was determined through the number of years of schooling and was divided into 3 categories: <7 years, 7 to 10 years and >10 years. Working status was classified as 'working' or 'not working'. Participants were categorised as 'current smoker' or 'current drinker' if they answered 'yes' to the question, 'Do you smoke now?' or 'Have you consumed alcohol within the past three (03) months?', respectively.

Table 1. Characteristics of 7188 Subjects who Completed Either English or Chinese SF-36v2 in the Study

	English n = 6026	**Chinese n = 1162**
Age in years, mean (SD)	48.2 (12.5)	55.4 (11.2)
Sex, no. (%)		
Men	2917 (48.4)	492 (42.3)
Women	3109 (51.6)	670 (57.7)
Race, no. (%)		
Chinese	3704 (61.5)	1162 (100)
Malay	1166 (19.4)	
Indian	1156 (19.2)	
Marital status, no. (%)		
Never married	887 (14.7)	94 (8.1)
Currently married	4656 (77.3)	979 (84.3)
Separated/divorced/widowed	483 (8.0)	89 (7.7)
Years of education, no. (%)		
<7 years	1320 (21.9)	618 (53.2)
7 to 10 years	2165 (35.9)	400 (34.4)
>10 years	2541 (42.2)	144 (12.4)
Employed, no. (%)	4246 (70.5)	641 (55.2)
Housing type, no. (%)		
1 to 3 room flat	984 (16.3)	283 (24.4)
4 to 5 room flat	4019 (66.7)	771 (66.4)
Private condo/landed property	1023 (17.0)	108 (9.3)
Smoke, no. (%)	777 (12.9)	139 (12.0)
Drink, no. (%)	1400 (23.2)	226 (19.5)
Presence of disease, no. (%)	3344 (55.5)	746 (64.2)
Family functioning measure, mean (SD)	59.2 (18.2)	56.0 (16.7)

History of chronic diseases was captured through self-report data. Participants were asked whether they had ever been told that they had hypertension, diabetes mellitus or high cholesterol. History of coronary heart disease was defined as a positive response to any of the 3 questions, 'Has your doctor ever told you that you have blockage of the arteries to your heart?' or 'Have you had ever had a heart attack?' or 'Have you ever had angioplasty-ballooning or heart bypass operation procedures?' Participants were also asked whether they had ever been told by a physician that they had a cerebrovascular accident (stroke). Information on other chronic diseases (lung disease, cancer, musculoskeletal illness and mental illness) was also captured. For the health examination, participants were examined in the morning after a 10-hour overnight fast. Details of health examination, blood draw, sample preparation and biochemical analyses were previously published.[14]

Short Form-36 Version 2 (SF-36v2)

The Singapore version of the SF-36v2 was available in English, Chinese, Malay and Tamil. In this study, we analysed the data only for English and Chinese language SF-36v2 due to the small number of participants who completed the survey in Malay and Tamil. Data was pooled for the English and Chinese language surveys because it had been shown previously that the 2 languages were equivalent in our population.[15]

Construction of Singapore 12 Version 2

Individual items were recoded, summed and transformed as recommended in the SF-12v2 user manual.[16] In particular, the General Health (GH) item needed to be recalibrated to satisfy the assumption of a linear relationship between item scores and the latent trait defined by their scales. Participants with missing item scores, sociodemographic and clinical data were excluded listwise from the analysis.

We identified the 12 items that would optimally be used as the SG-12v2 questionnaire in 2 steps using the methodology described in the SF-12v2 manual:[16] First, we used forward stepwise regressions of the SF-36v2 items on Singapore SF-36v2 PCS and MCS to select items with the largest variance explained (R^2) on the 2 summary scales. By definition, 2 items each from physical functioning (PF) and mental health (MH) scales were selected and 1 item each from role-physical (RP), bodily pain (BP), social functioning (SF), and role-emotional (RE) scales was selected. Second, forward stepwise regressions were conducted by adding general health item, 'In general, would you say your health was: (GH1)' and the items from the first forward regression. The combination of items from RP, RE and vitality (VT) that explained the greatest variation on the Singapore SF-36v2 PCS and MCS scores would determine the makeup of the SG-12v2. It is an international quality of life assessment (IQOLA) criterion that GH1 be included in all country-specific questionnaires because of its use as a single-item overall health measure in many HRQoL instruments.[1] We then compared the items selected against items selected in various other countries derived from the published literature.[1,17]

Non-parametric bootstrapping was conducted, after the SG12v2 items were identified, to determine if there was a real difference between the SG-12v2 and SF-12v2 in their ability to explain variation (R^2) in PCS and MCS scores. Since the SG-12v2 items were selected using the same data used to calculate the R^2, an over-fitting factor was calculated to adjust the 95% confidence intervals (CI) of the difference in R^2 between the SG-12v2 and SF-12v2 items. The bootstrap and over-fitting factor procedures are described in Appendix 1. Upon adjustment of the 95% CI, should the 95% CI contain the value 0, it would mean that there was no difference between the 2 sets of instruments and that the SF-12v2 might be applicable to the multi-ethnic Asian Singapore population.[18]

Statistical Analyses

The bootstrap procedure was conducted using R version 2.14.2 (R Development Core Team, 2012)[19] whilst the rest of the data analyses were performed using Stata version 10 (StataCorp LP).

Assessment of Content, Criterion and Known-Group Validity of the SG12v2

We assessed content validity in 2 ways: First, by the percentage of variance (R^2) of the SF-36v2 PCS and MCS scores explained by the SG-12v2 items. The expected standard was ≥ 0.9;[1,16] Second, Pearson correlations between the SG-12v2 and the SF-36v2 PCS and MCS are expected to achieve ≥ 0.9.[1,16] Bland-Altman plots were used to enable us to visually assess the agreement between the SG-12v2 and the SF-36v2 PCS and MCS values. Criterion validity was determined using effect size differences between the SF-36v2 and SG-12v2. Effect size difference was calculated by dividing the differences in the mean scores by the standard deviation (SD) of the SF-36v2 summary score.[20] The SG-12v2 is considered to give similar results as the SF-36v2 if the effect size difference is smaller than the minimum important difference (MID; i.e., Cohen's effect size of 0.3-0.5).[17,21,22] We extended the evaluation of effect size difference to patients with and without a specific disease to determine how the instrument performed for the various groups. This would also provide evidence for known-group validity of the SG-12v2, where we expect participants without chronic diseases to have higher PCS and MCS compared to participants with any chronic diseases.

Comparison of the Singapore and the United States' (US) Instrument

Items from the 2 instruments were compared to determine the differences. Exploratory factor analyses was also conducted, for each of the instrument, and compared. The 8 domains of the SF-36v2 were expected to resolve into 2 factors, with PF and MH expected to have positive and strong association with their respective summary measures and negative association with the other summary measure.

Results

Content, Criterion and Known-Group Validity of SG12v2

The SG-12v2 explained 94% and 79% of the total variance of the SF-36v2 PCS and MCS, respectively (Table 2). The

Table 2. Items Comparison between the Singapore 12 Version 2 and the SF-12 Version 2

Singapore			**SF-12v2**		
Domain Item No.	SF-36 Item No.	Item Wording	Domain Item No.	SF-36 Item No.	Item Wording
PF3*	3c	Lifting/carrying groceries	PF2	3b	Moderate activities
PF8*	3h	Walking several hundred metres	PF4	3d	Climbing several flights
RP2	4b	Accomplished less	RP2	4b	Accomplished less
RP3	4c	Limited in kind of work	RP3	4c	Limited in kind of work
BP2	8	Extent pain interfered with work	BP2	8	Extent pain interfered with work
GH1	1	Your health is excellent…poor	GH1	1	Your health is excellent…poor
VT4*	9i	Feel tired	VT2	9e	Have a lot of energy
SF1*	6	Extent social activities interfered	SF2	10	Frequency social activities interfered
RE2	5b	Accomplish less	RE2	5b	Accomplish less
RE3	5c	Didn't do work as carefully	RE3	5c	Didn't do work as carefully
MH3	9d	Felt calm and peaceful	MH3	9d	Felt calm and peaceful
MH5*	9h	Been a happy person	MH4	9f	Felt downhearted and low
PCS R^2	MCS R^2				
0.938	0.788				

MCS: Mental-component summary; PCS: Physical-component summary
*Item differs from the SF-12v2.

Pearson correlations between SG-12v2 and the SF-36v2 PCS and MCS achieved the expected standard of 0.9 (PCS: 0.96; MCS: 0.88). In addition, we plotted Bland-Altman diagrams to assess the agreement between the SG-12v2 and SF-36v2 PCS and MCS (Figs. 1 and 2). The plots showed that a majority of the data spread for both PCS and MCS was within the 95% confidence interval (CI) band, indicating good agreement between the 2 instruments.

Effect size difference between the SG-12v2 and the SF-36v2 scores were conducted across several chronic disease groups in Table 3. The SG-12v2 and SF-36v2 detected similar significant differences between each of the chronic disease group as well as the 'no chronic disease' group. The largest effect size between the SG-12v2 and SF-36v2 scoring algorithms was the MCS score for subjects reporting psychological diseases, with a value

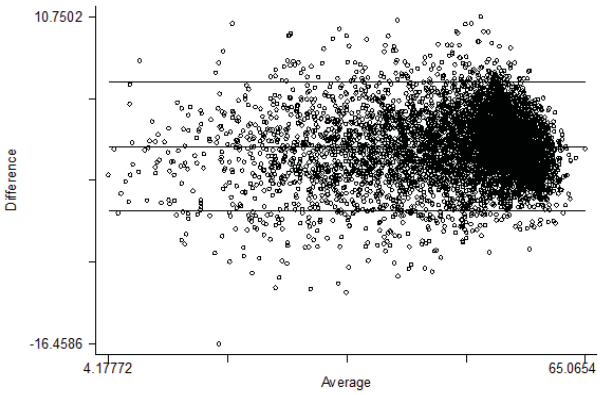

Fig. 1. Bland-Altman diagram shows the measure of agreement between Singapore SF-12v2 and SF-36v2 physical-component summary (PCS). Majority of the data lies within 95% confidence band indicating good agreement between the 2 instruments. Pitman's test of difference in variance: r = 0.063, indicating concordance between the 2 instruments.

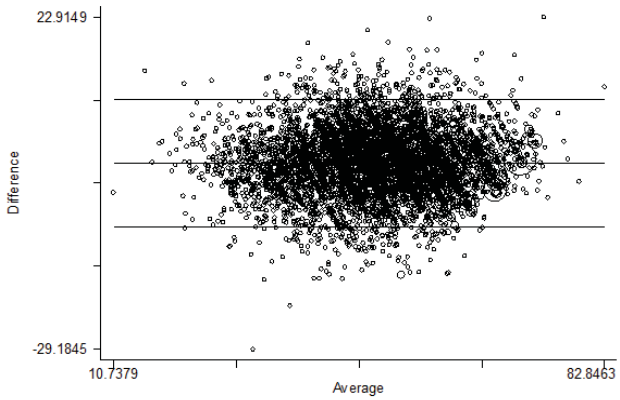

Fig. 2. Bland-Altman diagram shows the measure of agreement between Singapore SF-12v2 and SF-36v2 mental-component summary (MCS). Majority of the data lies within 95% confidence band indicating good agreement between the 2 instruments. Pitman's test of difference in variance: r = 0.004, indicating concordance between the 2 instruments.

Table 3. Effect Size Comparison between Singapore SF-12v2 and SF-36v2 Scores across Chronic Disease Groups

	PCS		MCS	
	Mean	**SD**	**Mean**	**SD**
All subjects (n = 7188)				
SF-36	50.1	9.8	50.0	10.0
SF-12	50.0	9.9	50.0	10.0
Effect size*	0.01		0.00	
No chronic disease (n = 3098)				
SF-36	52.2	8.1	50.6	9.6
SF-12	51.9	8.3	50.9	9.4
Effect size*	0.04		0.03	
Any chronic disease (n = 4090)				
SF-36	48. 5	10.6	49.5	10.2
SF-12	48.6	10.8	49.3	10.4
Effect size*	0.01		0.02	
Cardiovascular disease (n = 3276)				
SF-36	48.7	10.6	50.0	10.3
SF-12	48.9	10.8	49.8	10.4
Effect size*	0.02		0.02	
Stroke (n = 111)				
SF-36	40.0	14.0	46.8	13.0
SF-12	40.3	14.3	46.3	11.9
Effect size*	0.02		0.04	
Diabetes mellitus (n = 754)				
SF-36	46.6	12.4	48.0	11.0
SF-12	46.9	12.8	47.5	11.1
Effect size*	0.02		0.04	
Psychological diseases (n = 74)				
SF-36	38.1	14.3	42.3	11.7
SF-12	37.0	15.0	43.8	12.2
Effect size*	0.08		0.13	
Pulmonary disease (n = 336)				
SF-36	46.7	11.8	48.5	10.5
SF-12	47.0	11.9	48.2	10.7
Effect size*	0.03		0.03	
Joint disease (n = 1535)				
SF-36	45.7	11.5	47.7	9.8
SF-12	45.7	12.0	47.6	9.9
Effect size*	0.01		0.01	

MCS: Mental-component summary; PCS: Physical-component summary
*Calculated by taking the difference between the mean SF-36v2 and SF-12v2 and dividing by the standard deviation (SD) of the SF-36v2.

of 0.13, which nonetheless was smaller than the definition of MID. In terms of known-group validity, as expected, subjects who reported as having some form of disease had lower PCS and MCS scores as compared to the subjects who indicated 'no chronic disease'.

Comparison of the Singapore and US Instruments

The composition of SG-12v2 and SF-12v2 items can be found in Table 2. Five items (in asterisk) from the SG-12v2 differed from the SF-12v2. Results from bootstrap analysis confirmed that the items from SG-12v2 were indeed different

Table 4. Factor Score Coefficients‡ for the Singapore SF-12v2, the original SF-12v2 and the SF-36v2 PCS and MCS

Scales†	Factor Score Coefficients					
	SF-36v2		SG-12v2		SF-12v2	
	PCS	MCS	PCS	MCS	PCS	MCS
Physical functioning	0.251*	-0.0618	0.273*	-0.121	0.0314	0.250*
RP	0.349*	-0.147	0.305*	-0.101	0.341*	-0.107
BP	0.251*	-0.0609	0.245*	-0.0258	0.294*	-0.0859
GH	-0.187	0.464*	-0.143	0.528*	-0.179	0.471*
Vitality	-0.146	0.457*	-0.0473	0.413*	-0.124	0.451*
Social functioning	0.221*	0.0204	0.215*	0.0316	0.241*	0.0128
RE	0.323*	-0.124	0.298*	-0.101	0.373*	-0.172
Mental health	-0.093	0.389*	-0.105	0.507*	0.0269	0.260*

BP: Bodily pain; GH: General health; MCS: Mental-component summary; PCS: Physical-component summary; RE: Role-emotional; RP: Role-physical

*Indicate the higher weightage placed on the respective summary score.

†8 Domains of the SF-36v2.

‡Obtained from exploratory principal component factor analysis with Varimax rotation and used to compute the Physical- and Mental-component scores.

from the SF-12v2, since the 95% CI did not contain the value 0 for both PCS and MCS.

Table 4 shows the results of the factor weights obtained from the exploratory principal component factor analysis for the 2 countries. Only the 12 items from Singapore were able to converge into the PCS and MCS. When we selected the same items as the SF-12v2 and subjected them to exploratory principal component factor analysis, 2 factors were derived but could not be meaningfully labelled as PCS and MCS because both PF and MH load on to the same factor.

Discussion

Our study showed that the 12 items selected to optimally represent the PCS and MCS of the SF-36v2 differed by 5 items from those in the US. This was confirmed through the bootstrap analysis. The SG-12v2 showed good content, criterion and known-group validity and attained good agreement with the SF-36v2 with no evidence of a systematic bias.

In trying to understand why the SF-12v2 could not converge into the summary component scores, we replaced 1 of the 2 original SF-12v2 PF items with items from the SF-36v2 PF scale. We found that as long as the following items were not simultaneously selected, the modified SF-12v2 would resolve into PCS and MCS (results available on request): 'Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports' (vigorous activities), 'Moderate activities, such as moving table, pushing a vacuum cleaner, bowling, or playing golf' (moderate activities) and 'Climbing several flights of stairs' (climbing stairs). Two possible explanations could help explain our findings. First, although these 3 items measure physical activities, these activities require a certain level of mental toughness. Gerber et al, had previously showed a significant association between higher levels of physical activities and increased mental toughness scores.[23] Second, Singaporeans were found to be generally inactive, with activity levels below the recommended levels of physical activity (Fig. 3).[24] Hence, these 3 items might be more difficult for respondents in the Singapore context.

In the mental health scale, we found that 1 of the 2 mental health items differed between Singapore and the original instrument. We suspect that the reason why the item 'During the past 4 weeks, have you felt downhearted and depressed' was not appropriate in determining the mental health status of our local population was because Singaporeans, being Asians, still hold strongly to their culture and roots (depending on their ethnic background). As such, mental health issues are regarded as taboo topics for fear of facing discriminations by others.[25,26]

An important strength of this study lies in the bootstrap analyses. To our knowledge, this is the first study to use a 1000 bootstrap with replacement analysis to measure the difference in R^2 between a localised version of SF-12v2 and the original instrument. Furthermore, to ensure that the data was not over-fitted, we adjusted the 95% CI with an over-fitting factor. The fact that the results still showed that the SG-12v2 and the SF-12v2 were significantly different accentuated the need for a local version of the SF-12v2. Our findings are important for researchers and clinicians

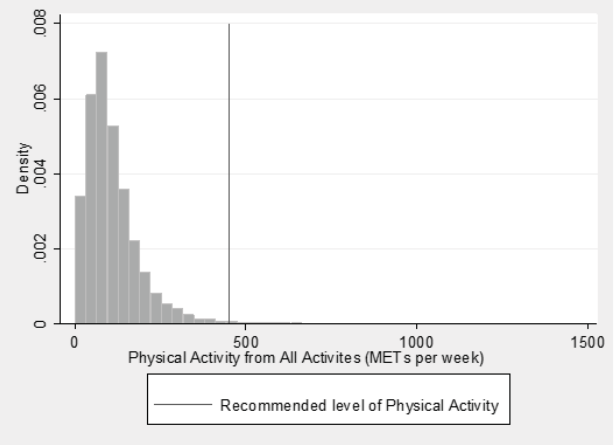

Fig. 3. Graph showing the levels of physical activity engaged by the 7188 Singaporean respondents. All activities: includes Transportation, Household, Leisure and Occupation activities. Recommended level of Physical Activity = 450 METs per week (Source: Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007;39:1423-34). MET: Metabolic equivalent of task.

who are interested in studying HRQoL in Singapore and around Southeast Asia.

This study is not without limitations. First, we restricted the item selection methodology to match what was used to develop the original SF-12. An alternative method might have been to use item response theory to customise the instrument to the population studied.[27] Nonetheless, we chose to follow the same method as the original SF-12 to ensure that we did not deviate too much from the original intent of the developers.

Conclusion

In this multi-ethnic Asian population, the SG-12v2 had been showed to be as good a measure when compared to the SF-36v2. In addition, it is a necessary replacement of the SF-12v2 since the SF-12v2 cannot be resolved into PCS and MCS. Thus, in local clinical trials and large population-based studies, should the SF-36v2 be deemed too long an instrument to be used, researchers can consider using the SG-12v2 as a substitute instrument.

Acknowledgement

This work was supported by grants from the National Medical Research Council, Singapore (Grant number: 0838/2004, IRG07nov013, and NMRC/0863/2004); the Biomedical Research Council (Grant number: 03/1/27/18/216 and 08/1/35/19/550) and the Diabetes Research Fund, National University of Singapore.

REFERENCES

1. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998;51:1171-8.
2. Griva K, Goh CS, Kang WC, Yu ZL, Chan MC, Wu SY, et al. Quality of life and emotional distress in patients and burden in caregivers: a comparison between assisted peritoneal dialysis and self-care peritoneal dialysis. Qual Life Res 2016;25:373-84.
3. Yang F, Wang VW, Joshi VD, Lau TW, Luo N. Validation of the English version of the Kidney Disease Quality of Life questionnaire (KDQOL-36) in haemodialysis patients in Singapore. Patient 2013;6:135-41.
4. Ho RC, Fu EH, Chua AN, Cheak AA, Mak A. Clinical and psychosocial factors associated with depression and anxiety in Singaporean patients with rheumatoid arthritis. Int J Rheum Dis 2011;14:37-47.
5. Lim L, Jin AZ, Ng TP. Anxiety and depression, chronic physical conditions, and quality of life in an urban population sample study. Soc Psychiatry Psychiatr Epidemiol 2012;47:1047-53.
6. Yap KB, Niti M, Ng TP. Nutrition screening among community-dwelling older adults in Singapore. Singapore Med J 2007;48:911-6.
7. Luo N, Wang P, Fu AZ, Johnson JA, Coons SJ. Preference-based SF-6D scores derived from the SF-36 and SF-12 have different discriminative power in a population health survey. Med Care 2012;50:627-32.
8. Ho RC, Giam YC, Ng TP, Mak A, Goh D, Zhang MW, et al. The influence of childhood atopic dermatitis on health of mothers, and its impact on Asian families. Pediatr Allergy Immunol 2010;21:501-7.
9. Ng TP, Feng L, Chiam PC, Kua EH. Psychiatric morbidity and acute hospitalization in elderly people. Int Psychogeriatr 2006;18:701-11.
10. Cutter J, Tan BY, Chew SK. Levels of cardiovascular disease risk factors in Singapore following a national intervention programme. Bull World Health Organ 2001;79:908-15.
11. Hughes K, Aw TC, Kuperan P, Choo M. Central obesity, insulin resistance, syndrome X, lipoprotein(a), and cardiovascular risk in Indians, Malays, and Chinese in Singapore. J Epidemiol Community Health 1997;51:394-9.
12. Hughes K, Yeo PP, Lun KC, Thai AC, Sothy SP, Wang KW, et al. Cardiovascular diseases in Chinese, Malays, and Indians in Singapore. II. Differences in risk factor levels. J Epidemiol Community Health 1990;44:29-35.
13. Tan CE, Emmanuel SC, Tan BY, Jacob E. Prevalence of diabetes and ethnic differences in cardiovascular risk factors. The 1992 Singapore National Health Survey. Diabetes Care 1999;22:241-7.
14. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III).Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation 2002;106:3143-421.
15. Tan ML, Wee HL, Lee J, Ma S, Heng D, Tai ES, et al. The Short Form 36 English and Chinese versions were equivalent in a multiethnic Asian population. J Clin Epidemiol 2013;66:759-67.

16. Ware JE KM, Keller SD. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. 2nd ed. Boston, MA: The Health Institute, New England Medical Center; 1995.
17. Lam CL, Tse EY, Gandek B. Is the standard SF-12 health survey valid and equivalent for a Chinese population? Qual Life Res 2005;14:539-47.
18. Julious S. Using confidence intervals around individual means to assess statistical significance between two means. Pharm Stat 2004:217-22.
19. R Development Core Team (2011). R: A Language and Environment for Statistical Computing. Vienna, Austria: the R Foundation for Statistical Computing; 2012.
20. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27:S178-89.
21. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1988.
22. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care 1999;37:469-78.
23. Gerbera M, Kalakb N, Lemolac S, Cloughd P, Pühsea U, Elliota C, et al. Adolescents' exercise and physical activity are associated with mental toughness. Ment Health Phys Act 2012;5:35-42.
24. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007;39:1423-34.
25. Peng T. Mental Illness May Go Untreated in Asian-Americans. Newsweek. Available at: www.newsweek.com/mental-illness-may-go-untreated-asian-americans-87613. Accessed on 7 January 2016.
26. Yang K. Chinese personality and its change. In: Bond MH, editor. The psychology of the Chinese people. Hong Kong: Oxford University Press; 1986.
27. Lai JS, Cella D, Chang CH, Bode RK, Heinemann AW. Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item bank from the FACIT-Fatigue Scale Qual Life Res 2003;12:485-501.

Appendix 1

The bootstrap procedure is as follows:	
0.	Start with K = 1.
1.	Generate a bootstrap sample with replacement from observed data.
2.	Fit to the bootstrap sample, the Singapore PCS and MCS models using regression coefficients that have been optimised from the observed data. Call the R^2 from these fits $r^{2(K)*}_{PCS}$ and $r^{2(K)*}_{MCS}$.
3.	Using the bootstrap sample, refit the Singapore PCS and MCS models. Call the R^2 from these fits $r^{2(K)}_{PCS}$ and $r^{2(K)}_{MCS}$.
4.	Fit the models obtained in step 3 onto the observed data. Call the R^2 from these fits $r^{2(K)PRED}_{PCS}$ and $r^{2(K)PRED}_{MCS}$.
5.	Calculate $\Delta r^{2(K)*}_{PCS} = r^{2(K)*}_{PCS} - r^2_{PCS,US}$ and $\Delta r^{2(K)*}_{MCS} = r^{2(K)*}_{MCS} - r^2_{MCS,US}$ where $r^2_{PCS,US}$ and $r^2_{MCS,US}$ are r-sq for US models from the observed data.
6.	Calculate the over-fitting factor for PCS as $\eta^{(K)}_{PCS} = r^{2(K)}_{PCS} - r^{2(K)PRED}_{PCS}$. Do similarly for MCS model.
7.	Redo steps 1 to 7 for K = 2,3…1000.

MCS: Mental-component summary score; PCS: Physical-component summary score; US: United States

Calculate the average over-fitting factor for PCS model as $\eta PCS = ave(\eta^{(K)}_{PCS})$ and calculate the bootstrap estimate of 95% confidence intervals for Δr^2_{PCS} using the percentile method and call the lower and upper bounds as $\Delta r^{2,LB}_{PCS}$ and $\Delta r^{2,UB}_{PCS}$. The over-fitting-corrected 95% confidence intervals for Δr^2_{PCS} is then calculated as ($\Delta r^{2,LB}_{PCS} - \eta_{PCS}$; $\Delta r^{2,UB}_{PCS} - \eta_{PCS}$). Procedure for calculating the over-fitting-corrected confidence intervals for the MCS model follows similarly.