David Wilson[1], Graeme Tucker[2], Catherine Chittleborough[2]

[1] *University of Adelaide*
[2] *Department of Human Services, Adelaide*

Rethinking and rescoring the SF-12

Summary

Objectives: To derive and assess the validity of an Australian version of the SF-12 quality-of-life questionnaire.
Methods: Using regression methods and structural equation modelling to obtain item weights, an Australian version of the SF-12 was derived from Australian population survey data and compared to the existing United States (US) SF-12 variable set.
Results: The Australian version of the SF-12 explained 94% of the variation for physical components summary (PCS) and the mental components summary (MCS) of the SF-36 questionnaire. There was high level of agreement on the MCS and PCS summary scores between both versions of the SF-12 and the SF-36.
Conclusions: Although it is possible to derive a valid Australian version of the SF-12 it is concluded the US version of the SF-12 be used for reasons of international comparability, but using item weights derived from structural equation modelling.

Keywords: Quality-of-life – SF-12 – Validity.

The SF-12 is a summary quality-of-life questionnaire measuring physical and mental health with the physical component summary (PCS) and the mental component summary (MCS). It is useful for measuring these health dimensions in population or subpopulation groups. The SF-36 MCS and PCS formed the basis on which the shorter version SF-12 question items were derived in the United States (Ware et al. 1996). The 12 items chosen as the US SF-12 achieved R^2 values of 0.91 for the PCS and 0.92 for the MCS (Ware et al. 1996). Explanation of 90% of the variance in the SF-36 PCS and MCS measures in the US population were deemed to be adequate decision-making criteria in accepting a 12-item measure of quality-of-life (Ware et al. 1995). This was an important development for researchers conducting surveys and other studies in which time and cost are important in measuring health status. The SF-12 is an instrument that can be administered in three minutes with a small trade off between brevity and precision.

A recent study has, however, questioned the scoring of the SF-36 PCS and MCS (Wilson et al. 2000). This showed that when coefficients are derived using structural equation modelling (rather than by principal components analysis and orthogonal rotation) and fit to South Australian population survey data in a confirmatory factor analysis, the summary PCS and MCS more accurately reflect the underlying subscales (general health, physical health, role physical, bodily pain, role emotional, vitality, social health, and mental health) from which they are derived. The structural equation method produced SF-36 summary scale scores which had good validity across age and implied health groups when compared with scores produced by orthogonal rotation (Wilson et al. 2000). The work of rescoring the SF-36 using structural equation modelling followed on from the previous work of Simon et al. who recommended caution in the interpretation of the PCS and MCS when the condition or treatment of interest had strong effects on scales with negative scoring coefficients (1998). This means that the derivation of the SF-12 question items and the construction of this shorter quality-of-life instrument was based on problematic SF-36 summary scores. Given this previous work this study used Australian population data to derive the SF-12 on the basis of an SF-36 that was scored using structural equation modelling. As a result of this and other factors a version of the SF-12 was produced that differed to some extent in the questions that formed the US SF-12.

Soz.- Präventivmed. 47 (2002) 172–177
0303-8408/02/030172-06 $ 1.50 + 0.20/0

Historically the US version of the SF-12, using US regression weights, has been used in Australian population studies. This is inappropriate given that the weights were derived from US survey data and are now of a considerable age. An earlier attempt to develop an Australian SF-12 failed on the basis that an instrument could not be derived that adequately explained the variance of the SF-36 summary scales (McCallum 1996). Since then, however, higher quality Australian population data has become available and was used in this study to revisit the issue. The research questions asked were: 1) can a valid SF-12 be derived that adequately explains the SF-36 summary scales based on structural equation modelling, and 2) how does this Australian version of the SF-12 compare with the US version in both question content and in assessing the health of various age and health groups.

Method

Two population data sources were used for this study. The first was the 1995 Australian National Health Survey (NHS) (1995) which was the second in a series of five-yearly population surveys designed to obtain national benchmark information on a range of health issues. The sampling method of the NHS is a self-weighting multistage clustered area sample based on Australian Bureau of Statistics (ABS) collector districts in which households are selected with equal probability. In this survey n = 23800 households were selected and all adults aged 15 years or older were interviewed. A subset of n = 19785 were asked to complete the SF-36 quality-of-life questionnaire. Of those interviewed, n = 18141 provided sufficient data to be included in the analyses of the SF-36 subscale and summary scores (Ware et al. 1993). These data were used in the paper by Wilson et al. to rescore the SF-36 summary scales using structural equation modelling (2000). The NHS data were again used in this study to derive an Australian SF-12 that best explained the SF-36 PCS and MCS and also to produce regression coefficients that could be used to score the US SF-12. Having derived the Australian SF-12 items from the NHS data set the South Australian Health Omnibus (SAHOS) data set was used to confirm the validity of both the Australian and US versions of the SF-12.

Derivation of the Australian SF-12 from the NHS data set occurred after imputation of data items by mean substitution, where not more than half the data items in a subscale were missing, as set out in the SF-36 scoring manual (Ware et al. 1993). Regression methods (involving backward elimination of the least significant term) were used to select from among the SF-36 items. Adjusted R^2 values were used to determine the amount of variance explained in this Australian SF-12. The NHS data set was also used to compare the variance explained in the US version of the SF-12. Because PCS and MCS are now linear combinations of contributing variables (for each dimension) (Wilson et al. 2000), a full backward elimination model could not be fitted (one term needs to be eliminated to avoid a perfect fit). By first removing one term at random, then identifying the least significant term remaining, eliminating that term and replacing it with the first term removed, backward elimination methods can then proceed as normal.

The Australian SF-12 variables newly derived from the NHS data and the US SF-12 were then fitted to data obtained in the independent 1998 SAHOS using a sample of n = 3007 (70% response rate) in a confirmatory factor analysis. The SAHOS is a statewide survey designed to obtain benchmark information on a range of health issues. The SAHOS like the NHS Survey is also a self-weighting multistage clustered area sample selected from ABS collector districts of people aged 15 years or more who live in metropolitan Adelaide or country towns with populations over 1000. The survey is conducted annually and the method has been extensively published (Wilson et al. 1992). Households are selected with equal probability of selection within each collector district and then one adult in each household, aged 15 years or older is selected for interview according to the most recent birthday. Data from the survey are weighted to accurately represent the age, sex, household size, and geographic area of the South Australian population. Having confirmed the factor structure on the SAHOS data the same models were fitted to the NHS data to produce regression coefficients representative for Australia. Wilson et al. (2000) provide full rationale for the use of structural equation modelling and details of the methods used.

In addition to the SF-36, a range of other health and demographic questions were asked in the 1998 SAHOS and from these data it was possible to validate both versions of the SF-12, with Australian regression coefficients, using a range of age and implied health groups. In the 1998 SAHOS respondents were asked whether or not in the previous twelve months they had used any medication for a chronic physical condition: that is, one that has lasted for, or is likely to last for, six months or more. They were also asked the same question for depression (medication such as tranquillisers or anti-depressants) or a diagnosed mental illness (such as schizophrenia). From these data the study population were divided into four implied health groups according to no medication, physical health medication only, mental health medication only or both physical and mental health medication. In the previous paper by Wilson et al. (2000) these

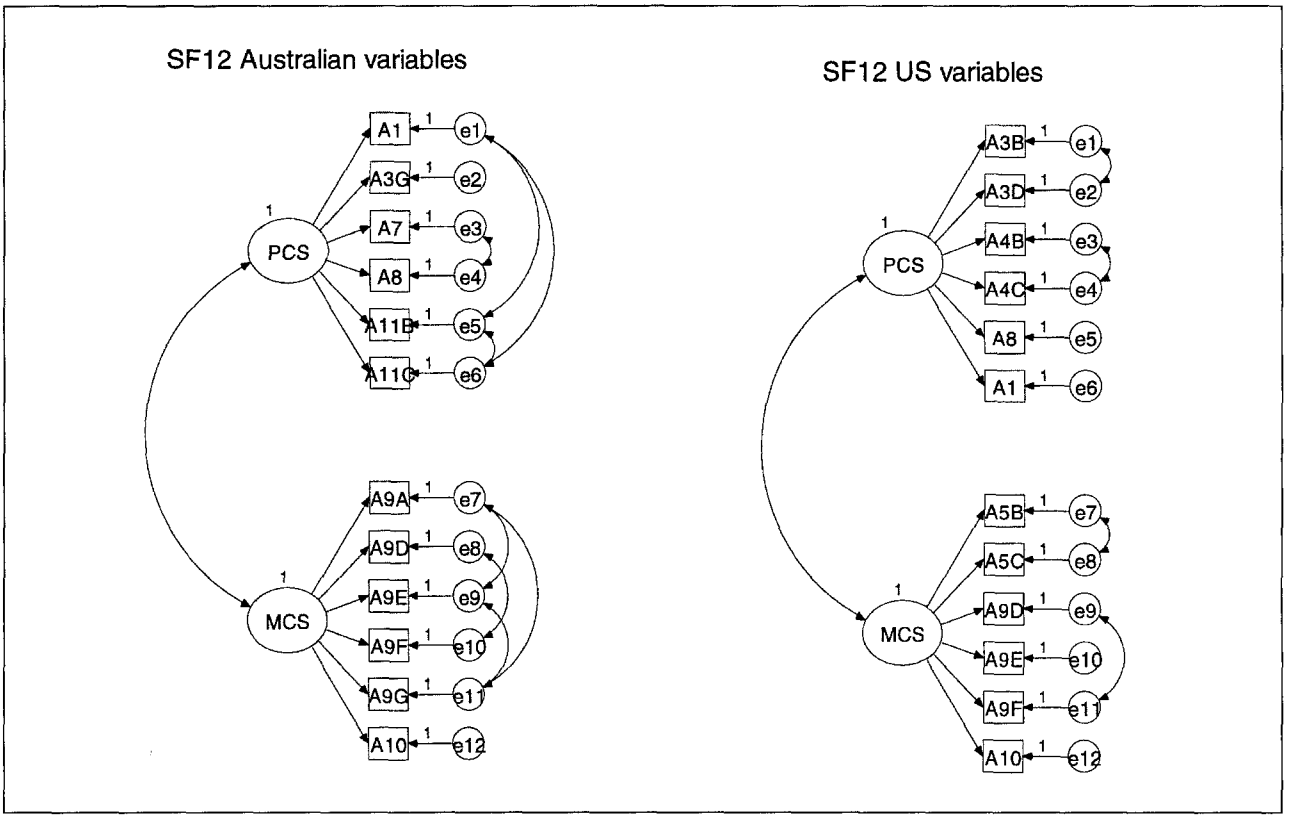

Figure 1 SF-12 model fitted

implied health groups were used to assess the external validity of two versions of the SF-36 PCS and MCS by comparing the scores with their underlying subscale scores. These versions of the PCS and MCS were derived from both structural equation modelling and principle components analyses. Scoring the SF-36 using structural equation modelling had greater validity than scoring by principal components analysis. The present study compares the Australian and US versions of the SF-12 PCS and MCS scores with the summary SF-36 PCS and MCS, using structural equation modelling, for age groups and implied health groups as an assessment of SF-12 validity.

The models fitted to the data using structural equation modelling are shown in Figure 1. The figure also shows that the covariance terms between errors were applied to items from the same subscales on the basis that these items would be expected to be more closely correlated with each other than with items from other subscales.

Table 1 SF-36 question items derived for the Australian and United States SF-12 PCS and MCs

	Australian SF-12	United States SF-12
PCS	A1 Evaluation of general health	US1 Evaluation of general health
	A8 Pain interferes with normal work	US8 Pain interferes with normal work
	A3g Walking more than a mile	US3b Moderate activities
	A7 Pain magnitude	US3d Climbing several flights
	A11b As healthy as anybody I know	US4b Accomplished less
	A11c Health to get worse	US4c Limited in kind of work
MCS	A9d Calm and peaceful	US9d Calm and peaceful
	A9e Lot of energy	US9e Lot of energy
	A9f Downhearted and blue	US9f Downhearted and blue
	A10 Social-time	US10 Social-time
	A9a Feel full of life	US5b Accomplished less
	A9g Feel worn out	US5c Not careful

Results

The twelve Australian items selected for the PCS and MCS together with their US counterparts are shown in Table 1. It can be seen that, in total, six questions from the SF-36 are

Table 2 Comparisons of Australian and United States versions of the SF-12 PCS and MCS scores with the SF-36 summary scores by age groups

	Age group				
	<30 years	30–49 years	50–69 years	70+ years	Total
SF-36 PCS	52.8	51.8	48.2	44.4	50.3
SF-36 MCS	52.1	51.1	51.3	49.9	51.2
SF-12 PCS (Aus)	52.7	51.5[a]	48.3	45.9[a]	50.4
SF-12 MCS (Aus)	52.3[a]	50.6[a]	50.4[a]	48.5[a]	50.7[a]
SF-12 PCS (US)	52.8	51.8	48.4	44.8[a]	50.4
SF-12 MCS (US)	52.5[a]	51.0	50.9[a]	49.0[a]	51.1

[a] Statistically significantly different from the SF-36 comparable scale.

Table 3 Comparisons of Australian and United States versions of the SF-12 PCS and MCS scores with the SF-36 summary scores for implied health groups

	Medication taken				
	None	Physical only	Mental only	Both	Total
SF-36 PCS	53.1	43.6	47.1	34.6	50.3
SF-36 MCS	52.8	48.9	41.6	37.3	51.2
SF-12 PCS (Aus)	53.0[a]	44.2[a]	47.7	35.6[a]	50.4
SF-12 MCS (Aus)	52.4[a]	48.0[a]	41.9	37.5	50.7[a]
SF-12 PCS (US)	53.1	44.2[a]	47.2	35.6[a]	50.4
SF-12 MCS (US)	52.8	48.5[a]	42.4	37.7	51.1

[a] Statistically significantly different from the SF-36 comparable scale.

common to both the US and Australian version of the SF-12. For the Australian SF-12 PCS and MCS the SF-36 variation explained, as determined by the adjusted R^2, was 94% for both scales. The root mean square error (RMSEA) for the Australian model was 0.07. For the US version of the SF-12 the variance explained was 91% for the PCS and 90% for the MCS. The RMSEA, which assesses the fit of the model in relation to the degrees of freedom, was 0.08 for this model. We would not employ a model with a RMSEA equal to or greater than 0.1 (Arbuckle & Wothke 1999).

Table 2 shows the SF-12 PCS and MCS age group scores calculated in this study and compared to the SF-36 summary scores calculated in the previous SF-36 study (Wilson et al. 2000). It can be seen from Table 2 that there is a high level of agreement between both versions of the SF-12 and also a high level of agreement between these and the SF-36 summary scores. Some statistically significant differences were found for some age groups for the Australian and US SF-12 PCS and MCS when compared with the SF-36 PCS and MCS However, these differences were relatively small and statistical significance reflects the large sample sizes used. Overall only the Australian SF-12 MCS was significantly different from the SF-36 MCS. The maximum *relative* differences amounted to 3% between paired scores for the Australian SF-12 MCS and SF-36 MCS in the 70 years and older age group. This was calculated by subtracting the SF-12 score from the SF-36 score and dividing by the SF-36 score. The majority of other paired differences were less than 2% and some were less than 1%.

When compared with the theoretical health groups, based on medication taken, there was again a high level of agreement between the scores of the SF-12 versions and the SF-36 summary scales. Significant differences were observed for some implied health groups, but again the maximum relative difference did not exceed 3%. The conclusion that should be reached is that despite some statistically significant differences in these SF-12 scores when compared with the SF-36 scores, these differences are not significant in practical or clinical terms.

Discussion

In this study the SF-12 reproduced the average summary scores of the SF-36 with over 90% of the variance explained using Australian regression weights. This addresses Ware's criteria for acceptance of the instrument. In selecting the best set of twelve explanatory variables for the SF-36 PCS and MCS it should be borne in mind that all question items are part of a summary scale constructed to take advantage of the high correlation between each of the variables. Thus the variables are collinear and all strongly related to the dependent variable. Selection of question items for the Australian SF-12 are therefore based on very small differences between each item and a number of different combinations explained more than 90% of the variation in the SF-36 summary scales in the Australian data. It is also highly probable that selection of an SF-12 will be data set specific. Collinearity diagnostics conducted in this study all produced conditioning index values in excess of 20 indicating collinearity between the variables in the SF-36 subsets (Gebski et al. 1992).

The MCS and PCS summary scores for both versions of the SF-12 compared well with the SF-36 summary scores produced in the previous study by Wilson et al. (2000) and had higher agreement with the underlying scale scores (validity) than did the MCS and PCS scores obtained from orthogonal rotation. We cannot assert that the validity of results obtained in this Australian study apply to other countries. It is reasonable, however, to suggest that users of the SF-36 and SF-12 in countries other than the United States use weights that are derived in those countries as has been done in this study.

Until this study the US SF-12 has been scored using regression coefficients derived in the United States. This study has corrected the situation and shown that the US SF-12 is a valid instrument in the Australian context. For other countries using the US SF-12 question items, regression weights would also best be obtained from endogenous population data.

The Australian version of the SF-12 explained more of the variance for both the PCS and the MCS than did the US version and the goodness of fit diagnostics (RMSEA) may suggest that the US version is not quite as good a model as the Australian version. Given that it does explain 90% of the variation in the SF-36 PCS and MCS with a RMSEA of 0.08 it is, however, a very adequate model. With this in mind we can conclude that the preferred version of the SF-12 for the Australian context should be the US SF-12. This is based on both the model adequacy and the fact that this version has international comparability in quality-of-life scores. In addition, despite better statistical diagnostics for the Australian variable set in these analyses, the US variable set actually performed better when compared to the SF-36 and provided smaller comparative differences in the validity study. Further support for using the US SF-12 comes from the fact that the Australian SF-12 did not cover all of the subscales in the questions derived. For the PCS role-physical was omitted and for the MCS role-emotional was omitted. Overall therefore the US SF-12 is a better instrument for both the Australian and international research context.

Zusammenfassung

Überdenken und Neuauswertung des SF-12-Fragebogens

Fragestellung: Eine australische Version des SF-12-Fragebogens zur Erfassung der Lebensqualität ableiten und dessen Validität beurteilen.

Methoden: Unter Verwendung von Regressions- und Strukturgleichungsmodellen zur Bestimmung von Gewichtungsfaktoren wurde eine australische Version des SF-12-Fragebogens abgeleitet. Hierzu wurden Daten einer Befragung der australischen Bevölkerung herangezogen. Der australische Fragebogen wurde anschliessend mit dem bereits bekannten U.S. SF-12-Variablensatz verglichen.

Ergebnisse: Die australische Version des SF-12 erklärte 94% der Variation für die physische (PCS) und psychische Summenskala (MCS) des SF-36-Fragebogens. Zwischen den beiden Versionen des SF-36 und SF-12 stimmten die MCS- und PCS-Summenwerte sehr gut überein.

Schlussfolgerungen: Es ist möglich, eine valide australische Version des SF-12 zu erhalten. Aus Gründen der internationalen Vergleichbarkeit ist es jedoch besser, die U.S. Version des SF-12 einzusetzen, aber unter Verwendung von Gewichtungsfaktoren, die anhand struktureller Vergleichsmodelle abgeleitet wurden.

Résumé

Repenser et redéfinir le score du SF-12

Objectifs: Etablir une version australienne du questionnaire SF-12 sur la qualité de vie et évaluer sa validité.

Méthode: A partir de méthodes de régression et de pondération par des modèles d'équations structurelles, une version australienne du SF-12 a été établie à partir d'une enquête de population et comparée avec l'ensemble des variables composant la version nord-américaine existante du SF-12.

Résultats: La version australienne du SF-12 expliquait 94% de la variation du score synthétique physique (PCS) et mental (MCS) du questionnaire SF-36. Il y avait un haut niveau de concordance pour ces deux scores entre les deux versions du SF-12 et du SF-36.

Conclusions: Bien qu'il soit possible d'établir une version australienne valide du SF-12, l'utilisation de la version américaine du SF-12 peut être utilisée pour des raisons de comparabilité internationale, mais en utilisant un système de pondération spécifique basé sur des modèles d'équation structurelles.

Soz.- Präventivmed. 47 (2002) 172–177

References

Arbuckle JL, Wothke W (1999). Amos 4.0 user's guide. Chicago: SmallWaters Corporation.

National Health Survey (1995). SF-36 population norms Australia. Canberra: Australian Bureau of Statistics.

Gebski V, Leung O, McNeil D, Lunn D (1992). Spida users manual. New South Wales: Statistical Computing Laboratory.

McCallum J (1996). The shorter form: analysis of SF12 items in Australian data: proceedings. Canberra: Integrating Health Outcomes Measurement in Routine Health Care Conference.

Simon GE, Revicki DA, Grothaus L, Von Korff M (1998). SF-36 summary scores: are physical and mental health truly distinct? Med Care *36*: 567–72.

Ware J, Kosinski M, Keller SD (1996). A 12-Item Short Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care *34*: 220–33.

Ware JE, Kosinski M, Keller SD (1995). SF-12: how to score the SF-12 Physical and Mental Health Summary Scales. Boston, Mass: The Health Institute, New England Medical Center.

Ware JE, Snow KK, Kosinski MA, Gandek B (1993). SF-36 Health Survey manual and interpretation guide. Boston, Mass: The Health Institute, New England Medical Center.

Wilson D, Parsons J, Tucker G (2000). The SF-36 summary scales: problems and solutions. Soz Praventivmed *45*: 239–46.

Wilson D, Wakefield M, Taylor A (1992). The South Australian Health Omnibus Survey. Health Promot J Aust *2*: 47–9.

Address for correspondence

A/Professor David Wilson
Department of Medicine
University of Adelaide
Queen Elizabeth Hospital
Woodville, South Australia 5011

Tel.: +1 61 8 82226047
Fax: +1 61 8 82226042

e-mail: David.Wilson@dhs.sa.gov.au