Volume 13 • Number 4 • 2010
VALUE IN HEALTH

Dimensionality and Summary Measures of the SF-36 v1.6: Comparison of Scale- and Item-Based Approach Across ECRHS II Adults Population

Mario Grassi, Andrea Nucera on behalf of the European Community Respiratory Health Study Quality of Life Working Group

Dipartimento di Scienze Sanitarie Applicate, Sezione di Statistica Medica e Epidemiologia, Università di Pavia, Pavia, Italy

ABSTRACT

Objectives: The objective of this study was twofold: 1) to confirm the hypothetical eight scales and two-component summaries of the questionnaire Short Form 36 Health Survey (SF-36), and 2) to evaluate the performance of two alternative measures to the original physical component summary (PCS) and mental component summary (MCS).

Methods: We performed principal component analysis (PCA) based on 35 items, after optimal scaling via multiple correspondence analysis (MCA), and subsequently on eight scales, after standard summative scoring. Item-based summary measures were planned. Data from the European Community Respiratory Health Survey II follow-up of 8854 subjects from 25 centers were analyzed to cross-validate the original and the novel PCS and MCS.

Results: Overall, the scale- and item-based comparison indicated that the SF-36 scales and summaries meet the supposed dimensionality. However, vitality, social functioning, and general health items did not fit data optimally. The novel measures, derived a posteriori by unit-rule from an oblique (correlated) MCA/PCA solution, are simple item sums or weighted scale sums where the weights are the raw scale ranges. These item-based scores yielded consistent scale-summary results for outliers profiles, with an expected known-group differences validity.

Conclusions: We were able to confirm the hypothesized dimensionality of eight scales and two summaries of the SF-36. The alternative scoring reaches at least the same required standards of the original scoring. In addition, it can reduce the item-scale inconsistencies without loss of predictive validity.

Keywords: quality of life, SF-36, construct validation, optimal scaling, summary scores.

Introduction

The health-related quality of life (HRQoL) measurement is often used in clinical trials, quality control programs, and health-care system. Several methods to correctly judge HRQoL data are available in clinical practice, and these should be validated and standardized to compare results coming from different studies via a rigorous development and users feedback [1,2]. Among the different questionnaires for assessing HRQoL, the Short Form 36 Health Survey (SF-36), developed by the Medical Outcome Study (MOS), is the most used worldwide [3–5]. For over 15 years, the SF-36 has been proven useful in comparing general and specific populations, in measuring the health deficit and the treatment efficacy, and for screening individual patients. Moreover, it has been found to correlate with the frequency and severity of specific symptoms and disease, as reported in more than 5000 papers on MEDLINE.

The SF-36 questionnaire has been evaluated and proposed in different (SF-36v1, SF-36v2) and shorter (SF-12; SF-8) versions with the aim to better measure HRQoL and to be easily applied in clinical trials. The SF-36 questionnaire describes eight scales with scale score ranges from 0 to 100 (percent of maximum sum score); it covers four physical health perceptions (physical functioning—PF, role limitations because of physical health problems—RP, bodily pain—BP, general health—GH), and four mental health concepts (vitality—VT, social functioning—SF, role limitations because of personal or emotional problems—RE, and mental health perceptions—MH). Successively, two global measures, depending on the height scales, have been derived and referred to as physical component summary (PCS) and as mental component summary (MCS).

The strategy of summary development was set up on the data-driven analysis of the 8×8 Pearson's correlations matrix of the scale scores by means of principal component analysis (PCA). The underlying dimensions have been counted by eigenvalues rules, and both Varimax (orthogonal = uncorrelated components) or Promax (oblique = correlated components) rotations were performed to confirm the hypothesis of the two high order underlying dimensions. Finally, a weighted sum of the eight scales based on the rotated "component score coefficients" have been proposed [6]. The recommended standard MOS system is based on the aim of providing the maximally independent measurement of physical and mental health domains, thus, the scoring method forces the PCS and MCS to be uncorrelated by orthogonal weights (MOS_{UC}). As the physical and mental health are often empirically related, and disease may influence both of them at different extents, an optional MOS scoring method with correlated oblique weights (MOS_C) has been further proposed by the SF36 developers [6], although it is not currently recommended.

A number of studies have confirmed the validity of the dimensional structure SF-36 applying MOS strategy [7–9]. Other studies, via confirmatory factor analysis and structural equation modeling, have introduced additional factors (components), or residual pairwise item correlations, with contrasting results [10–16].

Also, there is an ongoing debate about the summary scoring to be applied. Specifically, the uncorrelated summaries would seem not in agreement with the empirical evidences that mental

Address correspondence to: Mario Grassi, Dipartimento Scienze Sanitarie Applicate, Via Bassi, 21- 27100 PAVIA, Italy. E-mail: mario.grassi@unipv.it

10.1111/j.1524-4733.2009.00684.x

and physical health might strongly interact one to each other [17–18]. Some authors [19,20] highlighted discrepancies between scores on individual scales and components summaries. Taft et al. [21,22] suggested that these discrepancies are attributed to the effects of negatively weighted scales used in the PCS and MCS scoring algorithm. Three "mental" scales are negatively weighted for PCS, while for MCS four "physical" scales are negatively weighted. Thus, the higher the mental health scale scores the lower the PCS, and the higher the physical health scores the lower the MCS (and vice versa). In its extreme, PCS is primarily measuring-impaired mental health, and MCS-impaired physical health! The negative loadings were also assigned by the correlated solution.

With these caveats in mind, by means of the European Community Respiratory Health Survey (ECRHS) data [23,24], the purpose of this study was: 1) to confirm the hypothetical eight scales and two summaries of the SF-36 questionnaire based on a data-driven (exploratory) analysis of the 35 items recoded by "optimal" weights; and 2) to propose two new summary measures to avoid the negative weightings of the MOS (uncorrelated and correlated) component scoring.

Methods

Sample

Data for this report have been taken from ECHRS, an international longitudinal population-based study (25 centers) of more than 10,000 young adults, initially aged between 20 and 44 years in 1991–1993, randomly selected, and followed-up 9 years later using the same standardized protocol, in all the centers. At follow-up, validated Quality of Life Questionnaires, including the SF36, were administered.

Methods of the ECRHS trial are described in detail elsewhere [23,24]. Briefly, subjects were recruited from the ECRHS follow-up (ECRHS II), a longitudinal assessment between 1998 and 2002 of the subjects who participated in the second stage of the ECRHS I. During this stage, two samples have been investigated: a *random sample*, including those subjects who replayed to a regular mail short screening questionnaire who had reported none respiratory symptoms, and a *symptomatic sample*, including the responders to the screening questionnaire who had reported nocturnal shortness of breath or asthma attacks in the last 12 months or asthma treatment.

In all the centers, the SF-36 v.1.6 questionnaire was self-administered after the main ECRHRS clinical interview and before lung function testing. Self-answers to the following long-standing illnesses binary (yes/no) questions were preliminary recorded before administration of the SF-36 questionnaire: "Do you have any long term limiting illness?" and "Do you have any of the following conditions?" using a checklist of eleven chronic illnesses.

Dimensionality Analysis

The SF-36 data was recoded as described in detail in the SF-36 user's manual. Twenty-eight items are in ordinal type following the Likert format (cf. for example PF items: yes, limited a lot; yes, limited at little, no, not limited at all, recoded as 1-2-3), seven items are in binary format (yes–no recoded 1–2), and one item, investigating the health changes over the past year is not used for HRQoL evaluation. Therefore, to investigate the questionnaire dimensionality we rescaled the Likert/binary points of the 35 items of the SF-36 using an "optimal scaling" method draw from multiple correspondence analysis (MCA).

MCA can be introduced in many different ways (see the extensive review in [25,26]), we have considered the Guttman's approach where the first MCA dimension quantifies the rows (subjects) and columns (items) of tabular questionnaire data in such a way that an optimally "internal consistency criterion" is satisfied [27]. This method uses the items as categorical (nominal) variables input, and produces quantifications for each option of the items (called "optimal weights"); and consequently, rescaled item scores (called "optimal quantified variables"), and an "optimal score" computed as sum of the rescaled item scores for each subject can be derived.

By Guttman's optimal scaling approach, the first MCA dimension assigns option quantifications maximizing the total variance of the quantified variables (called Guttman's eta), and the reliability of the scores computed by Cronbach's alpha coefficient. Thus, acceptable internal consistency of the first MCA dimension follows the criteria that has been suggested by Cronbach's alpha >0.70 and >0.90 for groups and for individual comparisons, respectively [28].

Successively, to evaluate the SF-36 questionnaire structure, the PCA and Varimax/Promax rotations were performed considering the 35×35 Pearson's correlation matrix computed on the 35 items after MCA item rescaling. To evaluate the similarity between scale- and item-based approaches, PCA and Varimax/Promax rotations were also performed considering the 8×8 Pearson's correlation matrix computed on the original eight scale scores.

By examining whether the eigenvalues were greater than unity and by looking for sharp breaks in the size of the eigenvalues using a Scree plot, the number of components were counted at the first-order level analysis (i.e., scales); while, by the "Goodness of Fit Index" (GFI), a measure of agreement between the observed and expected, based on PCA, correlation matrix (good if >90% and excellent if >95%) the number of components were counted at the second-order level analysis (i.e., summaries) [28]. After Varimax/Promax rotation, item- and scale-component loadings greater than 0.40 in absolute value were chosen to identify a simple component structure, i.e., component with no overlapping clusters of SF-36 items/scales.

To compare scale versus item values, the item-component loadings and the proportion of item-variance explained by the components were evaluated after averaging within each scale the item-loadings, and the item-variances. Additionally, to get a single number describing the relationship between scale- and item-based loadings, we used the vector correlation coefficient (RV), a generalization of the Pearson's determination coefficient, R^2 [29]. Also, as the determination coefficient, RV is bounded between 0 and 1, and Good, Strong, Excellent agreement between the two matrices has been suggested if RV > 0.50; >0.70, and >0.90, respectively.

Finally, in support of the SF-36 questionnaire structure, we expected comparable second-order structure for scale- and item-based analysis. Specifically, we hypothesized that: 1) the PF items/scales would correlate highest with the physical component, followed by RP and BP items/scales, and all three items/scales would correlate lowest with the mental component; 2) the MH items/scales would correlate highest with the mental component, followed by RE and SF items/scales, and all three items/scales would correlate lowest with the physical component; and 3) the GH and VT items/scales would correlate moderately with both physical and mental components, with GH items/scales correlating higher on the physical component, and the VT items/scales correlating higher on the mental component.

Summary Measures Comparison

We constructed the PCS and the MCS using the general US population means/standard deviations, and the "component

Table 1 US population scale means, standard deviations, and scale weights for the MOS (uncorrelated, and correlated), summary scores, reproduced from Saris-Baglama et al. [**6**: pp. 99, 100, and 111]

	US population		MOS_{UC} (Varimax)		MOS_C (Promax)	
	Mean	SD	PCS	MCS	PCS	MCS
PF	84.52404	22.89490	0.42402	–0.22999	0.34450	–0.10655
RP	81.19907	33.79729	0.35119	–0.12329	0.30379	–0.02356
BP	75.49196	23.55879	0.31754	–0.09731	0.27858	–0.00766
GH	72.21316	20.16964	0.24954	–0.01571	0.23562	0.05247
VT	61.05453	20.86942	0.02877	0.23534	0.09233	0.23434
SF	83.59753	22.37642	–0.00753	0.26876	0.06661	0.25667
RE	81.29467	33.02717	–0.19206	0.43407	–0.06539	0.36583
MH	74.84212	18.01189	–0.22069	0.48581	–0.07870	0.40787

MOS, Medical Outcome Study; UC, uncorrelated; C, correlated; SD, standard deviation; PCS, physical component summary; MCS, mental component summary; PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health.

score coefficients" of PCA on the eight scales derived from the MOS scoring, as illustrated in the SF-36 user's manual (cf. Table 1). Three "mental" scales (SF, RE, and MH) in the PCS, and four "physical" scales (PF, RP, BP, and GH) in the MCS have negative scoring coefficients using the MOS_{UC} method, and two "mental" scales (RE, and MH) in the PCS, and three "physical" scales (PF, RP, and BP) in the MCS have negative scoring coefficients using the MOS_C method.

The MOS scales-based procedure is overall a weighted sum of the eight scales with equal (unit) weights of the items within scales. Thus, our proposal is to assume a scoring system with (–1, 0, 1) weights assigned to each of the 35 items, i.e., an item-based scoring.

Two 0–100 sum scores of the 35 items using (–1, 0, 1) unit-strategy were planned. This strategy uses the "component score coefficients" (i.e., the item weights of the component scores) after Varimax/Promax rotation, in which the salient weights were replaced with 1 or –1 in a manner consistent with their original signs, while no salient weight were replaced with zero. Weights in absolute values that were 1/3 as large or larger than the largest absolute weight per component were considered salient. For theoretical account of unit strategy and comprehensive comparison of 1/3-rule of thumb versus alternative rules using simulated data the reader is referred to [30–32].

Specifically, our PCS and MCS item-based scoring was defined as the simple sums of the 35 Likert/binary items according the unit-strategy for physical and mental components. These two summaries were rescaled 0–100, as the eight SF-36 scales, and were expressed in percent, with 100% indicating the most favorable level of physical/mental health and 0% the most unfavorable.

To assess the similarity between scoring methods (MOS_{UC}, MOS_C, and item-based) the Pearson's correlation coefficients (r) between the summary scores were computed. To evaluate the differences in the scoring methods, the summary scores were compared assessing: 1) the potential influence of outlier profiles by sensitivity analysis; and 2) the clinical (criterion-based) validity by means of know-groups comparison. To carry out the sensitivity analysis, six hypothetical scale profiles (minimum and maximum scale scores; scores of 0 for physical scales, of 100 for mental scales, and vice versa; scores of 1SD for physical scales, of 0.3SD for mental scales, and vice versa) were derived. To carry out validation analysis, subjects were assigned to mutually exclusive groups differing in self-reported asthma-like symptoms, long-term limiting illness, and depression conditions. It was expected that the "physical" components would score worse in the group with long-term limiting illness group, and that the "mental" components would score worse in the depression group; and that both physical and mental components would score worse in the asthma-like symptom group.

Multivariate analysis of variance (MANOVA) models with the PCS and MCS scores as response variables on the explanatory variables defined by the know subject groups, and controlling for ECRHS II centers, were fitted. Each scoring system (MOS_{UC}, MOS_C, and item-based) was evaluated in separate models, and all the scores were compared on the norm-based units linearly transforming the derived summaries in scorings with mean of 50 and 10 as standard deviation, in the general US population. The *P*-values for the parameter estimates (the group mean differences) were evaluated by *F*-test. The significance level was set at $P < 0.05$, two-sided. The "half a standard deviation" rule [33], i.e., a variation of five points for the SF-36 norm-based score, was considered as the minimally clinically important difference for comparing MANOVA parameter estimates.

Descriptive data analyses, MCA, PCA, and MANOVA were performed using SPSS software, version 15.0 (SPSS Inc., Chicago, IL).

Results

Sample

Overall, 29 centers participated in ECRHS II, and 10,933 subjects completed the main questionnaire; 1961 subjects belonging to four centers that did not collect any HRQoL data and 118 subjects who did not answer to any of the SF-36 questions were excluded from the present analyses. Consequently, the SF-36 questionnaire was completed in 8854 subjects from 25 centers; among them, 6611 also completed the questionnaire of long-standing conditions. Twenty-three centers (10 countries) were European and two centers/countries were extra European. Switzerland, Spain, and France covered about half of the included subjects (19.4%, 19%, and 12.1%, respectively). The remaining countries contributed to about 5.5%. The symptomatic sample represented the 16.8% of the total (cf. Table 2).

Dimensionality Analysis

According to Guttman, internal consistency criteria, the recoded options of the 35 items with the optimal weights computed by first MCA dimension, are displayed in Table 3. The homogeneity (Guttman's eta) of the total score and the reliability (Cronbach's alpha) were equal to 0.40 and 0.94, respectively. Across the eight scales, the Guttman's eta of the MCA scaling ranged from 0.56 (GH) to 0.88 (BP), thus the Cronbach's alpha coefficient varied from 0.80 (GH and SF) to 0.92 (PF), indicating an excellent optimal scaling. Generally, the transformation plots displayed that the equal spacing Likert points were not fitted by the optimal weights, indicating a better recoding of MCA (data not shown).

Table 2 Frequencies distributions across countries and random/symptomatic samples of the ECHRS II populations

ECRHS countries	Random		Symptomatic		Total	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Belgium	533	7.2	64	4.3	597	6.7
Spain	1220	16.6	463	31.1	1683	19.0
France	1033	14.0	34	2.3	1067	12.1
Italy	491	6.7	55	3.7	546	6.2
England	530	7.2	129	8.7	659	7.4
Iceland	455	6.2	64	4.3	519	5.9
Norway	588	8.0	0	0.0	588	6.6
Swiss	1346	18.3	369	24.8	1715	19.4
Sweden	368	5.0	79	5.3	447	5.0
USA	194	2.6	35	2.4	229	2.6
Australia	365	5.0	129	8.7	494	5.6
Estonia	243	3.3	67	4.5	310	3.5
Total	7366	100	1488	100	8854	100

ECRHS, European Community Respiratory Health Survey.

The dimensionality indices derived from the Pearson's correlation matrix of item-based strategy (35 items after optimal MCA recoding) were carried out as hypothesized (cf. Table 4). The PCA identified eight dimensions with eigenvalues between 11.7 and close to the unity (0.94), which explained almost the 67% of the observed total variance. The Scree plot, showing only one sharp break indicating eight underling components, also confirmed the SF-36 first-order structure. Only the first two eigenvalues were notably higher than the unity (11.7 and 4.03, respectively). Thus, the GFI signed two dimensions that fitted the 90% of the observed interitem Pearson's correlation matrix on the first component and reached the 90% adding the second component, confirming the two SF-36 underling second-order structure.

Table 3 MCA optimal weights for the item options (from 1 to 6) of the SF-36 questionnaire

Scale	Item	Options					
		1	2	3	4	5	6
PF	a	−1.53	−0.12	0.42			
	b	−2.59	−1.29	0.25			
	c	−2.53	−1.18	0.24			
	d	−2.34	−0.56	0.31			
	e	−2.69	−1.81	0.15			
	f	−2.18	−0.82	0.24			
	g	−2.64	−1.16	0.23			
	h	−2.77	−1.91	0.15			
	i	−2.35	−2.57	0.10			
	j	−2.08	−2.20	0.09			
RP	a	−1.77	0.19				
	b	−1.52	0.28				
	c	−1.67	0.24				
	d	−1.72	0.25				
BP	a	0.45	0.21	−0.15	−0.74	−1.46	−2.33
	b	0.39	−0.21	−1.01	−1.94	−2.53	
GH	x	0.64	0.38	−0.11	−1.44	−3.26	
	a	−1.91	−1.19	−0.59	−0.21	0.31	
	b	0.46	0.05	−0.40	−1.12	−0.91	
	c	−0.89	−0.58	−0.20	0.06	0.37	
	d	0.63	0.19	−0.41	−1.12	−2.03	
VT	a	0.61	0.46	0.13	−0.55	−1.45	−1.96
	e	0.58	0.52	0.23	−0.34	−1.11	−1.80
	g	−2.42	−1.60	−0.97	−0.23	0.31	0.50
	i	−2.03	−1.24	−0.62	0.07	0.48	0.60
SF	a	0.41	−0.44	−1.17	−1.98	−2.58	
	b	−1.37	−1.90	−1.00	−0.18	0.46	
RE	a	−1.49	0.19				
	b	−1.17	0.26				
	c	−1.26	0.22				
MH	b	−1.42	−1.32	−0.77	−0.26	0.12	0.37
	c	−1.64	−2.40	−1.83	−0.93	−0.25	0.32
	d	0.54	0.39	0.08	−0.59	−1.11	−1.04
	f	−1.94	−1.97	−1.50	−0.62	0.11	0.46
	h	0.52	0.38	0.00	−0.57	−1.29	−1.29

MCA, multiple correspondence analysis; SF, social functioning; PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; RE, role emotional; MH, mental health.

As hypothesized, from the Varimax rotation of the axes with eight dimensions being used as cutoff of 0.40 (cf. Table 4), the physical component items (PF, RP, BP, GH) represented four different health concepts. Considering the mental component items, the RE, and MH scales were also loaded on two different dimensions, as expected. Conversely, the SF and VT scales showed an alternative structure. The SFa item weighted on RE dimension, and the SFb item loaded on the MH one; finally, two distinct dimensions were reproduced by the VT's items with the VTg and VTi on one dimension, whereas, VTa and VTe, with the MHh item, on another one. Comparable results were recognized performing the Promax rotation (data not shown).

The scale- and the average (within each scale) item-component loadings of the Varimax/Promax rotation for two dimensions of the scale- and item-based strategies are reported in Table 5. The total variance in the SF-36 scale and the item scores displayed by the two components were 65.3% and 44.8%, respectively. Generally, the inspection of the component loadings, being used as cutoff of 0.40, strongly supported their interpretation as physical and mental health summary measures, in the various analysis conditions.

By means of orthogonal (Varimax) scale-based analysis, the variance explained in each SF-36 scale by the two components ranged from 0.53 (GH) to 0.78 (MH). The PF scale was more associated with the physical and less with the mental dimension (0.80 vs. 0.12). In contrast, the MH scale was linked with the mental, and less with the physical dimension (0.88 vs. 0.13). Both RP and BP scales had higher loadings on the physical dimension (0.75 and 0.76), as well SF and RE scales had stronger adhesions with the mental dimension (0.76 and 0.78). Finally, both the GH and the VT scales had moderate loadings, with the GH scale correlating higher with physical (0.61) than mental (0.40), and the VT scale correlating higher with mental (0.73) than physical (0.37). Similar pattern of scale-loadings was observed from the oblique (Promax) scale-based analysis with the components correlation estimated equal to 0.532.

Comparable pattern of the average (within each scale) item-component loadings were also recognized from the Varimax and Promax item-based analysis, except for the GH scale which loaded on the mental dimension. RV coefficients between the

Table 4 Item-based component loadings after the Varimax rotation of the SF-36 questionnaire

		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
PF	a	0.48							
	b	0.71							
	c	0.67							
	d	0.69							
	e	0.79							
	f	0.64							
	g	0.77							
	h	0.83							
	i	0.81							
	j	0.66							
RP	a		0.81						
	b		0.75						
	c		0.80						
	d		0.78						
BP	a							0.79	
	b							0.72	
GH	x				0.60				
	a				0.68				
	b				0.71				
	c				0.59				
	d				0.70				
VT	a					0.76			
	e					0.78			
	g								0.75
	i								0.74
SF	a			0.52			0.36		
	b			0.37			0.44		
RE	a			0.79					
	b			0.83					
	c			0.75					
MH	b						0.80		
	c						0.69		
	d					0.55	0.53		
	f						0.65		
	h					0.70	0.35		
Eigenvalue*		11.65	4.03	2.01	1.72	1.20	1.06	0.97	0.94
%cum var*		33.28	44.81	50.54	55.46	58.89	61.92	64.68	67.37
%com corr*		80.36	90.00	92.39	94.15	95.00	95.66	96.21	96,69

*Initial eigenvalues, cumulative% of total variance, and cumulative% of total correlation (Goodness of Fit Index) explained by components.
SF, social functioning; PC. principal component; PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; RE, role emotional; MH, mental health.

scale- and item-based results, considering the physical and mental component loading matrix after Varimax and Promax rotations showed an excellent agreement (0.91 and 0.92). The components correlation of Promax item-based analysis was 0.465.

Summary Measures Comparison

Summaries measures using the (–1, 0, 1)-weights of unit rule of item-based strategy were derived from Varimax/Promax rotations of two PCs of the 35 items, after MCA optimal rescaling. In Table 6 only the Promax results, and the linked unit weights are shown, being the positive weights more salient than the negative ones. Thus, the unit strategy becomes a straightforward off/on (0/1) rule, overlapping the controversial negative weightings of the MOS scoring components.

In general, the (0,1)-weights of the component score coefficients were closer to the physical and mental summary hypothesis: the items of the PF, RP, BP scales and the items of the VT, SF, RE, MH scales summed on the physical and mental dimensions,

Table 5 Rotated (orthogonal, and oblique) component loadings of SF-36 scales considering eight scales and 35 items scoring strategies

	Orthogonal (Varimax) rotation						Oblique (Promax) rotation					
	8 scales			35 items			8 scales			35 items		
Scale	PCS	MCS	R^2	PCS	MCS	R^2	PCS	MCS	R^2	PCS	MCS	R^2
PF	0.80	0.12	0.66	0.72	0.10	0.53	0.88	–0.13	0.69	0.76	–0.08	0.53
RP	0.75	0.25	0.63	0.52	0.36	0.4	0.78	0.03	0.62	0.47	0.25	0.39
BP	0.76	0.21	0.62	0.46	0.38	0.35	0.79	–0.01	0.64	0.40	0.29	0.35
GH	0.61	0.4	0.53	0.35	0.4	0.29	0.55	0.26	0.53	0.27	0.34	0.29
VT	0.37	0.73	0.66	0.23	0.64	0.46	0.17	0.71	0.65	0.08	0.64	0.46
SF	0.35	0.76	0.7	0.23	0.70	0.54	0.14	0.75	0.7	0.07	0.70	0.54
RE	0.16	0.78	0.64	0.1	0.65	0.44	–0.09	0.84	0.6	–0.07	0.69	0.44
MH	0.13	0.88	0.78	0.05	0.69	0.48	–0.15	0.96	0.79	–0.13	0.74	0.48
Total	0.308	0.345	0.653	0.219	0.229	0.448	0.434	0.407	0.653	0.279	0.269	0.448
r(PCS;MCS)			0			0			0.532			0.465

Notes: R^2, proportion of variance of each scale explained by the components. Total, proportion of total variance explained by the components; r(PCS,MCS), components correlation.
PCS, physical component summary; MCS, mental component summary; PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health.

Table 6 Component score coefficients standardized by the largest absolute weight per component of Promax (oblique) rotation, and (−1;0,1)-unit weights of the SF-36 questionnaire

Overall Items		Component score coefficients PCS	MCS	(−1;0,1) coding PCS	MCS
PF	a	0.78519	0.06332	1	0
	b	1,00000	−0.11307	1	0
	c	0.92059	−0.07104	1	0
	d	0.87773	−0.03706	1	0
	e	0.95188	−0.18263	1	0
	f	0.87724	−0.11254	1	0
	g	0.99919	−0.12897	1	0
	h	0.99441	−0.21483	1	0
	i	0.94636	−0.26953	1	0
	j	0.81911	−0.22899	1	0
RP	a	0.54103	0.25011	1	0
	b	0.50774	0.32871	1	0
	c	0.62750	0.23963	1	0
	d	0.60724	0.27649	1	0
BP	a	0.52672	0.29340	1	0
	b	0.60606	0.30792	1	0
GH	x	0.47785	0.42491	1	1
	a	0.27637	0.39195	0	1
	b	0.29779	0.38248	0	1
	c	0.17721	0.29867	0	0
	d	0.44189	0.47817	1	1
VT	a	0.10953	0.78772	0	1
	e	0.06319	0.81276	0	1
	g	0.11142	0.71905	0	1
	i	0.07709	0.75344	0	1
SF	a	0.06034	0.87862	0	1
	b	0.06040	0.84370	0	1
RE	a	−0.06406	0.81159	0	1
	b	−0.13734	0.89619	0	1
	c	−0.13291	0.85371	0	1
MH	b	−0.18490	0.78990	0	1
	c	−0.15526	0.93808	0	1
	d	−0.18693	0.92267	0	1
	f	−0.18013	1,00000	0	1
	h	−0.19533	0.92505	0	1

PCS, physical component summary; MCS, mental component summary; PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health.

respectively. Only one PCS' s item (BPb) reached a value closer to the selected cutoff (1/3) for the MCS, yet for conceptual framework was not included in the mental summaries. Vice versa, the items of the GH scale, as previously noted, had a nonzero cross-loadings GHa and GHb summed on the mental summary, GHx and GHd summed on both fields, and finally GHc had a borderline cutoff on the mental summary.

Therefore, via the item-based approach, 18 items are summated in physical health, and 19 in the mental health summary measures, as $41 = 59-18$ and $76 = 95-19$ represent the ranges = maximum–minimum of possible responses of PCS and MCS, respectively. Following the two examples above, the proposed scorings are given:

1. The PCS is composed of 18 items, with several option choices (min 2 and max 5); if the response profile of any subject sums to 56, the PCS = (56–18)/41% = 92.7%.
2. The MCS is formed by 19 items with several option choices (min 2 and max 6); if the response profile of any subject sums to 81, the MCS = (81–19)/76% = 81.6%.

Thus, the global scores of the subject are 92.7% and 81.6% of the most favorable level of physical and mental health, respectively.

Alternatively, after simple algebra, PCS and MCS can be also computed using the scale scores as:

$$PCS = \frac{20}{41}PF + \frac{4}{41}RP + \frac{9}{41}BP + \frac{8}{41}(GHx + GHd)$$

$$MCS = \frac{20}{76}VT + \frac{8}{76}SF + \frac{3}{76}RE + \frac{25}{76}MH + \frac{20}{76}GH$$

i.e., the item-derived summary measures are scale weighted sums, where the weights are the raw scale ranges, except for GH items on the physical health measure.

Using the US population means and the standard deviations of Table 1, the norm-based scores derived from item-based scoring were given by the *T*-score linear transformation:

$$PCS^* = \frac{PCS - 79.72451}{16.32461} \times 10 + 50$$

$$MCS^* = \frac{PCS - 71.54387}{15.80036} \times 10 + 50$$

where 79.72451 (16.32461) and 71.54387 (15.80036) were the US population means (standard deviation) resulting from the raw scale range weighting.

Pearson's correlations of the MOS_{UC}, MOS_C, and item-based alternative summaries are shown in Table 7. As expected, the correlation of Varimax (MOS_{UC}) mental and physical component scores was closer to zero; whereas, there was a moderate correlation (>0.50) of the two health components for the Promax (MOS_C and item-based) scorings. Considering the between scoring system correlations, the MOS_C and item-based physical (0.97) and mental (0.96) scores were collinear; while the correlations of MOS_{UC} and item-based physical (0.92) and mental (0.89) scores were slightly lower.

Table 8 summarizes the scoring method comparison between hypothetical scale profiles, and MANOVA mean estimates of know-groups with self-reported asthma-like symptoms, long term limiting illness, and depression conditions, adjusting for ECRHS centers.

Comparing hypothetical profiles the proposed item-based scoring is in line with the expected: when the scales has minimum (maximum) scores, the PCS and MCS are 0 (100). When the physical health scales are at 100, and the mental scale are at 0

Table 7 Means, SD, and Pearson's (r) summary measure correlations for the MOS (uncorrelated, and correlated), and item-based scoring methods

	PCS (MOS_{UC})	MCS (MOS_{UC})	PCS (MOS_C)	MCS (MOS_C)	PCS (item)	MCS (item)
PCS (MOS_{UC})	1.00					
MCS (MOS_{UC})	−0.05	1.00				
PCS (MOS_C)	0.94	0.29	1.00			
MCS (MOS_{cC})	0.18	0.97	0.50	1.00		
PCS (item)	0.92	0.28	0.97	0.48	1.00	
MCS (item)	0.33	0.89	0.62	0.96	0.61	1.00
Mean	51.95	50.26	51.95	50.78	52.27	50.86
SD	8.05	9.84	8.07	9.61	9.74	9.96

SD, standard deviation; PCS, physical component summary; MCS, mental component summary; MOS, Medical Outcome Study; UC, uncorrelated; C, correlated.

Table 8 Hypothetical scale profiles and MANOVA estimate mean scale profiles for the symptomatic, log-term illness, and depression groups of ECHRS II populations* comparing MOS with item-based scoring methods

	Hypothetical profiles						Symptomatic (n = 935)	Long term illness (n = 1087)	Depression (n = 763)
	min	max	100;0	0;100	1SD; 0.3SD	0.3SD; 1SD			
PF	0	100	100	0	100	91	79.3	75.6	78.9
RP	0	100	100	0	100	91	69.6	64.3	66.2
BP	0	100	100	0	99	83	63.4	57.5	61.1
GH	0	100	100	0	92	78	56.8	52.6	54.1
VT	0	100	0	100	67	82	50.3	49.0	45.3
SF	0	100	0	100	90	100	69.1	66.8	60.9
RE	0	100	0	100	91	100	66.2	66.6	54.6
MH	0	100	0	100	80	93	61.1	61.5	53.5
PCS (MOS_{UC})	20.2	57.9	75.0	3.2	59.3	51.0	46.8	44.1	47.2
MCS (MOS_{UC})	17.3	62.1	−1.3	80.7	50.9	60.2	42.9	43.5	38.1
PCS (MOS_C)	12.4	61.0	59.9	13.4	59.2	53.7	44.9	42.6	44.1
MCS (MOS_C)	10.5	63.8	7.4	66.9	53.4	60.1	42.3	42.2	37.8
PCS (item)	1.2	62.4	62.4	1.2	61.3	54.3	43.7	40.8	42.6
MCS (item)	4.7	68.0	12.6	60.1	55.0	60.9	41.3	40.5	37.2
PCS 0–100	0	100	100	0	98	87	69.4	64.6	67.7
MCS 0–100	0	100	13	88	79	89	57.8	56.5	51.3

*n = 6359 subjects and c = 20 centers, excluding the missing values from N = 6611 subjects who completed the questionnaire of long-standing illnesses conditions. All know-group (yes–no) mean difference estimates were statistically significant ($P < 0.05$) by MANOVA *F*-test.
MANOVA, multiple analysis of variance; ECHRS, European Community Respiratory Health Survey; MOS, Medical Outcome Study; SD, standard deviation; PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health; PCS, physical component summary; UC, uncorrelated; MCS, mental component summary; C, correlated.

scores, the PCS, and MCS are 100 and 13, respectively, and vice versa are 0 and 83 for the 0–100 profile. Examination of the 1SD–0.3SD profile, i.e., scores of 1SD above the mean for physical health scales, and 0.3SD above the mean for mental health scales, the PCS norm-based score is 61.3 (1.1 SD above the mean), and the MCS is 54.3 (0.4 SD above the mean), and vice versa for the 0.3SD–1SD profile are 54.3 and 60.9, respectively.

By contrast, the MOS_{UC} scores have shown inconsistent results; for example, considering the 1SD–0.3SD profile the PCS score is 59.3 (about 1SD above the mean), but MCS score is 50.9 (about equal to the mean); maximum/minimum scores are produced by 100–0 and 0–100 bipolar profiles, i.e., excellent physical health combined with poor mental health; vice versa. The MOS_C scores present the best results on the 1SD–0.3SD, and 0.3SD–1SD profiles, yet extreme scores for 100–0 and 0–100 profiles.

In general, there was a clear difference between know-groups in mean differences of MOS_{UC}, MOS_C, and proposed item-based scorings. As expected, symptomatic sample had lower average scores on both the physical and mental scores; subjects with long-term limiting illness had the lowest average profile on the physical summaries, while those in the depressive group reported poorer average heath status on mental ones. Comparing the three scoring systems, identical rankings were observed with mean scores that differed from the US norm means (50 for norm-based summaries, or 79.7 and 71.5 for 0–100-based PCS and MCS, respectively) by five points of the half a standard deviation rule or more, with decreasing scoring system order ($MOS_{UC} < MOS_C <$ item-based).

Discussion

The SF-36 is one of the widely used HRQoL measures, and the ECRHS II dataset, composed by 8854 valid questionnaires coming from 25 international centers is one of the most widespread dataset including SF-36 administration. The huge number of data gave us the opportunity to produce reliable results and to evaluate the measurement features of SF-36 questionnaire.

The first question we aimed to answer at was to confirm the eight first-order dimensions and the two second-order dimensions of the SF-36. To do this, we performed a PCA based on the 35 items, after MCA optimal quantifications. Scale-based analysis was used in several studies on SF-36, and item-based analysis is performed in confirmatory analysis by using structural equations modeling [12–16], only in small number of exploratory analysis studies an item-based level has been considered [34–37], but both exploratory or confirmatory studies have used the Likert/binary item response as continuous variables.

As the SF-36 is widely used in clinical practice, it was mandatory to investigate the dimensionality by different approaches. There is no reason to assume that the answers to the questionnaire queries such as: 1 = "definitely true," 2 = "mostly true," 3 = "don't know," 4 = "mostly false," and 5 = "definitely false" of items a–d of GH scale should have equal intervals as supposed by Likert recoding. As highlighted by our study on Likert/binary formats [38], linearity assumption among ordinal response points is often not respected in SF-36 items, and it is necessary to calibrate the Likert recoding. Thus, we use the methodology of MCA "optimal scaling" recoding before dimensionally testing via PCA.

Optimal scaling comes from psychometrics that assigns numeric values to categorical variables in an optimal way, and then the item responses are judged as continuous. As well detailed by De Leeuw [39], the single item quantifications derived by MCA linearize all the bivariate regressions in the Pearson's correlation matrix. In this way, MCA allows the management of the nonlinear information contained in the original data, and the performance of a suitable linear PCA on the 35 items of the SF-36 questionnaire, i.e., MCA/PCA produces a non linear multivariate analysis (see e.g., Gifi [25]).

After optimal item quantifications via MCA, the conventional PCA output presented here showed a positive response of the supposed dimensionality of the eight scales and the two summaries, and generally, support that the items of a scale, and the scales of a summary, loaded with high component loadings (>0.40) on the supposed underlining constructs. However, some discrepancies were noted.

Considering the eight dimensions (scales), items of the VT scale split up in items measuring positive (VTa and VTe, with MHh item) and negative (VTg and VTi) mental health status, and

indicated that the VT scale does not measure one single underlying construct in the ECRHS subjects. Also, the two items of the SF scale split up, but on other supposed constructs: SFa on the RE scale, and SFb on the MH scale. It is of note that positively and negatively worded items loaded on two components, suggesting that the subjects had difficulties or misleading in changing between these reversed answering formats.

The summary components defined by scale- and item-based analysis with orthogonal (uncorrelated) and oblique (correlated) rotations confirmed the underlying two-component structure of the SF-36 questionnaire. Nevertheless, item-based analysis suggested that the GH scale correlated (in average) with the mental rather than the physical component of health. Specifically, GHx and GHd items load on both the physical and mental components, the GHa, and GHb items on the mental components, while GHc on anyone.

These findings, also reported in other item-based studies [35–37], highlight the need to consider the VT, SF, and GH items more closely and possibly to modify the conceptual framework to improve the underlying dimensionality of the questionnaire.

The second aim of the present study was to compare the two global summary components, based on the eight scales (MOS approach), with an alternative one based on the 35 items (item-based approach). The two employed approaches have similar course about how the summaries should be handled. The MOS approach studies the scale dimensionality and structure via PCA using the Pearson's correlation matrix of eight scales. These scales were computed after Likert coding (from 1 to 6 points as maximum), recalibration, and sum of the item responses. Successively, the PCS and MCS summaries were obtained as a weighted sum of the eight scales. The item-based alternative approach starts considering the items as categorical (nominal) variables, and evaluates the item dimensionality and structure via PCA using Pearson's correlation matrix of 35 items after MCA optimal data coding of the item responses. Successively, the PCS and MCS summaries were obtained as on/off (0/1) sum of the 35 items. Thus, two steps should be processed to calculate the summaries for both approaches, but these steps are quite different in conceptual framework and operational procedure of scoring development.

According to the Likert model of the MOS first step, a construct is regarded as being latent continuous, and is operationalized to be measured by highly correlated and equally important items in order to increase the reliability and improve precision. It uses arbitrary numbers, which indicate the ordered structure of the alternative responses, and also assumes that the precision increases with the number of digits in the scale [28]. In contrast, according to the MCA model of the item-based first step, the ordinal/binary data of the SF-36 items were processed as nominal ones, and were transformed in continuous form by optimal quantifications. The MCA solution allows to define the optimal weights for the item options and their ranking, independently by an a priori recoding, enabling an optimal grading for each category response of the questionnaire.

Scores for the two summary measures in the MOS second step are generated in three stages. First, the 0–100 scale scores are standardized (*z* score transformation) by subtracting the US population mean for that scale and dividing the difference by the US population standard deviation for the scale. Next, *z* scores are multiplied by the respective principal component coefficients, derived from US population data, and summed. These are weighed scales sums of an equal weighting of items within each scale. Finally, these summary scores are linearly reexpressed to have a mean of 50 and a standard deviation of 10 (*T*-score transformation), in the general US population.

The item-based alternative scoring in the second step is very basic, just a simple sum (without weighting) of the Likert/binary responses of the items loading in the physical or mental components. To facilitate comparisons across scales and summaries, summaries are reexpressed in the 0–100 scale scores range. Otherwise, a weighted scales sums, where the weights are the raw ranges of the scales, except for GH scale, can be computed; thus, the summaries can be reexpressed in standard deviation units as *T*-scores, using US or other population norms.

Ware et al. [40,41] have provided extensive justification for two uncorrelated PCS and MCS solutions. The advantages include: easier modeling with respect to the additional factors (components), or residual pairwise item correlations. Independent components are more responsive to the distinction between psychical and mental health outcomes. A direct relationship between component loading and explained variance gives an easier interpretation; also, oblique components require negative scoring weights. Nevertheless, recent comparison studies of Farivar et al. [42], Hann and Reeves [43], and Anognostopoulos et al. [44] have recommended that users of the SF-36 adopt the oblique solution for calculating PCS and MCS, but the proposal SF-36 summary scores was similar to MOS_C or was structural equations-based.

Our item-based scoring is derived from a data-driven (exploratory) PCA correlated solution and use the 1/3-unit rule; thus, it fits several PCA advantages of the Ware's scoring system, and objectively overlaps the possible negative weightings of oblique solution with a posteriori rule. Consequently, our alternative scoring is equivalent to the RAND-36 method [45] based on the item response theory for item scoring, and on a correlated confirmatory factor analysis solution that conceptually force a priori the weighting of the four scales of mental health to zero in PCS, and vice versa the four scales of physical health to zero in MCS.

Similarly to the MOS_C strategy, the new scoring allows the physical and mental health summary scores to be somewhat correlated, and assess the extent of this correlation in each study population. As hypothesized in the ECHRS populations, the correlation between PCS and MCS, using scale- and item-based systems, get to 0.53 and 0.47, respectively, showing a moderate value in line with the previous studies using MOS_C. Moreover, the MOS_C and item-based physical and mental scores correlations were both equal to 0.97, suggesting that the scores were empirically the same. By contrast, the MOS_{UC} and item-based physical and mental scores correlations were less noteworthy (PCS: 0.92 and MCS: 0.89).

Our item-based scores can be expressed as 0–100 scores, matching to the original scale scores, or as norm-based scores, matching to the original summary scores. Comparing with the MOS_{UC} summaries, the item-based alternative ones are in line with the expected scores derived on hypothetical outlier scale profiles. Thus, the SF-36 scales scores and physical and mental health summary scores are in agreement, reducing the inconsistent results reported in some SF-36 studies. Additionally, clinical (criterion-based) validity of the proposed scores by means of know-groups comparison produces results supporting the hypotheses suggested, and are compatible with those of the MOS scores.

We would encourage other authors to investigate this alternative item-based scoring in addition to MOS uncorrelated and correlated scorings, to determine if our findings can be replicated using other populations and condition-specific samples. MCA, as PCA, is a procedure presents in all the general statistical packages (SPSS, SAS, Stata, R); in SPSS and R, the data matrix with optimal quantifications of the first MCA dimension is automati-

cally saved in a file as option. Lastly, future research should be dedicated to deriving the norms for various conditions, ages, genders, countries of the new scoring, but the norm-based rescaling in the US population means and standard deviations, or other specific populations can be employed until that new norms are available.

Conclusions

We suggest that the approach we developed may contribute to the improvement of the SF-36 scales and summary measures in HRQoL research. Our results supported the hypothesized dimensionality of the eight scales and the two component summaries of the SF-36 by using item-level data-driven (exploratory) analysis. The new physical and mental summary measures are simple item sums or weighted scale sums, where the weights are the raw scale ranges. The alternative scoring reach at least the same required standards of the original scoring avoiding the negative coefficients weighting produced by the MOS orthogonal (uncorrelated) and oblique (correlated) PCA solution. This can reduce inconsistent results between the SF-36 scale scores and summary scores reported in previous studies.

This study was partially supported by a grant from the MIUR-Italy. The authors wish to thank B. Leynaert at INSERM (Paris, France), and J.M. Anto at IMIM-IMAS (Barcelona, Spain), coresponsibles of Quality of Life Working Group of ECRHS II, and E. Omenaas at Armauer Hansens House (Bergen, Norway) for providing suggestions after draft reviewing. Also, the authors gratefully acknowledge the members of the ECRHS II groups who made the data of this research available.

Source of financial support: None.

References

1 Wilson IB, Cleary PD. Linking clinical variables with Health-Related Quality of Life: a conceptual model of patient outcomes. JAMA 1995;273:59–65.
2 Aaronson NK, Acquadro C, Alonso J, et al. International quality of life assessment (IQOLA) project. Qual Life Res 1992;1:349–51.
3 Ware JE, Sherbourne CD. The MOS 36-Item Short Form Health Survey (SF-36). I Conceptual framework and item selection. Med Care 1992;30:473–83.
4 McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247–63.
5 McHorney CA, Ware JE, Racchel L, Sherbourne CD. The MOS 36-Item Short Form Health Survey (SF-36). III. Tests of data quality, sckaling assumptions, and reliability cross diverse patient groups. Med Care 1994;32:40–66.
6 Saris-Baglama RN, Dewey CJ, Chisholm GB, et al. SF Health outcomes scoring software user's guide. Lincoln, RI: QualityMetric Inc. 2004.
7 Jenkinson C, Layte R, Lawrence K. Development and testing of the Medical Outcomes Study 36-item short form health survey summary scale scores in the United Kingdom: results from a large-scale survey and a clinical trial. Med Care 1997;35:410–16.
8 Ware J, 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;51:1159–65.
9 Gandek B, Sinclair S, Kosinski M, Ware J. Psychometric evaluation of the SF-36 Health Survey in medicare managed care. Health Care Financ Rev 2004;25:5–24.
10 Wolinsky FD, Stump TE. A measurement model of the Medical Outcomes Study 36-item Short-Form Health Survey in a clinical sample of disadvantaged, older, black, and white men and women. Med Care 1996;34:537–48.
11 Lewin-Epstein N, Sagiv-Schifter T, Shabtai EL, Shmueli A. Validation of the 36-item Short-Form Health Survey (Hebrew version) in the adult population of Israel. Med Care 1998;36:1361–70.
12 Reed PJ. Medical outcomes study short form 36: Testing and cross-validating a second-order factorial structure for health system employees. Health Serv Res 1998;33:1361–80.
13 Keller S, Ware J, Bentler P, et al. Use of structural equation modelling to test the construct validity of the SF-36 health survey in ten countries: Results from the IQOLA project. J Clin Epidemiol 1998;51:1179–88.
14 Anagnostopoulos F, Niakas D, Pappa E. Construct validation of the Greek SF-36 health survey. Qual Life Res 2005;14:1959–65.
15 Beals J, Welty TK, Mitchell CM, et al. Different factor loading for SF-36: The strong heart study and the national survey of functional health status. J Clin Epidemiol 2006;59:208–15.
16 Guthlin C, Walach H. MOS SF-36: Structural equation modelling to test the construct validity of the second-order factor structure. Eur J Psychol Assess 2007;23:15–23.
17 Dersh J, Polatin P, Gatchel R. Chronic pain and psychopathology: research findings and theoretical considerations. Psychosom Med 2002;64:773–86.
18 Schattner A. The emotional dimension and the biological paradigm of illness: time for a change. Q J Med 2003;96:617–21.
19 Simon G, Revicki D, Grothaus L, Vonkorff M. SF-36 Summary scores—Are physical and mental health truly distinct? Med Care 1998;36:567–72.
20 Nortvedt M, Riise T, Myhr K-M, Nyland H. Performance of the SF-36, SF-12, and RAND-36 summary scales in a multiple sclerosis population. Med Care 2000;38:1022–8.
21 Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? Qual Life Res 2001;10:395–404.
22 Taft C, Karlsson J, Sullivan M. Interpreting SF-36 summary health measures: A response – Reply. Qual Life Res 2001;10:415–20.
23 Burney PGJ, Luczynska C, Chinn S, et al. The European Community Respiratory Health Survey. Eur Respir J 1994;7:954–60.
24 European Community Respiratory Health Survey II Steering Committee. The European Community Respiratory Health Survey II. Eur Respir J 2002;20:1071–9.
25 Gifi A. Nonlinear multivariate analysis. Chichester: Wiley, 1990.
26 Greenacre MJ. Theory and applications of correspondence analysis. London: Academic Press; 1984.
27 Guttman L. The quantification of a class of attributes: a theory and method of scale construction. In: Horst P, ed., The Prediction of Personal Adjustment. New York: Social Science Research Council; 1941.
28 Mcdonald RP. Test theory. A unified treatment. London: Lawerence Erlbaum Associates; 1999.
29 Robert P, Escoufier Y. A unifying tool for linear-multivariate statistical methods: The RV-coefficient. Appl Stat 1976;25:257–65.
30 Ten Berge JMF, Knol DL. Scale construction on the basis of component analysis: a comparison of three strategies. Multivariate Behav Res 1985;20:45–55.
31 Grice JW, Harris RJ. A comparison of regression and loading weights for the conputation of factor scores. Multivariate Behav Res 1998;33:221–47.
32 Grice JW. A comparison of factor scores under conditions of factor obliquity. Psychol Methods 2001;6:67–83.
33 Samsa G, Edelman D, Rothman M, et al. Determining clinically important differences in health status measures: A general approach with illustration to the health utilities index mark II. Pharmacoeconomics 1999;15:141–55.
34 de Vet HCW, Adèr HJ, Terwee CB, Pouwer F. Are factor analytical techniques used appropriately in the validation of health status questionnaire? A systematic review on the quality of factor analysis of the SF-36. Qual Life Res 2005;14:1203–18.

35 Failde I, Ramos I. Validity and reliability of the SF-36 health Survey Questionnaire in the patients with coronary artery disease. J Clin Epidemiol 2000;553:359–65.
36 Thumboo J, Fong KY, 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. Qual Life Res 2001;10: 175–88.
37 Dallmeijer AJ, Dekker J, Knol DL, et al. Dimensional structure of the SF-36 in neurological patients. J Clin Epidemiol 2006;53: 541–3.
38 Grassi M, Nucera A, Zanolin E, et al. Performance comparison of Likert and binary forms of SF-36 v.1.6 across ECRHS II adults populations. Value Health 2007;10:478–88.
39 de Leeuw J. Multivariate analysis with linearizable regressions. Psychometrika 1988;53:437–54.
40 Ware J, Kosinski M. Interpreting SF-36 summary health measures: A response. Qual Life Res 2001;10:405–13.
41 Ware J, Kosinski M. Interpreting SF-36 summary health measures: A response -Supplemental documentation. Retrieved from SF-36. Available from: http://www.sf-36.org/news/responsetotaft.pdf [Accessed December 14, 2009].
42 Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 health Survey V.1. Health and Qual Life Res 2007;5:54 [PMCID: PMC2065865].
43 Hann M, Reeves D. The SF-36 scales are not accurately summarized by independent physical and mental component scores. Qual Life Res 2008;17:413–23.
44 Anagnostopoulos F, Niakas D, Tountas Y. Comparison between exploratory factor-analytic and SEM-based approaches to constructing SF-36 summary scores. Qual Life Res 2009;18:53–63.
45 Hays RD, Sherbourne CD, Mazel H. The RAND-36 item health survey 1.0. Health Econ 1993;2:217–27.