

A factor analytic study of the Italian National Institute of Health Quality of Life – Core Evaluation Form (ISSQoL-CEF)

M Lauriola¹
R Murri³
M Massella⁴
M Mirra⁴
S Donnini⁴
V Fragola⁴
J Ivanovic⁵
M Pavoni⁶
G Mancini²
R Bucciardini⁴

¹Department of Social and Developmental Psychology,

²Department of Infectious and Tropical Diseases, University of Rome “La Sapienza”, Rome, Italy; ³Catholic University of “Sacro Cuore”, Rome, Italy; ⁴Istituto Superiore di Sanità, Rome, Italy; ⁵National Institute for Infectious Diseases Lazzaro Spallanzani, Rome, Italy; ⁶Ospedale Civile Santa Maria delle Croci, Ravenna, Italy

Objectives: The Italian National Institute of Health Quality of Life – Core Evaluation Form (ISSQoL-CEF) is a specific questionnaire measuring health-related quality of life for human immunodeficiency virus-infected people in the era of highly active antiretroviral therapy. The main goal of this study was to examine the construct validity of this questionnaire by confirmation of its hypothesized dimensional structure.

Methods: Baseline quality of life data from four clinical studies were collected and a confirmatory factor analysis of the ISSQoL-CEF items was carried out. Both first-order and second-order factor models were tested: Model 1 with nine correlated first-order factors; Model 2 with three correlated second-order factors (Physical, Mental, and Social Health); Model 3 with two correlated second-order factors (Physical and Mental/Social Health); Model 4 with only one second-order factor (General Health).

Results: A total of 261 patients were surveyed. Model 1 had a good fit to the data. Model 2 had an acceptable fit to the data and it was the best of all hierarchical models. However, Model 2 fitted the data worse than Model 1.

Conclusions: The findings of in this study, consistent with the results of previous study, pointed out the construct validity of the ISSQoL-CEF.

Keywords: confirmatory factor analysis, HRQoL, patient-reported outcomes

Introduction

The Italian National Institute of Health Quality of Life (ISSQoL) instrument is a recently developed health-related quality of life (HRQoL) instrument specifically designed to meet with the needs of human immunodeficiency (HIV) patients, clinicians, and health care providers in the age of highly active antiretroviral treatment (HAART).¹ Since the advent of HAART, the survival rate for HIV patients has increased and the assessment of HRQoL has become a priority for patients, physicians and researchers.² The ISSQoL is comprised of two sections, which may be jointly or separately used: the Core Evaluation Form and the Additional Important Areas Form. The additional form was not intended as a measure of HRQoL, rather it was devised as an optional part of the ISS-QoL aimed at investigating health related aspects which are important for managing the patient care but are not considered as core aspects of HRQoL³ (eg, the quality of interactions involving HIV people with their own social support network, such as partner, family, and friends, or the evaluation of the relationship with their own medical caring staff, or the willingness to plan a future parenthood or motherhood). Differently, the Core Evaluation Form, hereafter referred to as ISSQoL-CEF (or just CEF), was intended as a psychometric tool measuring HRQoL characteristics

Correspondence: Raffaella Bucciardini
Therapeutic Research and Medicines
Evaluation, Istituto Superiore di Sanità,
Viale Regina Elena, 299, 00161 Rome, Italy
Email r.bucciardini@iss.it

emanating from the multidimensional definition of health provided by the World Health Organization (WHO).³ The ISSQoL-CEF included 37 items and it has been developed to measure patient HRQoL along nine domains: physical well-being (PW), role well-being (RW), energy and vitality (EV), social functioning (SF), sexual life (SL), satisfaction with quality of life (SQL), depression and anxiety (DA), health distress (HD), and cognitive functioning (CF).

The CEF was created to measure HRQoL in HIV-infected persons by taking into account their new needs and changes related to the introduction of HAART. For example, increasing attention is now paid to patients' sexual life and sexual dysfunctions. More often than in the pre-HAART era, HIV-infected patients can now study or work or have a relatively normal daily life. Social life is now more intense than in the past.⁴ Like all psychometric assessments, good reliability and validity make a HRQoL questionnaire valuable. However, whereas reliability (ie, the extent to which repeated administrations of a measurement device produced equivalent results under controlled circumstances) can easily be assessed by choosing an appropriate reliability coefficient, the appraisal of validity is a complex and often long-lasting process.^{5,6} In particular, since different types of validity exist (eg, content, face, and construct validity), each needing specific methodological options, no single study can address the issue whether a given HRQoL measure is valid. Rather, the validity of a HRQoL measure like the ISSQoL-CEF is often attained by composing information from different studies. As it regards content validity^{5,7,8} (ie, a type of validity which examines the extent to which a specific field of interest has been comprehensively sampled by the items, or questions included in the instrument), the ISSQoL-CEF items and domains can be reasonably considered as representative of the HRQoL HIV-infected people, since they were generated based on a comprehensive literature review of HRQoL facets as well as on a content analysis of existing specific questionnaires for HIV-infected people.¹ As it concerns face validity^{5,7,8} (ie, a type of validity which considers how a questionnaire measure appears to its end-users and respondents), we have refined the item wording and improved the questionnaire's perceived utility by discussing preliminary CEF versions with HIV-infected people in repeated focus groups.¹

Along with the appraisal of content and face validity, the psychometric theory^{3,5,7,8} calls, however, for a more compelling assessment of construct validity as a mandatory step prior to licensing a questionnaire as a "ready for use" instrument in research and clinical settings. Briefly, a construct is a theoretically derived notion of the concept (or concepts) that an

assessment instrument is intended to measure. In our specific case, the HRQoL construct was defined as the patients' perception of their physical, social and mental health status, through evaluation of nine specific facets. Because the assessment of construct validity seeks for the agreement between a such theoretically derived notion and the specific measuring device under investigation, a comprehensive test of the ISSQoL-CEF unavoidably passes through the examination of how well the 37 items converged on each of the nine specific domains. So far, the ISSQoL-CEF construct validity was examined based on the analysis of multi-trait/multi-item matrices,^{5,6} that is a descriptive analysis of convergent and divergent item-domain correlations. Albeit this analysis¹ demonstrated that all the CEF items were more correlated with the specific domain they conceptually belong to (convergent correlations) than with other specific domains measured by the instrument itself (divergent correlations), these results have still to be considered as preliminary evidence for construct validity. In fact, not only the analysis of multi-trait/multi-item matrices was merely descriptive but it also did not distinguished properly between the constructs level and the level of their indicators, as both the items and the total domain scores may be regarded as measured variables, rather than as "latent" and "observed" variables, respectively. Furthermore, the analysis of multi-trait/multi-item matrices did not formally test any hypothesis as to the latent structure of the ISSQoL-CEF nor it examined whether items belonging to conceptually similar specific domains converged on a higher level general domain construct, such as for instance the case of depression/anxiety, health distress and cognitive functioning items which are all expected to converge on the superordinate dimension of mental-health. The present study is aimed at expanding on existing psychometric work¹ by testing the ISSQoL-CEF construct validity on an independent sample of respondents and by an hypothesis testing approach, such as the confirmatory factor analysis (CFA). Beyond advantages over earlier descriptive analysis,¹ the CFA approach has also the following advantages over standard exploratory factor analysis (EFA).⁹ First, CFA is usually performed on covariance matrices rather than on correlation matrices, which facilitates the comparison of model parameters across samples (eg, testing for invariance of the measurement model across groups). Second, factor rotation issues, including the problem of establishing the "right" number of factors (that is the real "Achilles' heel" of EFA), are irrelevant in CFA, since this latter method needs prior knowledge of the factor model to be tested. Third, in CFA each item regress only on a single "latent" variable representing the construct that it is hypothesized to measure. As a result, the variance of each

item can be strictly broken down into independent “true” and “error” variance components. This partition closely resembles both classic and modern psychometric theories assumptions, and it allows for establishing the reliability coefficient for each of the hypothesized latent variables. Last but not least, CFA provides a formal goodness-of-fit tests of the ability of a given factor model to fit data from one or different samples as well as of different factor models to fit the data from a single sample. In summary, all these features makes the CFA method a comprehensive approach for examining the ISSQoL-CEF construct validity.

Methods

Patients

Baseline ISSQoL-CEF quality of life data from four clinical studies for HIV+ infected patients were used to perform the CFA. Baseline demographic and clinical data were also collected.

Clinical study 1

Baseline quality of life data were collected from 134 patients participating in a randomized 24 weeks, controlled, open-label study of immediate versus delayed treatment with lipofilling surgery in HIV-positive people with severe facial lipoatrophy. Enrolled HIV-infected patients meet the following requirements: aged 18 years or older; CD4 count $>100/\text{mm}^3$; HIV-RNA below 1000 copies/mL; HAART therapy permanent for at least six months were included in this analysis.

Clinical study 2

Data from an observational, not randomized, controlled, longitudinal study to assess the impact of enfuvirtide on HRQoL in HIV-positive patients were used in this study. Antiretroviral-experienced and fusion inhibitor-naïve patients and aged 18 years or older were eligible for this study. Baseline quality of life data of 25 patients were included in this analysis.

Clinical study 3

Data from a cohort observational study aimed to measure the HRQoL of life of HIV-positive people eligible to CD4-guided interruptions regimen was included in this analysis. To be enrolled, patients will had to be eligible at the treatment interruptions, according to the following criteria: CD4 >500 ; HIV-RNA <50 ; CD4 nadir >200 , aged over 18 years, with capacity to fill the questionnaire. For this analysis quality of life baseline data were available for 24 patients.

Clinical study 4

Baseline quality of life data of 78 patients enrolled in an observational study aimed to verify the responsiveness validity of ISSQoL questionnaire were considered in this analysis. HIV-positive infected patients aged 18 years or older taking antiretroviral treatments were eligible for this study.

Statistical analyses

Parameterization of factor models

In keeping with earlier research,¹ we first tested a model which assumed nine correlated factors, each representing one of the ISSQoL-CEF-specific domains (Model 1). The correlations among factors were unconstrained in this model (Figure 1, Panel a). However, since a comprehensive test of construct validity also required the assessment of how well different specific domains regressed on more general domains, it was worthwhile to constrain the factor correlations by setting appropriate second-order factors in the model. In so doing, the correlations among first-order factors were no longer parameterized, as they were fully explained by structural paths linking second-order factors to first-order factors. Among all possible hierarchical arrangements of factors, we tested a model with three correlated second-order factors (see Figure 1, Panel b) which resembled the WHO definition of health³ (Model 2). As to the operative definition of which first-order factor was expected to regress on which second-order factor, we have made the following choices. First, we have hypothesized that physical well-being, role well-being, and energy/vitality regressed on the general Physical Health factor. Second, we have hypothesized that depression/anxiety, health/distress, and cognitive functioning regressed on the general Mental Health factor. Finally, we have assumed that both social functioning and sexual life regressed on the Social Health factor. While no explanation is needed for why the social functioning was linked to the general Social Health factor, we have considered the sexual life as a one of the facets of the general Social Health based on the fact that sexual relations of HIV infected people in the HAART age have several social and relational implications. Though one can regress a first-order factor (eg, the Sexual Life) on more than one second-order factor (eg, the Physical and the Social Health), this practice is strongly discouraged in second-order factor models¹⁰⁻¹³ as it may lead to empirical identification problems of the upper part of the model, especially if second-order factors were correlated. Similarly, though the satisfaction with overall quality of life could have been related to all three second-order factors we have decided to regress it on the Social Health factor, since this second-order factor

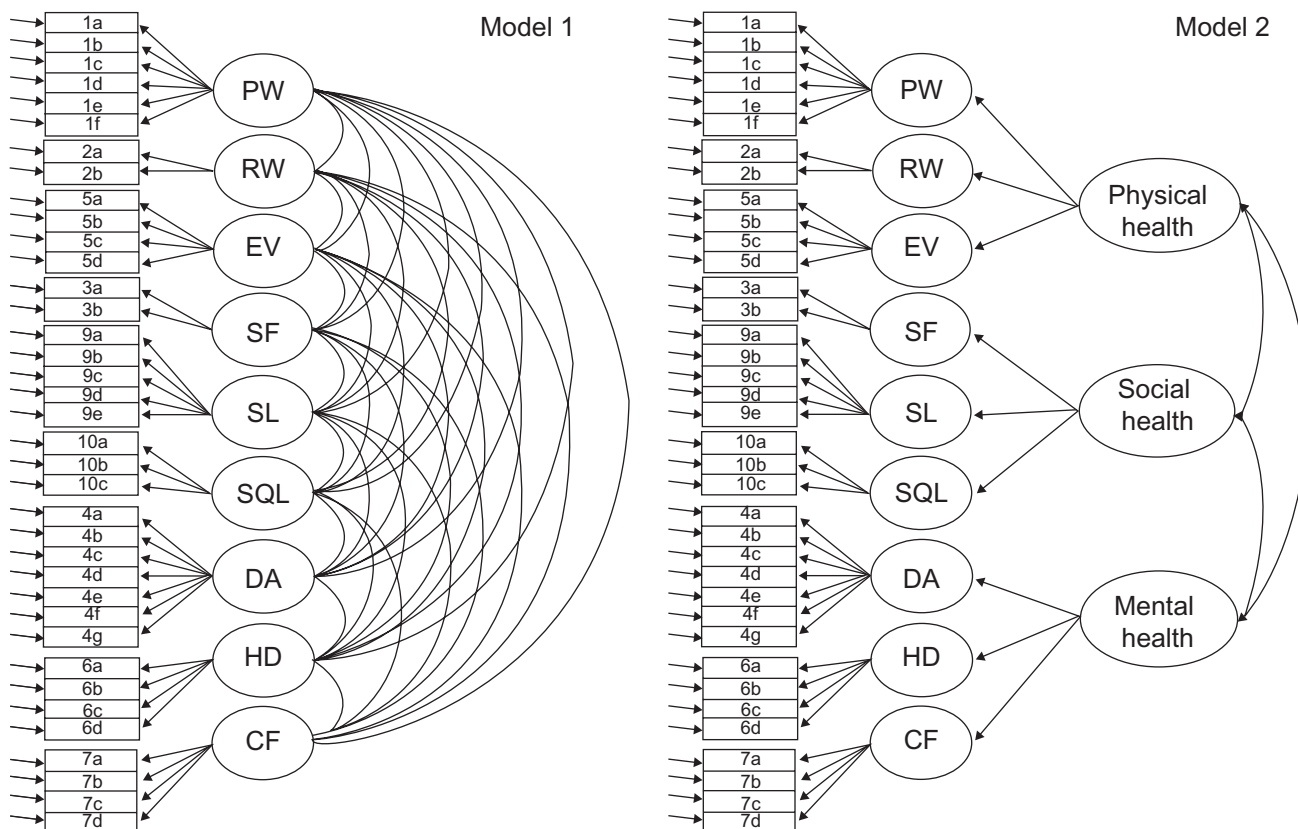


Figure 1 Path diagrams representing alternative factor models of the Italian National Institute of Health Quality of Life Instrument, Core Evaluation Form.

was otherwise represented by only two indicators, where at least three indicator variables for each second-order factor are recommended.^{10,13}

Depending on size and direction of the second-order factor correlations resulting from Model 2, two alternative factor models can be hypothesized. Model 3 who equated the Mental and the Social Health factors by setting to 1.00 their correlation and by imposing equality constraints to the correlations of both Mental and Social health with the Physical Health factor. Model 4 was even more restrictive as it equated all the second-order factor model by setting to 1.00 all the correlations among them. Models 3 and 4 are formally equivalent to models in which there are two and one second-order factors, respectively¹⁰ (Figure 1).

Assessment of model fit

Structural equations were used to estimate parameters and to test hypotheses on the goodness of fit of alternative factor models by use of the program EQS (v. 6.1, Multivariate Software, Inc., Encino, CA, USA).^{11,12} Specifically, the maximum likelihood (ML) robust method was used to estimate the models, as the observed data significantly violated the assumptions of multivariate normality (Mardia's normalized coefficient = 28.30). This method not only

provides researchers with unbiased parameter estimates and corrected standard errors for non-normal data, but it also helps in correcting many of the model's fit indices when deviation from multivariate normality occurred¹³. Historically, the model's fit was evaluated by the maximum likelihood chi-square statistic (or alternatively by the Satorra–Bentler¹⁴ scaled chi-square if the robust method was applied) in order to test the null hypothesis that the observed item covariance matrix and the reproduced covariance matrix are equal. However, because virtually any factor model could be rejected if the sample size is large enough, many authors^{15,16} recommended to supplement the evaluation of the model's fit by more "practical" indices of fit. A relative chi-square (ie, the chi-square statistic divided by the model's degrees of freedom is one of such indicators) lesser than 2:1 indicates substantial good fit.¹⁷ The comparative fit index¹⁸ (CFI) and the root mean square error of approximation¹⁹ (RMSEA) are among the most widely used fit indices.¹⁹ They assess the model's fit by different perspectives and their joint use is strongly recommended.¹⁹ The CFI is an incremental fit index which compares the hypothesized model's chi-square with that resulting from the independence model (ie, the model assuming that all relationships among measured variables are 0). By convention⁸ a CFI greater

than 0.90 indicates an acceptable fit to the data, with values greater 0.95 being very valued. The RMSEA measures the difference between the reproduced covariance matrix and the population covariance matrix, so that sampling variability is controlled for. The RMSEA is indeed a ‘badness of fit’ index, with values very close to 0 indicating almost perfect fit and with greater RMSEA indicating worse fit. By convention²⁰ a RMSEA lesser than 0.06 is considered as “good” fit, while a RMSEA lesser than 0.08 corresponds to an “acceptable” fit. The 90% confidence interval (CI) around the RMSEA point estimate is also commonly reported to indicate the possibility of close or exact fit.

Comparison of models

The Akaike information criterion²¹ (AIC) is specific a goodness-of-fit measure which corrects the model’s chi-square to penalize for model complexity. Unlike fit indices reported above, the AIC has no intuitive value nor recommended standards. The AIC is, however, useful for comparing alternative factor models under the expectation that the lower the AIC the better the model’s fit. Though, the AIC allows for quantitative comparisons, it does not allow for testing whether a model fitted the data significantly better than a competing one. Alternative factor models can be compared statistically if they are nested (ie, if one factor model can be derived by placing restrictions on another model). This is the case of models 3 and 4 who were both derived from Model 2. A second-order CFA model is also nested within the corresponding first-order model, provided that all the first-order factors were correlated. This the case of models 2, 3, and 4, who are all nested within Model 1.¹⁰ Under these specific circumstances, a chi-square difference test may be used to compare nested models. The goal of this test is to verify whether placing restrictions on a model (ie, reducing the number of free parameters) worsened the model’s fit significantly. If the comparison turns out to be statistically significant the less restrictive of the two models is preferred. It’s worth noting that, when comparing nested models which have been estimated by the maximum likelihood robust method, it is not correct to carry out the chi square difference test by merely replacing the standard chi square value with the Satorra–Bentler one, rather one should apply appropriate correction factors devised by Satorra and Bentler themselves.¹⁴

Testing factorial invariance

The CFA approach allows for comparing factor models estimated by different groups. Researchers interested in construct validity typically seek for invariance of the measurement model

across groups as a proof of crossvalidation of questionnaire data. In so doing, the equality of parameters sets (eg, factor loading paths, factor covariances, structural regression paths) across groups is tested in a logically ordered and increasingly restrictive fashion.¹³ The initial step requires only that the same number of factor and factor-loadings be the same across groups. This analysis, that is just a multigroup-representation of the best fitting model resulting from the whole sample, serves as a baseline onto which next analyses will be compared. The test of factor-loading equality follows that of configuration equality. Next, it follows the test of factor covariances and structural regression paths equality depending on whether the model is a single-order factor model or a second-order one.

Missing data handling

No imputation of missing values was done in this study, rather we have dealt with this problem by considering traditional approaches, such as the complete case analysis (listwise method) and the computation of sample’s statistics based on all available information (pairwise method). Though the listwise option is one the most widely used options, it dramatically reduces the sample size, especially if the analysis involves many variables. This may cause a substantial loss of precision in estimation as well as in power. Differently, the pairwise approach uses all the available univariate and bivariate information in the data to compute summary statistics. The main inconvenient with the pairwise method is, however, that there might be some inconsistencies in the estimated item covariance matrix as each of these statistics might have been computed based on a different number of patients. In our specific case, if the listwise approach would be chosen the single-group analyses would have been completed on a total of 210 patients (ie, 79.5% of the total cases) with a noticeable loss of precision. Differently, if the pairwise approach would be chosen the single-group analyses would have been completed on average on 254 patients (ie, 97.3% of the total cases) with a negligible loss of precision. Initially, we have carried out the analysis with both methods. However, since the results obtained with different approaches did not change substantially, we have presented in this study those obtained by using all available information. In addition, the pairwise was the only viable approach for assessing the invariance of the factor models in multi-group analyses.

Results

Baseline characteristics of the study population

Of the 261 persons sampled for this work, 258 reported baseline demographic and clinical data. In this study 84%

of the subjects were male; 38% were heterosexual and 37% homosexual. The median age was 44 years. The subjects were HIV-infected for a median of 11 years and about 70% of them had a history of a symptomatic condition. Median HIV viral load was 1.7 log and median CD4 cell count was 539 cells/ μ l (Table 1). The median baseline ISSQoL-CEF domains ranged from 50 to 100. Three of the domains (physical well-being, role-well-being and social functioning) had relatively high percentages (>20) of persons scoring at the highest possible scale level (ceiling effects) (Table 2).

Confirmatory factor analysis: single group analyses

Fit indices for alternative factor models of the CEF have been reported in Table 3. The inspection of the goodness of

Table 1 Demographic and clinical characteristics of patients at baseline

Variables	Statistics
N°, patients	261
Gender	
– Female: <i>n</i> (%)	41 (15.7)
– Male: <i>n</i> (%)	217 (83.1)
– missing information:	3 (1.2)
Age (years):	
Mean \pm SD (<i>n</i> , range)	45.0 \pm 8.2 (251, 24–71)
median	44.0
Predominant risk factor: <i>n</i> (%)	
– Homo/bisexual:	96 (36.8)
– I.V. drug use:	58 (22.2)
– Heterosexual:	99 (37.9)
– Hemophilic:	1 (0.4)
– Other:	4 (1.5)
– missing information:	3 (1.2)
CD4+/mm^3	
mean \pm SD (<i>n</i> , range)	613 \pm 572 (256, 16–7769)
median	539
HIV-RNA cp/ml (log₁₀)	
mean \pm SD (<i>n</i> , range)	2.2 \pm 1.2 (255, 0–5.7)
median	1.7
Stage at randomization: <i>n</i> (%)	
– CDC A:	66 (25.6)
– CDC B:	104 (40.3)
– CDC C:	85 (32.9)
– missing information:	3 (1.2)
Time from HIV diagnosis (years)	
mean \pm SD (<i>n</i> , range)	11.8 \pm 6.0 (253, 1–23)
median	11.0

Table 2 Summary of quality of life scores at baseline

Domain scores	Total sample
	Mean \pm SD (<i>n</i> , range); median % floor; % ceiling
Physical well-being	72 \pm 28 (257,0–100); 79.0.4; 21.0
Role well-being	77 \pm 30 (244,0–100); 100.5.3; 50.4
Energy/Vitality	65 \pm 22 (258,6–100); 69.0.0; 7.8
Social functioning	76 \pm 28 (248,0–100); 88.2.4; 43.5
Sexual life	54 \pm 27 (246,0–100); 60.3.7; 2.0
Satisfaction with quality of life	52 \pm 21 (259,0–100); 50.1.5; 2.7
Anxiety/Depression	67 \pm 22 (258,11–100); 68.0.0; 6.6
Health distress	63 \pm 26 (259,0–100); 67.2.7; 12.0
Cognitive functioning	70 \pm 23 (258,0–100); 75.0.4; 17.1

fit statistics, revealed that all the models had a statistically significant $\text{SB}\chi^2$ statistic, regardless of the number of factors and of whether models assumed a hierarchical arrangement of factors or not. The relative chi-square ($\text{SB}\chi^2/\text{df}$) was, however, lesser than 2 for all models, thus indicating that the statistical significance of the $\text{SB}\chi^2$ was mostly due to its sensitivity to relatively large sample size, rather than to substantially bad fit. The inspection of both incremental and absolute fit indices (Table 3) revealed that Model 1, with nine correlated first-order factors, not only had a quite good fit to the data (ie, CFI = 0.923, RMSEA = 0.055), but it also fitted better the data than all the models with second-order factors. In fact, not only Model 1 met with all the conventional standards (ie, CFI > 0.90, RMSEA < 0.06), while Models 2, 3, and 4 in some cases failed to do so, but it also had a more negative (ie, lesser) AIC than all hierarchical factor models. This latter finding indicated that though Models 2, 3, and 4 had lesser parameters than Model 1, their relative gain in parsimony was not counterbalanced by a negligible loss of fit, which unavoidably resulted from imposing restrictions to the first-order factor correlations.

The inspection of Model 1 parameters (Table 4, Panel a) showed, however, that the factor-correlations assessed for Model 1 were moderately high (median $r = 0.54$; $Q_3 - Q_1 = 0.66 - 0.43$). Whereas one may read this finding just as an empirical support for the convergent validity of the nine specific domain factors, such correlations might be also

Table 3 Fit Indices for alternative factor models of the ISSQoL-CEF

Models	ML χ^2	AIC	SB χ^2	df	SB χ^2/df	CFI	RMSEA	RMSEACI
1. Nine first-order factors, No second-order factors	1352.57	-121.56	1064.43	593	1.79	0.923	0.055	0.050–0.060
2. Nine first-order factors, Physical, Mental and Social Health second-order factors	1502.06	-68.342	1165.65	617	1.89	0.911	0.058	0.053–0.063
3. Nine first-order factors, Physical and Mental/Social Health second-order factors	1520.49	-60.050	1177.95	619	1.90	0.909	0.059	0.054–0.064
4. Nine first-order factors. General Health second-order factor	1540.85	-55.248	1184.75	620	1.91	0.908	0.059	0.054–0.064

Note: All fit indices were corrected based on the maximum likelihood robust method.

Abbreviations:

viewed as an indication that one or more superordinate factors might have been influenced the patient ratings. Despite this hypothesis was rejected based on the mere statistical lecture of the AIC indices (Table 3) and despite Model 2 also resulted in a significant decline of the model's fit relative to Model 1 ($\Delta SB\chi^2 = 108.26$; $df = 27$; $P < 0.001$), it looks like that Model 2 still had an acceptable fit to the data. Differently, Models 3 and 4, who also had an acceptable fit to the data, resulted in a significantly worse fit than Model 2 as it concerns both the AICs and the chi-square scaled difference tests, which turned out to be significant when comparing Model 2 to both Model 3 ($\Delta SB\chi^2 = 9.54$; $df = 2$; $P < 0.01$) and Model 4 ($\Delta SB\chi^2 = 10.49$; $df = 3$; $P < 0.01$).

Once we have established that Model 1 was the overall best fitting model and that Model 2 provided a more articulated view of the ISSQoL-CEF factorial structure, while maintaining an acceptable fit, we inspected the standardized solution resulting from both models. As it concerns the first-order measurement model, whose implication for reliability and validity of the ISSQoL-CEF domains are essential, our inspection of the standardized solutions obtained for Model 1 (Table 4, Panel b) and Model 2 (Table 4, Panel a) provided very consistent estimates. Not only the factor loading pattern was the same, but also the coefficients did not differ substantially across models. The coefficients were in most cases larger than 0.70 (ie, more than 50% of common variance between items and factors). These finding indicated that, whatever the model chosen to represent the factorial structure of the questionnaire, the reliability of its domain factor scores was expected to be high and relatively unbiased. Accordingly, the reliability coefficients assessed for the nine first-order factors were in most cases much above the required psychometric standards²² (Table 4, Panel a and b).

As to the second-order measurement model (Table 4, Panel b), whose parameters are of a some importance in

order to get to global health indicators, each of the first-order factors regressed on the appropriate second-order factor, significantly. The loadings were particularly high for the Mental Health and the Physical Health factors, which represented the cornerstones of any HRQoL instrument. With regard to the Social Health factor, the overall satisfaction with quality of life regressed on this second-order factors as much as the social functioning did, while the factor loading for the sexual life was slightly lower. The reliability coefficients assessed for each of the second-order factors were again much above the required psychometric standards²² (Table 4, Panel b).

Confirmatory factor analysis: multiple-group analyses

In the present study, we have merged data from four different clinical studies in order to get to an acceptable sample size for a CFA study (ie, at least five times the number of items). So far, the analyses considered all four clinical samples as being drawn from the same population of patients. However, there is some potential for heterogeneity in the dataset which might be exploited in order to gather some information on whether the ISSQoL-CEF factor models fitted to the whole sample data may be generalized to different subpopulations of patients. In this case, the sample size was suboptimal for conducting such a rigorous test of factorial invariance (ie, all clinical studies included less than five patients per item), we have compared the factorial invariance of Models 1 and 2 considering $N = 134$ (Clinical study 1) and $N = 127$ (Clinical studies 2, 3, and 4) patients. We started our analysis with evaluating the equality of configuration for both models by a test of the assumption that the number of factor and the factor-loadings was the same in the two groups. As to Model 1 (Figure 1, Panel a), the configurational equality analysis resulted in a good fit to the data ($SB\chi^2 = 1702.92$; $df = 1186$; $AIC = -669.07$; $CFI = 0.919$; $RMSEA = 0.058$). Differently,

Table 4 Standardized factor solutions for models 1 and 2

b) Model 2																			
First-order factor loadings																			
Item/ Factor	PW	RW	EW	SF	SL	SQL	DA	HD	CF	Item/ Factor	PW	RW	EW	SF	SL	SQL	DA	HD	CF
1a	,47	-	-	-	-	-	-	-	-	1a	,47	-	-	-	-	-	-	-	-
1b	,90	-	-	-	-	-	-	-	-	1b	,90	-	-	-	-	-	-	-	-
1c	,85	-	-	-	-	-	-	-	-	1c	,85	-	-	-	-	-	-	-	-
1d	,84	-	-	-	-	-	-	-	-	1d	,85	-	-	-	-	-	-	-	-
1e	,88	-	-	-	-	-	-	-	-	1e	,87	-	-	-	-	-	-	-	-
1f	,84	-	-	-	-	-	-	-	-	1f	,83	-	-	-	-	-	-	-	-
2a	-	,91	-	-	-	-	-	-	-	2a	-	,91	-	-	-	-	-	-	-
2b	-	,84	-	-	-	-	-	-	-	2b	-	,84	-	-	-	-	-	-	-
5a	-	-	,64	-	-	-	-	-	-	5a	-	-	,68	-	-	-	-	-	-
5b	-	-	,76	-	-	-	-	-	-	5b	-	-	,72	-	-	-	-	-	-
5c	-	-	,61	-	-	-	-	-	-	5c	-	-	,55	-	-	-	-	-	-
5d	-	-	,82	-	-	-	-	-	-	5d	-	-	,87	-	-	-	-	-	-
3a	-	-	-	,92	-	-	-	-	-	3a	-	-	-	,91	-	-	-	-	-
3b	-	-	-	,74	-	-	-	-	-	3b	-	-	-	,75	-	-	-	-	-
8a	-	-	-	-	,57	-	-	-	-	8a	-	-	-	-	,57	-	-	-	-
8b	-	-	-	-	,76	-	-	-	-	8b	-	-	-	-	,76	-	-	-	-
8c	-	-	-	-	,92	-	-	-	-	8c	-	-	-	-	,92	-	-	-	-
8d	-	-	-	-	,62	-	-	-	-	8d	-	-	-	-	,62	-	-	-	-
8e	-	-	-	-	,84	-	-	-	-	8e	-	-	-	-	,84	-	-	-	-
9a	-	-	-	-	-	,73	-	-	-	9a	-	-	-	-	-	,74	-	-	-
9b	-	-	-	-	-	,76	-	-	-	9b	-	-	-	-	-	,75	-	-	-
9c	-	-	-	-	-	,86	-	-	-	9c	-	-	-	-	-	,87	-	-	-
4a	-	-	-	-	-	-	,69	-	-	4a	-	-	-	-	-	-	,69	-	-
4b	-	-	-	-	-	-	,74	-	-	4b	-	-	-	-	-	-	,74	-	-
4c	-	-	-	-	-	-	,77	-	-	4c	-	-	-	-	-	-	,76	-	-
4d	-	-	-	-	-	-	,53	-	-	4d	-	-	-	-	-	-	,54	-	-
4e	-	-	-	-	-	-	,74	-	-	4e	-	-	-	-	-	-	,74	-	-
4f	-	-	-	-	-	-	,81	-	-	4f	-	-	-	-	-	-	,81	-	-
4g	-	-	-	-	-	-	,83	-	-	4g	-	-	-	-	-	-	,83	-	-
6a	-	-	-	-	-	-	-	,89	-	6a	-	-	-	-	-	-	-	,89	-
6b	-	-	-	-	-	-	-	,92	-	6b	-	-	-	-	-	-	-	,93	-
6c	-	-	-	-	-	-	-	,85	-	6c	-	-	-	-	-	-	-	,85	-
6d	-	-	-	-	-	-	-	,78	-	6d	-	-	-	-	-	-	-	,78	-
7a	-	-	-	-	-	-	-	-	,63	7a	-	-	-	-	-	-	-	-	,63
7b	-	-	-	-	-	-	-	-	,69	7b	-	-	-	-	-	-	-	-	,68
7c	-	-	-	-	-	-	-	-	,94	7c	-	-	-	-	-	-	-	-	,94
7d	-	-	-	-	-	-	-	-	,92	7d	-	-	-	-	-	-	-	-	,92

First-order factor correlations																
PW	1,00															
RW	,74	1,00														
EV	,44	,50	1,00													
SF	,46	,66	,64	1,00												
SL	,19	,29	,43	,45	1,00											
SQL	,43	,50	,72	,58	,42	1,00										
DA	,46	,56	,83	,70	,50	,74	1,00									
HD	,33	,44	,69	,62	,54	,65	,82	1,00								
CF	,35	,49	,66	,60	,31	,57	,82	,62	1,00							
PW		RW		EV		SF		SL		SQL		DA		HD		CF
Second-order factor loadings and correlations																
P																
S																
M																
Rho	,96	,88	,88	,70	,91	,85	,95	,96	,92	,91	,85	,95	,96	,92	,91	,89

Second-order factor loadings and correlations																
P	,74	,86	,65													
S										,52	,79					
M														,99	,85	,76
Rho	,96	,88	,87	,69	,91	,85	,95	,96	,92	,91	,85	,95	,96	,92	,91	,89

Abbreviations: PW, physical well-being; RW, role well-being; SF, social functioning; DA, depression/anxiety; EV, energy/vitality; HD, health/distress; CF, cognitive functioning; SL, sexual life; SQL, overall satisfaction with quality of life; P, physical health; S, social health; M, mental health; Rho, factor reliability coefficient.

Model 2 yielded a slightly suboptimal fit to the data as it concerns the CFI, while the RMSEA was in keeping with the conventional standard for an acceptable fit ($SB\chi^2 = 1888.92$; $df = 1240$; $AIC = -592.25$; $CFI = 0.899$; $RMSEA = 0.064$). The configurational equality analysis, not only provided evidence that the path model was the same in both group, but it also served as a baseline to evaluate whether more stringent forms of invariance were tenable. After we have imposed between groups equality constraints to the first-order factor loadings, both Model 1 ($SB\chi^2 = 1746.11$; $df = 1214$; $AIC = -681.88$; $CFI = 0.917$; $RMSEA = 0.058$) and Model 2 ($SB\chi^2 = 1812.51$; $df = 1268$; $AIC = -687.88$; $CFI = 0.897$; $RMSEA = 0.064$) did not differ from the baseline analysis significantly ($\Delta SB\chi^2 = 44.99$; $df = 28$; $P = 0.03$ and $\Delta SB\chi^2 = 40.25$; $df = 28$; $P = 0.06$, respectively for Model 1 and 2). This finding showed that the measurement model of the ISSQoL-CEF was the same between groups.

Next, we have imposed parameters equality to the factor covariance matrix on Model 1 and to the structural regression paths (ie, the second-order factor loadings) on Model 2. This analysis for both Model 1 ($SB\chi^2 = 1812.51$; $df = 1250$; $AIC = -687.88$; $CFI = 0.912$; $RMSEA = 0.059$) and Model 2 ($SB\chi^2 = 1975.03$; $df = 1277$; $AIC = -578.97$; $CFI = 0.891$; $RMSEA = 0.065$), resulted in a significant worse fit relative to the baseline model ($\Delta SB\chi^2 = 122.36$; $df = 64$; $P < 0.01$; $P = 0.03$ and $\Delta SB\chi^2 = 102.53$; $df = 37$; $P < 0.01$, respectively). Thus, we have conclude that the factor models of the CEV have passed two out of three tests of factorial invariance.

Discussion

The most recent international guidelines on the use of HAART²³ pointed out that one of the four main goals of the therapy is to improve patients' quality of life. A great number of specific questionnaires for HIV-infected people are available.²⁴⁻²⁸ However, since they were developed in the pre-HAART period, they did not include some currently important aspects of HRQoL, such as the quality of sexual life or the fatigue experienced in performing daily activities. In the present study we have tested the construct validity of a new HRQOL instrument for HIV patients in the HAART era, since providing empirical evidence supporting this type of validity is deemed^{3,5,7,8} as mandatory before licensing a any new psychometric tool as a valuable assessment in clinical practice and research.

Alternative factor models have been fitted to the data collected from 261 HIV patients participating to four different clinical studies. The factor structure with nine correlated factors, each representing one of the ISSQoL-CEF domains,

was the overall best fitting model. All fit indices were much above the standard for an acceptable fit and in most cases they met with the requirements for concluding that the model had a good fit.

However, whereas this model had nice statistical properties and it was in keeping with previous studies of the ISSQoL,¹ it might have been useful to consider a higher-order factor structure to account for the correlation among lower order factors. However, as it is common in CFA studies, higher-order factor models like this are likely to result in a worse fit to the data than models who assume a single-order structure with correlated factors. Reasons for such declining of fit might be merely found in the objective lesser number of parameters specified for higher-order factor models which, in turn, lead to an oversimplified representation of the collected data (ie, models with more parameters fit the data better) as well as in the fact that there might some residual correlation among first-order factor that second-order factors are unable to capture (ie, second-order factors can explain only a limited part of the covariance among first-order factors). Thus, the issue of comparing a first-order factor models with second-order factor models its unlikely to result in the choice of a hierarchical structure based on the lecture of their statistical properties, unless each of the first-order factors has a near perfect regression coefficient on only one of the second-order factors. In spite of these technical hitches, we have considered a second-order model with three higher-order factors which mirrored the Physical, Social, and Mental components of the WHO multidimensional definition of health.³ As expected, the second-order factor model had a relatively worse fit than the single-order factor model. However, it is worth noting that the factor model with Physical, Social, and Mental health as second order factors was quite defensible, as it resulted in an overall acceptable fit to the data, despite the choice of modeling the correlations among the nine first-order factors by a few structural paths linking second-order factors to first-order factors.

Some important conclusion can be drawn regardless of which is the preferred model. First, the factor-loading matrix of the ISSQOL-CEF items on the nine first-order factors yielded fairly high coefficients, regardless of whether we have inspected the single order or the second-order standardized solution. This finding not only supported the validity of the CEF but it also has demonstrated that its nine domain factors were measured with a remarkably high degree of reliability. In fact, the item-factor relations were so empirically robust that they were relatively unbiased by how we have modeled the covariance among the nine factors (ie, letting the nine factors

freely correlate or constraining their correlation structure by positing higher order factors). The robustness of item-factor relations was also corroborated by multisample analyses who were aimed at investigating whether each of the factor models fitted to the whole sample could have been generalized to different clinical populations. Though carried out with a sub-optimal sample size (ie, the two groups of patients were not as large as required for a such rigorous test and not all the samples drawn from different clinical studies could have been compared due to their relatively small sample size), these analyses supported the invariance of the model's configuration as well as that of the factor-loading parameters, while only most restrictive forms of invariance, such as that of factor-covariance or that of structural paths were not tenable.

In summary, all the findings reported in this study provided strong support to the construct validity of the ISSQoL-CEF, especially as it regards the profiling of HIV patients along its nine domains. The most important restriction of this study was, however, represented by the suboptimal sample size, which was a critical aspect especially when multisample factor analyses were carried out. While the size of the whole sample was acceptable for a CFA study, it was barely enough to test the multisample invariance of alternative factor models. Despite this limitation, we may conclude that we have successfully attained the main goal of the study, to show that the ISSQoL-CEF, a recently developed questionnaire specific to HIV patients in the HAART age, is a valid and reliable tool.

As pointed out in psychometric and HRQoL literature,²⁹ only once the questionnaire construct validity has been verified by confirmation of its hypothesized dimensional structure, other types of scale refinement might be considered, including the cross-cultural adaptation of our instrument or the development of a short-form version by distilling the ISSQoL-CEF into a few key questions. The next steps of our psychometric research should go in both these directions.

Acknowledgments

The authors would like to thank S Merli, O Viganò ("L Sacco" Hospital, Milan) for their contribution in acquisition of data of clinical study 4. The authors report no conflicts of interest in this work.

References

1. Bucciardini R, Murri R, Guarinieri M, et al. ISS-QoL: a new questionnaire for evaluating the quality of life of people living with HIV in the HAART era. *Qual Life Res.* 2006;15(3):377-390.
2. Palella FJ Jr, Delaney KM, Moorman AC, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *N Engl J Med.* 1998;338:853-860.

3. Spilker B. *Quality of Life Assessment in Clinical Trials*. New York, NY: Raven Press; 1996.
4. Wu AW. Quality of life assessment comes of age in the era of highly active antiretroviral therapy. *AIDS*. 2000;14:1449–1451.
5. Nunnally JO. *Psychometric Theory*. New York, NJ: McGraw-Hill; 1978.
6. Hays RD, Hayashi T. Beyond internal consistency reliability: rationale and user's guide for the multitrait analysis program (MAP) on the microcomputer. *Behavioral Research Methods, Instrumentation and Computers*. 1990;22:167–175.
7. Anastasi A. *Psychological Testing*. New York, NY: Macmillan; 1988.
8. Fink A. *How to Measure Survey Reliability and Validity v. 7*. Thousand Oaks, CA: Sage Publications; 1995.
9. Stevens J. Exploratory and confirmatory factor analysis. In: *Applied Multivariate Statistics for the Social Sciences*. 4th eds. Lawrence; 2002.
10. Rindskopf D, Rose T. Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behav Res*. 1988;23:51–67.
11. Brown MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. *Testing Structural Equation Models*. Newbury Park, CA: Sage Publications; 1993:136–162.
12. Bentler PM. *EQS structural equations modeling software (Version 6.1) [Computer software]*. Encino, CA: Multivariate Software Inc.; 2004.
13. Byrne BM. *Structural equation modeling with EQS: Basic concepts, applications and programming*. Newbury Park, CA: Sage Publications; 2006.
14. Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 2001;66:507–514.
15. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychol Meth*. 1999;4:272–299.
16. McDonald RP, Ho MR. Principles and practice in reporting structural equation analyses. *Psychol Meth*. 2002;7:64–82.
17. Kline RB. *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford; 1998.
18. Widaman KF, Thompson JS. On specifying the null model for incremental fit indices in structural equation modeling. *Psychol Meth*. 2003;8:16–37.
19. Hu L, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*. 2004;6:1–55.
20. Bentler PM. Comparative fit indices in structural models. *Psychol Bull*. 1990;107:238–246.
21. Bozdogan H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*. 1987;52:345–370.
22. Carmines EG, Zeller RA. *Reliability and Validity Assessment*. Newbury Park, CA: Sage Publications; 1991.
23. US Department of Health and Human Services. Guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents. 2005. Available from: <http://www.aidsinfo.nih.gov/>. Accessed on January 10, 2005.
24. Wu AW, Rubin HR, Mathews WC, et al. A health status questionnaire using 30 items from the Medical Outcomes Study. Preliminary validation in persons with early HIV infection. *Med Care*. 1991;29:786–798.
25. Fang CT, Hsiung PC, Yu CF, Chen MY, Wang JD. Validation of the World Health Organization quality of life instrument in patients with HIV infection. *Qual Life Res*. 2002;11:753–762.
26. Starace F, Cafaro L, Rucci P, et al. Quality of life in HIV seropositive persons: application and validation of the WHOQOL-HIV, Italian version. *AIDS Care*. 2002;14:405–415.
27. De Boer JB, Sprangers MA, Aaronson NK, Lange JM, van Dam FS. A study of the reliability, validity and responsiveness of the HIV overview of problems evaluation system (HOPES) in assessing the quality of life of patients with AIDS and symptomatic HIV infection. *Qual Life Res*. 1996;5:339–347.
28. Peterman AH, Cella D, Mo F, McCain N. Psychometric validation of the revised Functional Assessment of Human Immunodeficiency Virus Infection (FAHI) quality of life instrument. *Qual Life Res*. 1997;6:572–584.
29. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med*. 1993;118:622–629.

Appendix I

Tentative translation of the ISSQoL-CEF items

1. [PW] In the past 4 weeks, did you feel your health restricted your ability to perform any of the following physical activities?
 - a) Physically demanding activities such as running for a long time, lifting heavy objects or engaging in strenuous sports
 - b) Everyday activities, like moving a table, carrying shopping bags, cleaning the house
 - c) Walking uphill or climbing a few flights of stairs
 - d) Bending over, getting up or crouching down
 - e) Walking at a leisurely pace for a half an hour
 - f) Eating, dressing, bathing, or getting up and down from the toilet
2. [RW] In the past 4 weeks, did you feel your health restricted your ability to perform the following activities?
 - a) Performing normal tasks at work, home or school?
 - b) Looking for a job or keeping a job?
3. [SF] For each of the following questions, check the answer that best describes your situation in the past 4 weeks.
 - a) Did your health limit your social activities?
 - b) Did your prescribed drugs limit your social activities?
4. [DA] For each of the following questions, check the answer that best describes your situation in the past 4 weeks.
 - a) Did you feel unable to overcome possible moments of depression?
 - b) Did you have problems concentrating on what you were doing?
 - c) Did you feel that everything you did was an effort?
 - d) Did you have trouble sleeping?
 - e) Did you feel lonely?
 - f) Did you feel sad?
 - g) Did you feel as if you could not start any activity?
5. [EV] For each of the following questions, check the answer that best describes your situation in the past 4 weeks.
 - a) Did you have enough energy to perform your everyday activities?
 - b) Did you feel energetic and cheerful?
 - c) Did you feel exhausted?
 - d) Did you feel strong enough to do what you had planned to do?
6. [HD] For each of the following questions, check the answer that best describes your situation in the past 4 weeks.
 - a) Did your health condition make you feel burdened?
 - b) Did your health condition make you feel discouraged?
 - c) Did your health condition make you feel hopeless?
 - d) Did you feel frightened because of your health?
7. [CF] For each of the following questions, check the answer that best describes your situation in the past 4 weeks.
 - a) Did you have difficulty solving problems or, for example, making plans or decisions, or learning new things?
 - b) Did you have short-term memory loss, such as forgetting an appointment or where you put something, etc.?
 - c) Did you have difficulty keeping your attention on any activity for a long time?
 - d) Did you have difficulty performing activities that required concentration and mental effort?
8. [SL] For each of the following questions, check the answer that best describes your situation in the past 4 weeks.
 - a) Did you feel satisfied with your sexual activity?
 - b) Did being HIV positive limit your sexual life?
 - c) Did your health condition make your sexual life worse?
 - d) Did your antiretroviral therapy make your sexual life worse?
 - e) Did your sexual life become worse since the infection was diagnosed?
9. [SQL] For each of the following questions, check the answer that best describes your situation
 - a) Are you satisfied with your health?
 - b) Overall, are you satisfied with your mental and emotional state?
 - c) Overall, are you satisfied with your physical condition?

Patient Preference and Adherence

Publish your work in this journal

Patient Preference and Adherence is an international, peer-reviewed, open access journal that focusing on the growing importance of patient preference and adherence throughout the therapeutic continuum. Patient satisfaction, acceptability, quality of life, compliance, persistence and their role in developing new therapeutic modalities and compounds to

Submit your manuscript here: <http://www.dovepress.com/patient-preference-and-adherence-journal>

optimize clinical outcomes for existing disease states are major areas of interest. This journal has been accepted for indexing on PubMed Central. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Dovepress