Quality of Life Research **13**: 299–310, 2004.

The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial A Report from the WHOQOL Group☆

S.M. Skevington[1], M. Lotfy[2] & K.A. O'Connell[1]
[1]*WHO Centre for the Study of Quality of Life, Department of Psychology, University of Bath, Bath, UK (E-mail: s.m.skevington@bath.ac.uk);* [2]*Department of Mental Health and Substance Dependence, World Health Organisation, Geneva, Switzerland*

Accepted in revised form 5 March 2003

Abstract

Quality of life (QOL) assessments that are easily administered and which do not impose a great burden on the respondent are needed for use in large epidemiological surveys, clinical settings and clinical trials. Using data from the WHOQOL-BREF field trials, the objectives of this work are to examine the performance of the WHOQOL-BREF as an integrated instrument, and to test its main psychometric properties. The WHOQOL-BREF is a 26-item version of the WHOQOL-100 assessment. Its psychometric properties were analysed using cross-sectional data obtained from a survey of adults carried out in 23 countries (n = 11,830). Sick and well respondents were sampled from the general population, as well as from hospital, rehabilitation and primary care settings, serving patients with physical and mental disorders and with respect to quotas of important socio-demographic variables. The WHOQOL-BREF self-assessment was completed, together with socio-demographic and health status questions. Analyses of internal consistency, item–total correlations, discriminant validity and construct validity through confirmatory factor analysis, indicate that the WHOQOL-BREF has good to excellent psychometric properties of reliability and performs well in preliminary tests of validity. These results indicate that overall, the WHOQOL-BREF is a sound, cross-culturally valid assessment of QOL, as reflected by its four domains: physical, psychological, social and environment.

Key words: Assessment, Cross-cultural, Quality of life, WHOQOL-BREF

Introduction

Increasingly, health care planners are recognizing that measures of disease alone are insufficient determinants of health status. Over the past decades, two classes of complementary health status measures have emerged to fill the information gap – objective measures of functional health status and subjective measures of health and well-being (for reviews see Refs. [1–4]). These measures are multilevel and multi-dimensional. There are many published quality of life (QOL) measures but there is still a lack of consensus among researchers about its definition and this is reflected in the choice of items for their instruments.

The WHO defines QOL as 'an individual's perception of their position in life in the context of the culture and value systems in which they live, and in relation to their goals, expectations, standards and concerns' [5]. In measuring QOL therefore, the WHOQOL Group takes the view that it is important to know how satisfied or bothered people are by important aspects of their life, and this interpretation will be a highly individual matter. The

☆ See Appendix 1 for details.

World Health Organisation Quality of Life assessment – the WHOQOL-100 – is a cross-culturally valid assessment of well-being. Assessment is operationalized through 100 items representing 25 facets organised in six domains [6, 7]. The tool was developed through a collaboration of 15 sites around the world working in their own national language. Centres simultaneously used common protocols that were agreed through international consensus at each stage of development process. The WHOQOL collaboration pooled information throughout the project and this procedure not only permits a high level of semantic and conceptual equivalence to be achieved between language versions but also creates a 'fast track' to the rapid establishment of multi-lingual instruments [7]. This new procedure whereby centres work simultaneously on the same stage of instrument development, pooling their ideas and results centrally (through WHO Geneva) and communicating with each other to achieve equivalence has been described as a 'spoke-wheel' methodology, through analogy with the spokes and hub of a bicycle wheel [8].

The WHOQOL-BREF is being developed as a short version of the WHOQOL-100 for use in situations where time is restricted, where respondent burden must be minimised and where facet-level detail is unnecessary e.g. with large epidemiological surveys and some clinical trials. Using data from 15 centres collected for the WHOQOL-100 field trials, items for the WHOQOL-BREF were selected for their ability to explain a substantial proportion of variance within their parent facet and domain, for their relationship with the overall WHOQOL model and for their discriminant validity [9]. Analysis of these extracted items showed that a four-factor structure best fitted the data [9]. Although this contrasted with the original concept of a 6-domain model for the WHOQOL, it was consistent with empirical results from the previous WHOQOL-100 field trials [8]. Based on these results, the WHOQOL-BREF was developed in the context of four domains of QOL: physical, psychological, social and environment [9]. Although extensive analysis had been carried out on the WHOQOL-100 field trial data to reduce items and assess the preliminary psychometric properties of a short form [9], this extracted data was insufficient to confirm the WHOQOL-BREF's properties as an integrated instrument. For this reason, fresh data needed to be collected. In this study we present new survey data that represent the first field trial of the WHOQOL-BREF that seeks to confirm and extend information about the properties of the short form as a whole. Specifically we aimed to assess item–response distributions, internal consistency reliability and item–scale correlations with other aspects of construct and discriminant validity. Evidence of test–retest reliability for the WHOQOL-BREF is already known [9]. It was predicted that sick participants would report poorer QOL than well participants but no predictions were made for other socio-demographic and centre differences.

Methods

Design

Data for the WHOQOL-BREF field trial were collected using a cross-sectional design in 24 centres representing 23 countries. The centres were drawn from countries in all the WHO Regions of the world, as well as from diverse cultures and different levels of socio-economic development. Data were contributed from field sites in Argentina, Australia, Brazil, Bulgaria, China, Croatia, Germany, Greece, Hungary, Israel, Italy, India: Madras and New Delhi, Japan, Malaysia, Netherlands, Nigeria, Norway, Romania, Russia, Spain, Turkey, United Kingdom, United States. Data collection methods were similar to internationally agreed protocols designed during the development of the WHOQOL-100 [6, 7].

Adult participants (adult was culturally defined) were recruited from a variety of in-patient and outpatient health care facilities, and from the general population. Using a common and consensually agreed protocol, quota sampling was used to structure the sample so that equal numbers of each gender and the two age groups (bisected at 45 years) were targeted. Recruitment would span the continua of the adult age range, four educational levels and types of marital status. Well samples were targeted similarly. The sites aimed to recruit sick participants from all the main groups of health care users but did not use a quota for diagnosis or severity. Wellness or sickness was de-

fined by self-report, from diagnostic categories assigned by health professionals, and with reference to contextual knowledge about (non) patient status relating to the nature of the population(s) approached in collection sites set up by each centre. Together the centres would obtain a richly heterogeneous sample of sick people covering 28 groups of physical or mental health problems (linked to ICD-10 categories) and with varying levels of disease severity and functioning. A fully structured design and common protocol was not feasible for these variables due to an absence of relevant national statistics in some parts of the developing world, and limited resources for research.

Instrument

The WHOQOL-BREF is an abbreviated 26-item version of the WHOQOL-100 containing items that were extracted from the WHOQOL-100 field trial data. The WHOQOL-BREF contains one item from each of the 24 facets of QOL included in the WHOQOL-100, plus two 'benchmark' items from the general facet on overall QOL and general health (not included in the scoring) (see Table 1). The facets were originally subsumed within one of six domains but factor analysis of the WHOQOL-100 indicated that Domain 1 could be merged with Domain 3 (physical with independence), and Domain 2 with Domain 6 (psychological with spirituality, religion and personal beliefs) thereby creating four domains of QOL [8]. Similar results were found during the extraction of data for the WHOQOL-BREF [9] which is currently scored in four domains: Domain 1: Physical health, Domain 2: Psychological, Domain 3: Social relations and Domain 4: Environment, with all facet items scored as part of their hypothesised domain. Domains are not scored where 20% of items or more are missing, and are unacceptable where two or more items are missed (or 1-item in the 3-item social domain). The scores are transformed on a scale from 0 to 100 to enable comparisons to be made between domains composed of unequal numbers of items.

The WHOQOL-BREF was self-administered by respondents but exceptionally, an experienced interviewer assisted administration by reading items aloud where self-completion was not possible, usually for reasons of literacy or disability. Standard instructions, socio-demographic details and an item on current health status were completed before answering the 26 items of the WHOQOL-BREF.

During development of the WHOQOL-100, four types of 5-point Likert interval scale were designed and tested to reflect intensity, capacity, frequency and evaluation, and one of these was attached to each item [10]. These response scales were also used in the WHOQOL-BREF. Items inquire 'how much', 'how completely', how often', 'how good' or 'how satisfied' the respondent felt in the last 2 weeks; different response scales are distributed across the domains [10, 11]. The translation process used by the WHOQOL Group to develop linguistically and culturally appropriate new versions of the measure has been revised and updated from the WHO standard procedures and is reported elsewhere [6, 12, 13].

Analysis

Frequency, reliability and correlational analyses

Frequency analyses were performed to assess response distributions at the item level, globally and by country. In line with the WHOQOL-100 procedure, problematic items were identified as those where the response distribution was skewed such that fewer than 10% of responses fell in any two adjacent scale points for at least 12 of the 24 centres. Internal consistency was assessed using Cronbach's α and the contribution of each item to the total α. The average inter-item correlations for domains, and correlation of items with their intended domain (using corrected item–total correlations) were also calculated. Multi-trait/multi-item analyses were performed to assess internal consistency reliability, and to identify any items that were more highly associated with another domain than its intended domain, or those highly associated with both. (This analysis was based on the MAP – Multi-trait/Multi-Item Analysis Program – by Ware et al. [14] and has been used in previous WHOQOL work [7, 8]).

Discriminant validity

In a preliminary test of discriminant validity, the ability of the domain scores to discriminate between ill and well groups of respondents was tested

Table 1. Frequency responses (%) for items of the WHOQOL-BREF (n = 11,830)

Scale points/domains and facets	1 Poor QOL	2	3	4	5 Good QOL
Total	4.0	12.5	28.0	38.0	17.5
General QOL	2.3	9.1	35.0	43.4	10.1
General health	4.7	18.5	26.4	39.6	10.8
1. Physical health					
Pain and discomfort	3.6	12.9	21.4	26.8	35.2
Energy and fatigue	2.9	11.5	31.4	36.5	17.6
Sleep and rest	4.6	16.8	22.5	38.6	17.5
Dependence on medication[a]	5.9	15.9	18.6	24.4	35.2
Mobility[a]	3.7	10.8	21.9	36.1	27.4
Activities of daily living[a]	2.7	13.5	24.5	44.5	14.8
Working capacity[a]	4.8	15.5	24.6	40.2	14.9
2. Psychological					
Positive feelings	4.9	11.9	35.8	35.1	12.3
Negative feelings	3.2	12.7	25.4	41.7	17.0
Self-esteem	3.0	12.2	28.0	43.7	13.0
Thinking learning, memory and concentration	1.7	11.2	34.0	41.3	11.7
Body image	2.8	9.6	29.3	36.2	22.1
Spirituality, religion and personal beliefs[b]	4.1	10.7	27.9	38.8	18.6
3. Social relationships					
Personal relations	2.4	9.7	23.1	46.3	18.4
Sex	8.8	12.0	32.3	32.9	13.9
Practical social support	2.5	8.4	26.5	44.8	17.8
4. Environment					
Financial resources	8.1	19.1	37.0	25.0	10.9
Information and skills	2.9	11.2	32.1	38.3	15.4
Recreation and leisure	7.3	21.2	31.2	28.2	12.0
Home environment	3.5	9.3	21.0	43.4	22.9
Access to health and social care	2.9	10.2	26.0	45.7	15.2
Physical safety and security	3.8	9.9	32.1	41.9	12.3
Physical environment	3.9	9.6	36.8	38.2	11.5
Transport	4.4	11.3	22.5	42.1	19.8

[a] Independence domain; [b] Spirituality domain in 6-domain model.

by comparing mean scores in the two groups, using *t*-tests. This feature was also assessed by testing the relationship between domains and the two general facet items, using linear regression analysis. All domains were expected to be strongly and positively associated with the concept of overall QOL and health. The impact of gender and age on scores from those who were sick and well (dependent variable) was assessed through a hierarchical multiple regression where these socio-demographic variables were entered together as a block, followed by mean scores for the domains.

Data structure and model fit

Exploratory factor analyses (with Varimax rotation) were conducted to explore the factor structure of the data. Eigenvalues, relative magnitude and direction of factor loadings explaining variance and communality, were examined in these analyses. Confirmatory factor analyses (EQS.5.7b) using structural equation modelling [15] were conducted to obtain objective measures of model fit.

Results

Twenty-four centres contributed a total of 11,830 respondents to the WHOQOL-BREF data set (range 2408 (Germany) to 41 (Netherlands)). The study population consisted of adults aged 12–97 years, with a mean age of 45 (SD = 16) (mean

range 31 (Israel) to 61 (Madras; India)). Only 0.3% were under 16 years of age – in some parts of India, 12 years is considered adult. The recruited numbers in each decade of age were around 20%, from age 20, up to and including the oldest category of 60+. The majority were married or living as married (60%) and 25% were single. Fifty-three percent were women (range 32% Romania to 75% Argentina). Sex and age differences across centres were tested and found to be significant, respectively ($\chi^2 = 230.0$; $p < 0.001$) ($F = 3.6$; $p < 0.001$). More than one-third of respondents (36%) had not received full secondary school education, 40% completed secondary school and 24% had received tertiary education. Forty-seven percent were sick; they were recruited from primary care, in-patient, out-patient, rehabilitation and palliative care settings. Fifty-three percent had no ill health; they were not receiving health care, and in 14 centres were largely recruited from the community. Participants represented 28 identifiable physical and mental health conditions and disorders, drawn from 14 ICD-10 categories. The most prevalent conditions were cancer (17%), diabetes (11%), depression (11%), cardiovascular diseases and hypertension (11%) and musculo-skeletal problems (4%). Although group data on completion times are not available, WHOQOL-BREF can be self-administered in UK by well people, in less than 5 min. About 8% of respondents needed some assistance with completion.

Data quality

As expected from previous studies, responses to each item were distributed across the full range of the scale, with no evidence of ceiling or floor effects for any item for the total data set (Table 1). Among the 24 centres there was some variation and five items had marginally skewed distributions with few responses (<10%) in the two combined categories at the extreme lower end of the 5-point intensity scale (not at all/a little): concentration (cognitions), body image, information, personal relationships and access to health services. For these items, responses tended to group in the centre of the scale. Less than 1% of data were missing for all items, except for sex life (6%) and mobility (1.4%).

Internal consistency reliability

As a measure of the scale's internal consistency, Cronbach's α is reported for the total population and each centre (Table 2). For the total sample, values for Cronbach's α were acceptable (>0.7) for Domains 1, 2 and 4 i.e. physical health 0.82, psychological 0.81, environment 0.80, but marginal for social relationships 0.68. Across sites, results were consistently high with most of the α's in Domains 1 and 2 above 0.75, and in the range of 0.51–0.77 for Domain 3, and 0.65–0.87 for Domain 4. As the Domain 3 α calculation is based on only three items, but on six to eight items in the other domains, lower values in a majority of centres were not unexpected because the statistic is sensitive to numbers. Alpha analyses carried out by systematically removing then replacing each item showed that all 26 items made a significant contribution to the variance in the WHOQOL-BREF.

Discriminant validity

Discriminant validity is an important characteristic of any measure, and in 14 of the 24 centres, members of the general population were selected so that results from well people could be compared with those who were sick, in an initial examination of this property through a test of 'extreme' groups. A comparison of domain scores from sick and well respondents shows that for the majority of countries, discriminant validity was significant for each domain in the total population (Table 3). Discriminant validity was best demonstrated in the physical domain, followed by the psychological, social and environment domains. The results of the hierarchical multiple regression to assess the impact of gender and age on domain scores from sick and well people showed that gender and age together, only explained 2.7% of the overall variance (adjusted R^2), although this effect is significant ($F = 96.3$ (2,7007) $p < 0.0001$).

Construct validity

Domain concepts are more clearly represented and their scores are more easily interpreted when items are clearly identified with the intended domain. Analysis of correlations showed that in the total

Table 2. Internal consistency shown by Cronbach's α's for domains and centres (n = 11,830)

Centre	N	Domain 1: Physical (7 items)	Domain 2: Psychological (6 items)	Domain 3: Social (3 items)	Domain 4: Environment (8 items)
TOTAL		0.82	0.81	0.68	0.80
Argentina	106	0.55	0.79	0.65	0.76
Australia	211	0.73	0.81	0.65	0.72
Brazil	306	0.84	0.78	0.68	0.71
Bulgaria	216	0.83	0.83	0.56	0.74
China	50	0.82	0.89	0.76	0.70
Croatia	406	0.69	0.81	0.61	0.75
Germany	2408	0.88	0.83	0.76	0.78
Greece	48	0.84	0.84	0.71	0.87
Hungary	471	0.82	0.75	0.64	0.76
India, Madras	420	0.77	0.83	0.72	0.82
India, Delhi	1456	0.76	0.80	0.63	0.84
Israel	751	0.72	0.73	0.68	0.69
Italy	379	0.80	0.75	0.65	0.73
Japan	1453	0.71	0.75	0.66	0.75
Malaysia	320	0.81	0.65	0.59	0.74
Netherlands	41	0.79	0.75	0.77	0.85
Nigeria	50	0.85	0.85	0.67	0.83
Norway	1047	0.86	0.83	0.51	0.77
Romania	50	0.87	0.87	0.76	0.65
Russia	300	0.83	0.79	0.74	0.80
Spain	659	0.81	0.80	0.74	0.78
Turkey	48	0.88	0.85	0.58	0.67
UK	475	0.87	0.74	0.55	0.74
USA	159	0.87	0.87	0.69	0.84

Table 3. Discriminant validity: *t*-tests of domain scores for illness vs. well samples

Domain	Physical		Psychological		Social		Environment	
Mean sick	13.1		13.7		14.0		13.8	
Mean well	15.4		14.8		14.8		14.1	
Centre	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*
Total	39.2	<0.01	19.9	<0.01	13.0	<0.01	7.6	<0.01
Brazil	9.5	<0.01	4.7	<0.01	0.8	0.43	2.4	0.02
Bulgaria	10.9	<0.01	7.7	<0.01	6.6	<0.01	2.3	0.02
Croatia	1.8	0.08	6.6	<0.01	7.0	<0.01	1.8	0.07
Germany	33.0	<0.01	18.7	<0.01	9.8	<0.01	9.5	<0.01
Hungary	16.1	<0.01	7.7	<0.01	5.5	<0.01	2.4	0.02
Israel	1.9	0.05	1.3	0.21	2.6	<0.01	1.9	0.06
Italy	8.6	<0.01	2.4	0.02	1.1	0.28	1.2	0.23
Madras	1.8	0.07	1.6	0.18	1.5	0.13	3.0	<0.01
New Delhi	6.1	<0.01	7.1	<0.01	3.6	<0.01	4.7	<0.01
Malaysia	8.7	<0.01	5.0	<0.01	1.7	<0.01	3.4	<0.01
Norway	1.0	0.31	0.1	0.90	0.7	0.49	0.9	0.36
Russia	6.8	<0.01	5.4	<0.01	4.3	<0.01	1.5	0.14
Spain	1.2	0.20	0.1	0.95	1.9	0.06	1.0	0.31
UK	11.3	<0.01	3.5	<0.01	0.5	0.66	0.6	0.54

population, only seven items had strong correlations (>0.50) with domains other than their intended domain. Three of these items on energy, activities of daily living and work were from the

physical domain (Domain 1), and they correlated strongly with the psychological domain (Domain 2). The self-esteem item from Domain 2 was strongly correlated with all of the other domains. The other items were positive feelings, relationships and safety. Specific sites showed more extensive 'cross-domain' correlation, e.g. more than 30 out of 96 per centre in Argentina, Madras, Netherlands, Nigeria, Romania, USA, with most of these items arising from Domains 1 and 2 and most of these correlations occurring between items in these two domains. However in centres where sample sizes were small (as with most in the list above), this scale of cross-domain correlations would be expected, so these analyses should be interpreted with this in mind.

However, no item for the total sample correlated more strongly with another domain than with its own domain, but centre-specific analysis identified two items that occasionally correlated more strongly with domains other than their intended domain. In seven sites, the item on safety was more strongly correlated with the psychological domain than with its intended domain, environment, and in three sites, the energy item correlated more strongly with the psychological than the physical domain.

Because QOL is a complex construct that cannot be directly measured, to establish its construct validity, WHOQOL-BREF domain scores can be compared to general single-item QOL measures with evident face validity. It was predicted that all four domains would show a strong and significant association with overall QOL and health, and so construct validity was partly assessed by correlating the domain scores with each general item (Table 4). The overall assessment of QOL was most strongly associated with the psychological and environment domains, and the overall assessment of health with the physical domain, as predicted. A combined variable representing overall QOL and health through the sum of these two items, showed a strong association with the four domains indicating that each one should be considered when evaluating QOL. All final equation β values were significant.

A review of all the item–total correlations in the total population showed generally good results overall. Poor item–total correlations (<0.30) were only found for negative feelings and in one centre only. In 7 out of 24 centres, items on pain and/or dependence on medication were generally problematic in the physical domain, but no other items were consistently so by this criterion, across sites. Item–domain correlations ranged between 0.48 for pain, to 0.70 for activities of daily living (Domain 1), from 0.50 for negative feelings to 0.65 for spirituality (Domain 2), from 0.45 for sex to 0.57 for personal relationships (Domain 3) and from 0.47 for leisure to 0.56 for financial resources (Domain 4). Summary Pearson correlations (one-tailed test) between domains for the total sample were strong, positive and highly significant ($p < 0.0001$), ranging from 0.46 (physical vs. social) to 0.67 (physical vs. psychological).

Factor analysis

As mentioned earlier, the WHOQOL-100 was based on six theoretical domains that were subsequently reorganised into four domains during the development of the WHOQOL-BREF. Empirical evidence showed that facets from the independence and spirituality domains were associated with the physical and psychological domains respectively, and that a 4-factor solution fits the data better in both ill and well populations. Exploratory factor analyses (Varimax rotation) provided no evidence of a better model. Analysis of the total population data showed four factors (eigenvalues >1.0) that explained 53% of the variance in the data. Centre-specific analyses showed that most sites had four to six eigenvalues greater than 1.0 (mode = 5; range 3–7) and these explained 50–81% of the variance.

Table 4. Validity: association of domains with general facet items (standardized β's) (n = 11,830)

	R^2	Physical	Psychological	Social	Environment
Overall QOL	0.42	0.109	0.290	0.112	0.252
Overall Health	0.41	0.428	0.170	0.070	0.061
Health + QOL	0.52	0.323	0.258	0.102	0.171

Confirmatory factor analyses [14] were run to re-evaluate the fit values found for the original 4-domain model [9]. These were first conducted separately on two random, split-half samples of the data (n = 5133 and n = 5872). The results were acceptable, showing almost identical fit indices for each half and indicating a robust solution, (Table 5 and Figure 1) and supporting similar findings from previous studies [9]. Separate analyses conducted on sick (n = 3313) and well sub-samples (n = 3862) also demonstrated an acceptable fit for this model.

For each of the above analyses, the model for the 6-domain solution was also calculated and Table 5 shows that the fit indices are marginally poorer in each case. However this decrement is not so large as to discount the possibility that a 6-domain model might be used in appropriate contexts where it is justified. Some caution is urged in the interpretation of results in view of the substantial data contributions from five centres.

Table 5. Structural equation modelling fit indices of WHOQOL-BREF models

	4-domain model	6-domain model
Random split half sample A* (n = 5133)		
χ^2	6830.8 df = 249	7624.4 df = 248
CFI	0.863	0.847
RCFI	0.865	0.849
RMSEA	0.07	0.08
Random split half sample B (n = 5118)		
χ^2	6791.0	7132.3
CFI	0.864	0.857
RCFI	0.866	0.859
RMSEA	0.07	0.07
Sick sample (n = 3313)		
χ^2	3736.9	4418.6
CFI	0.876	0.851
RCFI	0.878	0.854
RMSEA	0.07	0.07
Well sample (n = 3862)		
χ^2	4991.3	4995.6
CFI	0.868	0.868
RCFI	0.872	0.871
RMSEA	0.07	0.07

CFI – Comparative Fit Index; RCFI – Robust Comparative Fit Index; RMSEA – Root mean square error approximation.
* Shown in Figure 1.

Preliminary data from the WHOOL-BREF

Mean domain scores for the total sample and for each centre were calculated and found to be relatively similar, with means ranging from 13.5 to 16.2 (SD: 2.6–3.2). Because earlier analyses showed some differences between centres in the age and sex of respondents, mean domain scores adjusted for these factors are presented in Table 6. Using 12.0 as the scale midpoint where QOL is judged to be neither good nor poor, inspection of the means shows that on average, QOL is acceptable to very good physically, psychologically and socially in all centres but is poorest where environmental QOL is considered.

Table 7 provides preliminary comparisons between groups, defined by age and sex. The psychological and social domains showed significant differences indicating that women have better social QOL but poorer psychological QOL than men. This table also shows that mean domain scores decrease with age, and the greatest changes are to be found in physical health.

Discussion

As with its more comprehensive counterpart – the WHOQOL-100 – the results from this field trial of the WHOQOL-BREF are noteworthy because they provide supportive evidence for the cross-cultural validity of this QOL measure. Although not designed to assess each of the 24 specific QOL facets in detail, with only 26 items the WHOQOL-BREF is short enough to be used where time is at a premium, where respondent burden is high or where facet detail is unnecessary. It has wide ranging uses in clinical settings and clinical trials. Although longer than some other short-forms, the WHOQOL-BREF covers a very broad range of facets that were agreed by international consensus. A noteworthy feature is the inclusion of social and environment domains for assessment.

The purpose of this paper was to examine the psychometric properties of the WHOQOL-BREF in terms of item–response distributions, internal consistency reliability, discriminant validity and construct validity. The results showed that the instrument performs well, although some areas deserve further attention. In particular, centre-

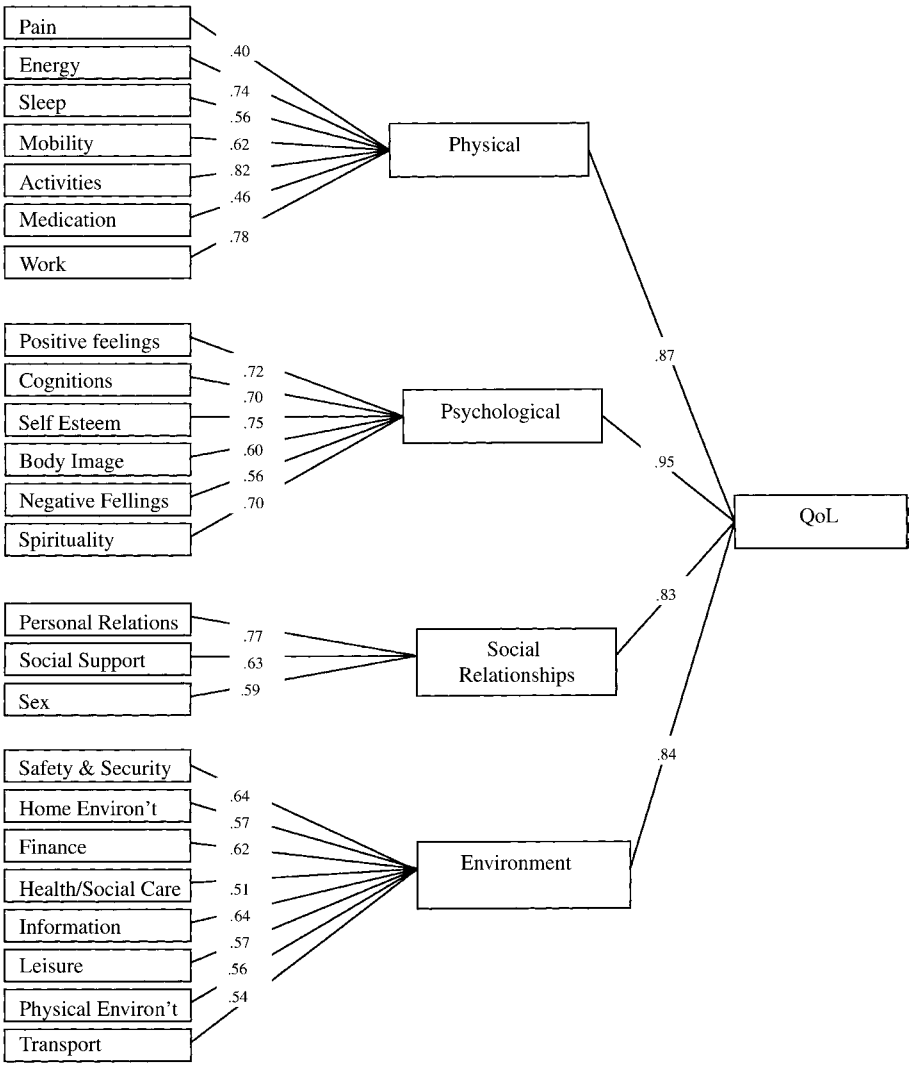

Figure 1. WHOQOL-BREF: 4-domain confirmatory factor model.

specific analyses identified some items that did not discriminate well between domains and two that had stronger correlations with domains other than their intended domain. While the identification of a particular item with its intended domain can be improved by changing the wording and semantics of the translation to reinforce the intended concept, the identification of such items also provides information relevant to construct validity. In some centres, items on safety and energy were more strongly associated with the psychological domain than their intended domains – environment and physical respectively – so in these centres the conceptualization of these issues appears to depart from the theoretical concept. Because no evidence of these associations was found during the development of the WHOQOL-100 or in the item-selection process, it is possible that the WHOQOL-BREF format may have led to a change in the context and thus the conceptualisation of the item by respondents. Further development of the translated instrument (including cognitive debriefing) should focus on this issue, particularly for the safety item.

Previous development work based on the extraction of item data from the WHOQOL-100 field

Table 6. WHOQOL-BREF domain scores (range 4–20) adjusted for age and sex, by center (n = 11,830)

	Physical domain 1		Psychological domain 2		Social domain 3		Environment domain 4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total	16.2	2.9	15.0	2.8	14.3	3.2	13.5	2.6
Argentina	12.1	2.2	10.6	2.9	10.8	3.5	10.7	2.3
Australia	15.7	3.2	15.6	2.7	15.3	3.4	13.1	2.3
Brazil	15.2	2.5	15.1	2.7	14.8	3.1	12.9	2.7
Bulgaria	17.9	2.9	16.5	2.6	15.3	3.0	14.8	2.3
China	15.8	2.9	14.3	2.5	13.7	3.0	13.2	2.4
Croatia	17.1	2.3	15.8	2.1	15.1	3.1	14.3	2.1
Germany	16.8	2.6	15.7	2.4	14.4	2.9	13.0	2.3
Greece	15.2	2.2	12.8	2.6	14.0	2.5	11.9	2.1
Hungary	14.9	2.5	15.4	1.8	14.8	2.1	13.6	1.8
India, Madras	14.8	2.3	15.4	2.2	14.8	2.9	14.8	2.5
India, New Delhi	15.9	2.9	14.2	2.7	13.9	3.7	12.1	2.8
Israel	15.5	3.0	14.2	3.0	13.0	3.8	12.6	2.6
Italy	16.8	3.3	15.0	2.4	15.1	2.6	14.3	2.0
Japan	16.0	2.4	14.2	2.4	12.8	2.5	12.4	2.2
Malaysia	15.6	2.5	13.9	2.8	12.7	3.3	13.5	2.2
Netherlands	18.3	3.0	16.6	2.8	15.8	3.3	15.9	2.8
Nigeria	15.4	3.2	13.9	3.1	13.3	3.6	12.8	2.0
Norway	17.0	3.5	14.7	3.2	13.9	4.7	13.8	3.4
Romania	15.6	2.6	14.2	2.8	13.8	3.1	12.7	2.8
Russia	16.2	3.4	14.8	3.2	14.8	4.0	15.7	3.0
Spain	16.0	3.3	14.8	3.3	13.6	3.4	12.4	3.0
Turkey	15.3	3.4	13.2	2.9	12.3	3.4	13.2	2.0
UK	15.8	3.8	14.7	3.4	14.2	3.5	14.1	2.3
USA	15.5	3.2	13.8	3.2	13.2	3.6	11.7	2.7

trials indicated that while the WHOQOL-100 was based on a theoretical model with six domains, empirical evidence supported a 4-domain model [8] and similar results were found in the development of the WHOQOL-BREF [9], so this was not unexpected given the conceptual similarity of the domains. Analysis of the present WHOQOL-BREF field trial data confirmed that overall, the 4-domain model fits the data well, and also for sick and well respondents separately. But further analysis showed that the 6-domain model was also a good fit – results which are generally consistent with the structure of the WHOQOL-100. Although the more parsimonious 4-domain model is applied, these results provide an empirical basis for extracting more information from the items/facets and scoring of six domains in situations where this is appropriate e.g. where spirituality needs assessment in palliative care.

Some caution is urged in the interpretation of results, given the substantial data contributions from five centres. Other biases from sampling may have arisen because centres did not contribute exactly the same profile of diagnostic groups and well people. Although these analyses provide good support for using the WHOQOL-BREF in QOL assessment, work needs to be carried out to further examine the test–retest reliability of the instrument (but see Ref. [9]), its concurrent validity in comparison to relevant other measures, and within a longitudinal design to examine sensitivity to changes in health states over time.

Conclusion

The WHOQOL-BREF arises from 10 years of development research on QOL and health care. It is a person-centred, multilingual instrument for subjective assessment and is designed for generic use as a multi-dimensional profile, so enabling a wide range of diseases and conditions to be compared. A new methodology has been created within this project whereby international centres

Table 7. Comparison of WHOQOL-BREF mean domain scores by gender and age group (n = 11,830)

	Physical	Psychological	Social	Environment
Gender				
Men				
Mean	14.3	14.2	14.1	13.8
SD	2.9	2.8	3.2	2.7
Women				
Mean	14.2	14.0	14.4	13.9
SD	3.1	2.8	3.1	2.6
F	2.3	17.5	27.1	2.0
p	0.13	0.001	0.001	0.16
Age				
12–20				
Mean	15.6	14.8	14.9	14.4
SD	2.9	2.8	3.1	2.4
21–30				
Mean	15.0	14.3	14.5	13.7
SD	2.9	2.8	3.4	2.6
31–40				
Mean	14.0	13.9	14.0	13.6
SD	3.0	2.8	3.2	2.7
41–50				
Mean	13.9	14.0	14.1	13.9
SD	2.9	2.7	3.1	2.6
51–60				
Mean	13.3	13.8	14.1	14.0
SD	2.9	2.8	2.9	2.6
61+				
Mean	14.2	14.1	14.2	13.8
SD	3.0	2.8	3.2	2.6
F	109.5	18.0	10.6	11.2
p	0.01	0.01	0.01	0.01

worked simultaneously from a common protocol at each stage of the development. International consensus was also obtained at each stage to guide the direction of the research. This replaces the standard serial translation method that is more commonly used in cross-cultural work and this new procedure has reduced some of the problems of obtaining semantic and conceptual equivalence between language versions of the instrument.

The WHOQOL-BREF has several strengths. It is based on a cross-culturally sensitive concept and is available in most of the world's major languages; hence it is appropriate for use in multinational collaborative research. It consists of QOL items that are concerned with the meaning of different aspects of life to the respondents, and how satisfactory or problematic is their experience of them. In addition, the WHOQOL-BREF can generate a profile of four domain scores within a relatively small item set of 26 items. This has implications for its use in research involving a variety of interventions, as well for applications in many service settings. More work on the remaining properties of validity (e.g. concurrent), sensitivity, and feasibility are required. Future research could obtain more comprehensive global survey data (e.g. including Arabia), of more consistent quality, and with structured diagnostic samples of patients. However its conceptual and methodological strengths, combined with the good psychometric properties described in this paper, suggest that WHOQOL-BREF may have a place among the leading generic QOL instruments.

Acknowledgements

The WHOQOL Group acknowledges the assistance of Dr C. Nelson in the early preparation of the manuscript.

Appendix 1

The WHOQOL group.
The WHOQOL Group comprises a co-ordinating group of collaborating investigators in each of the field sites and a panel of consultants. Dr D. Rex Billington directed this project that was initiated by Dr John Orley and Dr Norman Sartorius. The work reported here was carried out in 24 of the field sites: Professor H. Herrman, St. Vincent's Hospital, Australia; Dr S. Bonicatto, FUNDONAR, Argentina; Dr M. Fleck, University of the State of Rio Grande do Sul, Brazil; Dr V. Petkov, National Centre for Interdisciplinary Human Studies, Bulgaria; Professor S. Szabo, Prof Z. Metelko and Mrs M. Pibernik-Okanovic, University of Zagreb, Croatia; Professor M.C. Angermeyer and Dr R. Kilian, Department of Psychiatry, University of Leipzig, Leipzig, Germany; Dr L. Kullmann, National Institute for Medical Rehabilitation, Hungary; Dr S. Kumar, Chennai Medical College, Madras, India; Dr S. Saxena and Dr Rachna Bhargava, All India Institute of Medical Sciences, New Delhi, India; Dr M. Amir, Ben-Gurion University of the Negev, Beer-Sheva, Israel; Dr G. de Girolamo, Instituto Superiore di Sanità, Rome, Italy; Dr M. Tazaki, Science University of Tokyo, Japan; Dr H. Che Ismail, University of Sains Malaysia, Kelantan, Malaysia; Dr M. Kalfoss, Oslo College, Norway; Dr A. Lomachenkov, V.M. Bekhterev Psychoneurological Research Institute, St Petersburg, Russian Federation; Dr R. Lucas Carrasco, Barcelona, Spain; Dr C. Fidaner, Izmir Cancer Registry Center, Turkey; Professor S. Skevington, Ms K. O'Connell and Ms A. Wright, University of Bath, Bath, United Kingdom; Professor D. Patrick, Ms M. Martin and Dr D. Buesching, University of Washington, Seattle, United States. Some data was collected as part of the Longitudinal Investigation of Depression Outcomes (LIDO) study directed by Health Research Associates Inc., and in several field centres of the WHO Disability Assessment Schedule (WHODAS) validation study (ACE Group, WHO, Geneva). Further information on the WHOQOL can be obtained at http://www.who.ch.\msa\mnh\mhp\q1.htm

References

1. Wood-Dauphine S. Assessing quality of life in clinical research: From where have we come and where are we going? J Clin Epidemiol 1999; 52: 355–363.
2. McHorney CA. Health status assessment methods for adults: Past accomplishments and future challenges. Ann Rev Public Health 1999; 20: 309–335.
3. Muldoon MF, Barger SD, Flory JD, Manuck SB. What are quality of life measurements measuring? Br Med J 1998; 316: 542–545.
4. Guyatt GH, Naylor CD, Juniper E, et al. Users' guide to the medical literature XII: How to use articles about health–related quality of life. J Am Med Assoc 1997; 277: 1232–1237.
5. WHOQOL Group. Development of the WHOQOL: Rationale and current status. Int J Mental Health 1994; 23: 24–56.
6. WHOQOL Group. The World Health Organization Quality of Life assessment (WHOQOL): Position paper from the World Health Organization. Soc Sci Med 1995; 41: 1403–1409.
7. WHOQOL Group. The World Health Organization Quality of Life assessment (WHOQOL): Development and general psychometric properties. Soc Sci Med 1998a; 46: 1569–1585.
8. Skevington SM. Measuring quality of life in Britain: Introducing the WHOQOL-100. J Psychosom Res 1999; 47: 5449–5459.
9. WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998b; 28: 551–558.
10. Szabo S, Orley J, Saxena S on behalf of the WHOQOL Group. An approach to response scale development for cross-cultural questionnaires. Eur Psychol 1997; 2: 3270–3276.
11. Sartorius N, Kuyken W. Translation of health status measures. In: Orley J and Kuyken W (eds), Quality of Life Assessment: International Perspectives. Heidelberg: Springer Verlag, 1994.
12. WHO – The World Health Organization. WHOQOL User Manual. Geneva, World Health Organization, 1998.
13. Skevington SM. Advancing cross-cultural research on quality of life: Observations drawn from the WHOQOL development. Qual Life Res 2002; 11: 135–144.
14. Ware JE, Harris WJ, Gandek B, Rogers BW, Reese PR. MAP-R for Windows: Multi-trait/Multi-Item Analysis Program-Revised. Users Guide. Boston, MA: Health Assessment Lab, 1997.
15. Bentler, PM. EQS Structural Equation Program Manual. Encino, C.A: Multivariate Software, Inc, 1995.

Address for correspondence: Suzanne Skevington, WHO Centre for the Study of Quality of Life, Department of Psychology, University of Bath, Bath BA2 7AY, UK
Phone: +44-1225-826830; Fax: +44-1225-826752
E-mail: s.m.skevington@bath.ac.uk