....

European Perspectives

....

The EUROHIS-QOL 8-item index: psychometric results of a cross-cultural field study

Silke Schmidt[1], Holger Mühlan[1], Mick Power[2]

Background: Survey research including multiple health indicators requires brief indices for use in cross-cultural studies, which have, however, rarely been tested in terms of their psychometric quality. Recently, the EUROHIS-QOL 8-item index was developed as an adaptation of the WHOQOL-100 and the WHOQOL-BREF. The aim of the current study was to test the psychometric properties of the EUROHIS-QOL 8-item index. **Methods:** In a survey on 4849 European adults, the EUROHIS-QOL 8-item index was assessed across 10 countries, with equal samples adjusted for selected sociodemographic data. Participants were also investigated with a chronic condition checklist, measures on general health perception, mental health, health-care utilization and social support. **Results:** Findings indicated good internal consistencies across a range of countries, showing acceptable convergent validity with physical and mental health measures, and the measure discriminates well between individuals that report having a longstanding condition and healthy individuals across all countries. Differential item functioning was less frequently observed in those countries that were geographically and culturally closer to the UK, but acceptable across all countries. A universal one-factor structure with a good fit in structural equation modelling analyses (SEM) was identified with, however, limitations in model fit for specific countires. **Conclusions:** The short EUROHIS-QOL 8-item index showed good cross-cultural field study performance and a satisfactory convergent and discriminant validity, and can therefore be recommended for use in public health research. In future studies the measure should also be tested in multinational clinical studies, particularly in order to test its sensitivity.

Keywords: cross-cultural, EUROHIS, health indicator, quality of life, short version

....

In epidemiologic surveys of population health, it is often necessary to include a broad range of very different indicator sets, ranging for instance from chronic condition checklists to various indicators of risk behaviour.[1,2] The term 'indicator' was first defined (outside its physical context) within the 'social indicator movement' in the United States in the 1960s, which has been politically driven to measure societal change processes and was connected to ongoing changes in society and production. Indicators were defined as '. . . a statistic of direct normative interest which facilitates concise, comprehensive, and balanced judgements about the condition of major aspects of a society'.[3] Recently, the meaning of the term 'indicator' has changed, and has been used in a more general way as a derived simple measure of a more complex whole of physical, socioeconomic or subjective variables. International research groups developed and partly succeeded in finding measures for a broad range of health indicators.[4] These efforts on the development of measures were primarily seen as a descriptor of health status, functioning and well-being of populations with and without health conditions to be used in epidemiological studies. However, health indicators can also be used as outcome criteria in clinical intervention studies and as aids for political decision making in the health-care field.

In addition to research on the development of subjective and objective indicators for use in health monitoring, there has been

....

1 Center of Psychosocial Medicine, University Hospital of Hamburg, Germany
2 Section of Clinical and Health Psychology, University of Edinburgh, Edinburgh, UK
Correspondence: Dr Silke Schmidt, University Hospital of Hamburg Eppendorf Institute and Clinic for Medical Psychology, Center of Psychosocial Medicine, Martinistr. 52, Haus S35, D-20246 Hamburg, Germany, tel: +49 (0) 40 42803 6206, fax: +49 (0) 40 42803 4940, e-mail: sischmid@uke.uni-hamburg.de

a recent shift in quality of life (QOL) research to develop short versions of QOL measures that are not only used as monitoring instruments in health surveys but also for screening purposes in clinical studies. The most common example is the development of short measures from the SF-36, e.g. the SF-12.[5,6] General methodological suggestions have been provided; however, short-form development strategies and methods vary with respect to the intended use of the instrument.[7]

The EUROHIS-QOL 8-item index is a QOL measure that has been derived from the WHOQOL project, as an economic screening measure with a particular focus of the short version of the WHOQOL-100, the WHOQOL-BREF.[8–10] In order to develop the measure, multiple datasets were used and multiple strategies were employed.[11] Conceptually, each domain of the original WHOQOL-100 as well as of the WHOQOL-BREF—the psychological, physical, social and environmental domains—is represented in the short from by two items. The development of the EUROHIS-QOL 8-item index for the QOL measure has been based on three large, multinational samples of the WHOQOL-100 and the WHOQOL-BREF ($n > 20\,000$). To extract items for a short version, different methods were employed. First, basic multitrait analyses and analyses of the importance of items within certain domains were carried out. Analyses were based on a data bank of multinational importance ratings of all items of the WHOQOL-OLD. Secondly, Rasch analyses, and confirmatory and exploratory factor analyses were used to derive items that showed the best overall fit for a single factor. The derived version demonstrated good internal consistencies in a pilot study for three countries (UK, France, Germany), but still required psychometric testing in a larger multinational study before it could be employed in cross-cultural surveys. However, in order to be used in multinational projects on population health, epidemiology, cross-cultural and clinical studies, further cross-national psychometric testing is required.

The aim of the current study is to test the psychometric properties of the EUROHIS-QOL 8-item index in terms of its reliability, convergent and discriminant validity, and in terms of its cross-cultural performance, e.g. by conducting analyses on differential item functioning (DIF). The index will also be tested in terms of its Rasch properties, the overall structure of the eight items and QOL.

Materials and methods

Sample and data collection

The study was part of the EUROHIS field study, a European project on health determinants and outcomes for use in health surveys.[12,13] All measures employed have undergone international instrument development, e.g. translation, pilot testing and linguistic validation.

Participants were adminstered a range of measures, which were assessed by telephone or individual interviews. Samples were recruited from 10 countries (UK, France, Germany, Croatia, Czech Republic, Romania, Slovakia, Lithuania, Latvia and Israel; table 1). They either aimed at representativeness, or adjusted according to sociodemographic data: while most samples were based on population registers, the Romanian sample was only based on a specific region. Concerning the mode of administration, all countries used telephone or face to face interviews except for Lithuania, where mail surveys were employed.

Instruments

The EUROHIS-QOL 8-item index—a QOL measure. The EUROHIS-QOL 8-item index is an 8-item measure for QOL, derived from the WHOQOL-100 and the WHOQOL-BREF. The overall QOL score is formed by a simple summation of scores on the eight items, with higher scores indicating better QOL. However, conceptually the psychological, physical, social and environmental domains are each represented by two items. All answer scales have a 5-point response format on a Likert scale, ranging for instance from 'not at all' to 'completely'. Pilot study analyses of the performance of the overall scale showed satisfactory Cronbach alpha values (0.80) for internal consistency in three centres (UK, Germany, France[11]). Except for the health-related QOL items, all items showed significant cross-cultural variation.

General health. As an overall indicator of morbidity, respondents rated their general health ('How is your health in general?') on a 5-point scale that ranged from 'very good' (score 1) to 'very bad' (score 5). It has been shown that this widespread measure acts as an index of overall physical health perception, in contrast to chronic condition checklists, which are considered to be more objective measures, and in contrast to mental and social domains of health.[14–18] On the basis of an overview of 27 studies that included the general health single-item measure, Idler and Benyamini[16] identified only two studies that did not show a prognostic validity for other morbidity indicators.

Chronic physical conditions check list. Respondents were asked to indicate the lifetime incidence of any kind of long-standing chronic condition as well as the 12-month incidence (yes/no) in a chronic condition checklist that included the following diseases: asthma (allergic or otherwise), diabetes, cataract, elevated blood pressure (hypertension), heart attack, stroke, chronic bronchitis, arthritis, osteoporosis, gastric or duodenal ulcer, migraine (frequent headache) and anxiety/depression.[19] A variety of studies have compared chronic morbidity indicators measured through self-report with physical examinations or general practitioner registers.[20–22] While the prevalence of general morbidity does not show discrepancies between self-report and physician's evaluation, the agreement between both perspectives varies considerably as a function of the type of condition. High concordances have been shown for diabetes,

epilepsy, cancer and myocardial infarct.[20,22] Lower concordances were found for hypertension, migraine, muscosceletal diseases and mental disorders. For this reason, the following specifications regarding the chronic condition checklist have been included in the EUROHIS study: (i) whether or not these conditions were diagnosed by a doctor; (ii) whether there was a recent onset of this medical condition; and (iii) whether or not individuals are under medical treatment because of the condition. The association between these variables and the long-standing condition item was high, varying between 0.70 and 0.85.[13,23] Given this concordance, the single item on the lifetime prevalence of a chronic condition was included, also because it showed the lowest amount of missing data across all physical condition variables.

Mental health indicators. In the area of mental health, a general distinction has been made between general indicators enhancing the understanding of population health and indicators that can be used as screening indices for mental disorders.[24] Population-based mental health measures are primarily dimensional indicators, while screening indices are clinical and categorical in estimating the prevalence of mental disorders in the population. The dimensional indicators of mental health in the EUROHIS study comprised the following three measures in the cross cultural analyses.

(i) Psychological distress, as measured by the SF-36 mental health index (MHI5), which has been used to screen for mental health disorders in populations.[25] The MHI5 has been shown to have the highest correlation with the mental health summary score of the SF-36 ($r = 0.87$), and the lowest correlation with the physical health component ($r = 0.17$) of all SF-36 subscales.[26,27]

(ii) Role limitation was measured by the three questions of the Role Emotional Subscale (ROLEM) from the SF-12.[26,28,29]

(iii) Social support was measured by the Oslo Social Support Scale,[30,31] which comprises three items on the primary support group, interest and concern shown by others, and ease of obtaining practical help. Instead of calculating an overall social support scale, the aspects of quantity and quality of social support were entered separately into the overall analyses.

Further measures on risk behaviour, health behaviour and health-care utilization,[32] as well as physical health, were developed in the EUROHIS study, but were not included in the current study.[12]

Data analytic strategy

Multiple statistical procedures were involved in the psychometric analyses of the EUROHIS-QOL 8-item index. Statistical analysis employed basic descriptive psychometric properties, scale properties using classical psychometric theory, specifically internal consistency, validating using other measures, discriminant validity, testing of the scale structure according to structural equation modelling and psychometric analyses using probabilistic measurement theory (unidimensionality). Furthermore, item bias was explored using different methods for the detection of DIF, a logistic regression approach and an ANOVA significance test of the residuals.

Descriptive and basic statistical analyses of the data were performed using SPSS 11.0 computer software. Different software packages for investigating structural equation modelling (AMOS,[33] EQS[34]) and models of item response theory (RUMM,[35] WINMIRA[36]) were employed.

Results

Sample characteristics

The total number of respondents from the 10 countries was 4849, with 1203 individuals from the UK, France and Germany, a combined total of 1876 from Croatia, the Czech Republic,

Table 1 Sample characteristics (n = 4849)

Country group	(I) Western European			(II) Baltic states		(III) Eastern European				(IV) Israel	Total
Country n (%)	France 400 (8.2)	Germany 401 (8.3)	UK 402 (8.3)	Lithuania 455 (9.4)	Latvia 323 (6.7)	Croatia 361 (7.4)	Romania 602 (12.4)	Slovakia 401 (8.3)	Czech Republic 512 (10.6)	Israel 992 (20.5)	4849 (100.00)
Gender											
Female	242 (60.4)	217 (54.0)	218 (54.2)	245 (53.9)	234 (72.4)	194 (53.9)	317 (52.7)	196 (48.9)	269 (52.6)	618 (62.4)	2.750 (56.7)
Male	158 (39.6)	184 (46.0)	184 (45.8)	210 (46.1)	89 (27.6)	167 (46.1)	285 (47.3)	205 (51.1)	243 (47.7)	374 (37.6)	2.099 (43.3)
Education											
0–11 years	184 (46.8)	190 (47.5)	179 (44.6)	159 (34.9)	80 (24.8)	95 (26.5)	240 (39.9)	128 (34.0)	139 (27.0)	263 (26.8)	1.657 (34.5)
12 years	59 (15.3)	54 (13.4)	66 (16.6)	72 (16.4)	52 (16.1)	138 (38.2)	133 (22.1)	90 (24.0)	158 (31.0)	314 (32.3)	1.136 (23.7)
>12 years[a]	151 (37.9)	157 (39.1)	157 (38.8)	216 (48.7)	190 (59.0)	52 (35.4)	228 (37.9)	164 (42.0)	214 (42.0)	403 (40.9)	2.007 (41.8)
Age group											
36–45 years	50 (12.5)	50 (12.5)	51 (12.7)	278 (61.1)	213 (65.9)	126 (34.9)	280 (46.5)	262 (65.3)	288 (56.3)	571 (57.6)	2.169 (44.7)
46–55 years	100 (25.0)	100 (24.9)	102 (25.4)	96 (21.1)	45 (13.9)	74 (20.5)	125 (20.8)	85 (21.2)	97 (18.9)	177 (17.8)	1.001 (20.6)
56–65 years	200 (50.0)	201 (50.1)	198 (49.3)	80 (17.6)	37 (11.5)	73 (20.2)	96 (15.9)	53 (13.2)	54 (10.5)	99 (10.0)	1.091 (22.5)
66–75 years	50 (12.5)	50 (12.5)	51 (12.7)	1 (0.2)	28 (8.7)	88 (24.4)	101 (20.8)	1 (0.2)	73 (14.3)	145 (14.6)	588 (12.1)
Marital status											
Married or living as	315 (79.3)	306 (76.3)	263 (66.1)	315 (69.7)	138 (42.7)	230 (63.6)	432 (71.8)	264 (64.7)	301 (58.7)	670 (67.7)	3.234 (66.8)
Separate	8 (2.0)	10 (2.5)	9 (2.3)	9 (2.0)	14 (4.3)	2 (0.6)	4 (0.7)	92 (23.4)	11 (2.2)	7 (0.7)	166 (3.4)
Divorced	24 (5.6)	40 (10.1)	52 (13.2)	44 (9.4)	26 (8.0)	27 (7.5)	26 (4.3)	26 (6.8)	44 (8.6)	45 (4.6)	354 (7.3)
Widowed	30 (7.7)	28 (7.1)	42 (10.4)	24 (5.4)	25 (7.7)	40 (11.1)	69 (11.5)	16 (4.2)	47 (9.2)	69 (7.0)	390 (8.1)
Never married	23 (5.4)	17 (4.0)	33 (8.1)	61 (13.5)	120 (37.2)	62 (17.2)	71 (11.8)	3 (0.8)	109 (21.3)	196 (20.1)	695 (14.4)
Employment											
Employed or self-employed	136 (34.0)	167 (41.6)	180 (44.8)	264 (58.0)	160 (49.5)	151 (41.8)	305 (50.7)	315 (78.6)	300 (58.6)	416 (41.9)	2.394 (49.4)
Unemployed	14 (3.5)	17 (4.2)	20 (5.0)	71 (15.6)	17 (5.3)	28 (7.8)	26 (4.3)	35 (8.7)	19 (3.7)	81 (8.2)	328 (6.8)
Student	1 (0.3)	2 (0.5)	4 (1.0)	8 (1.8)	74 (22.9)	18 (5.0)	14 (2.3)	17 (4.2)	47 (9.2)	27 (2.7)	212 (4.4)
Retired	188 (47.0)	178 (44.4)	158 (39.3)	50 (11.0)	46 (14.2)	149 (41.3)	193 (32.1)	21 (5.2)	118 (23.0)	43 (4.3)	1.144 (23.6)
Housewife/houseman	51 (12.8)	36 (9.0)	33 (8.2)	34 (7.5)	26 (8.0)	12 (3.3)	60 (10.0)	–	18 (3.5)	166 (16.7)	436 (9.0)
Other	10 (2.5)	1 (0.2)	7 (1.7)	24 (6.2)	–	3 (0.8)	4 (0.7)	26 (6.4)	20 (2.0)	159 (26.1)	335 (7.0)

a: Three categories were employed to differentiate between different levels of educational degrees and contained different examples for different languages (see Nosikov and Gudex[12])

Romania and Slovakia, 778 from the Baltic States Lithuania and Latvia, and 992 from Israel (table 1).

The number of missing values varied between different health indicators; in addition, non-response rates were lowest in the psychometrically developed indicator sets, i.e. the QOL and mental health indicator data. The highest rate of missing data was observed in the use of preventive care indicator (non-response rate of 5.8%). Missing data varied considerably across countries, particularly in the physical activity and alcohol consumption indicator set. With respect to sociodemographic data, the respondents in the western European countries, being based on representative data (western European countries and Israel), displayed a higher age range ($F = 105.6$, 3, 4831; $p = 0.06$), showed fewer years of schooling ($\chi^2 = 207$, 5, 3; $P < 0.001$) and higher rates of employment ($\chi^2 = 111.9$, 3; $P < 0.001$). The rates of female participants were higher in the two Baltic states ($\chi^2 = 35.1$, 6; $P < 0.001$), particularly in the Latvian sample. Participants were more frequently married in the western and eastern European groups than in the Baltic countries and Israel ($\chi^2 = 51.7$, 3; $P < 0.001$) which was, however, associated with the lower age range in these countries. Basic sociodemographic properties (for gender, cohort, education, employment, marital status) of the whole sample and of country-specific subsamples are listed in table 1.

Cross-cultural applicability of items and scales

Descriptive country differences in the EUROHIS-QOL 8-item index. Overall, the rate of missing data for the eight items of the EUROHIS-QOL 8-item index was very low (below 1%; table 2). The scale showed a good internal consistency ($\alpha = 0.83$) and low to moderate floor and ceiling effects. Table 3 displays means and standard deviations of the total EUROHIS-QOL 8-item index, as well as the eight items, across the 10 countries. Country-specific analyses suggest considerable country differences on a descriptive level. This also holds for EUROHIS-QOL overall score means adjusted for core sociodemographic features, such as gender, age group, marital status, etc. (figure 1). For statistical analyses, four country groups were formed: the western countries (UK, France and Germany), the eastern countries (Czech Republic, Slovakia, Romania, Croatia), the Baltic states (Lithuania, Latvia) and Israel. There is a higher

QOL in the western European states and Israel and lower scores in southern and eastern European states, in particular for the Baltic states. Employing ANOVA, country effects were significant, with post-hoc Scheffé tests highlighting significant differences between Lithuania and Latvia on one hand, and the western European countries as well as Israel on the other. As hypothesized, the clear gap between eastern and western European countries allowed for grouping of these countries in mulitfactorial analyses. Gender, age and country effects of a combined univariate analysis of the EUROHIS-QOL 8-item index total score as well as the EUROHIS-QOL 8-item index subdimensions highlighted the same range of effects with participants from eastern and Baltic European countries (Latvia, Lithuania, Romania, Czech Republic, Slovakia), as well as females and younger persons reporting lower QOL on a single item level.

However, there were some interesting age effects on the level of single items. Country, age and gender effects of the EUROHIS-QOL 8-item index are displayed in table 3. In general, country effects were the largest effects on a descriptive level (partial $Eta^2 = 0.2$).

Convergent and discriminant validity

Convergent validity: intercorrelations between EUROHIS-QOL 8-item index and different measures for mental and physical health. In the total sample, the zero-order correlation between the EUROHIS-QOL 8-item index for QOL and the mental health index (MHI5) was $r = 0.49$, between QOL and the general health variable $r = 0.53$ and between QOL and social support (OSLO measure) $r = 0.36$. Comparing the interrelationship between these three measures across the four country groups the correlations between the QOL and general health showed correlations higher than $r = 0.50$ for all of these countries. In the Baltic states and southern/eastern European countries the correlation between the QOL and the MHI5 was $r = 0.40$ and $r = 0.39$, respectively.

Discriminant validity between healthy and ill people

Testing the discriminant validity of the measure, the EUROHIS-QOL index items were analysed with respect to whether they

Table 2 Selected descriptive properties for the eight items and total score of the EUROHIS-QOL 8-item index ($n = 4849$)

|	EUROHIS-QOL 8-item index	MD (%)[a]	Mean	SD	Floor (20%)	Ceiling (20%)	Skewness	α (*if del.)	$r_{item\ total}$[b]
1	How would you rate your quality of life	0.82	3.68	0.82	1.43	13.43	−0.55	0.80*	0.65
2	How satisfied are you with your health	0.84	3.60	0.96	2.95	13.69	−0.70	0.81*	0.59
3	Do you have enough energy for everyday life	0.74	3.80	0.91	1.83	21.63	−0.68	0.81*	0.59
4	How satisfied are you with your ability to perform your daily activities	0.72	3.79	0.87	1.85	16.87	−0.92	0.80*	0.61
5	How satisfied are you with yourself	0.99	3.66	0.88	1.77	14.41	−0.61	0.81*	0.60
6	How satisfied are you with your personal relationships	0.78	3.89	0.84	1.37	20.87	−0.91	0.82*	0.51
7	Have you enough money to meet your needs	0.80	3.14	1.09	9.09	9.54	−0.27	0.82*	0.49
8	How satisfied are you with the conditions of your living place	0.78	3.89	0.97	2.24	27.46	−0.89	0.83*	0.45
QOL	Total score	1.77	3.68	0.62	0.61	19.76	−0.54	0.83	−

a: Missing data listwise
b: corrected for overlap

Table 3 Selected descriptive and psychometric properties for the total score (range 1–5) of the EUROHIS-QOL 8-item index, by country ($n = 4849$)

	Country	n (%)	α	MD[a]	Total		Female		Male		Younger		Older	
					Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	France	400	0.74	2.25	3.84	0.46	3.80	0.50	3.92	0.40	3.90	0.53	3.83	0.45
2	Germany	401	0.80	1.25	4.08	0.49	4.03	0.52	4.14	0.44	4.07	0.54	4.08	0.48
3	UK	402	0.80	1.24	4.00	0.62	3.99	0.62	4.01	0.62	3.85	0.60	4.02	0.62
4	Lithuania	455	0.85	1.76	3.19	0.63	3.15	0.62	3.28	0.63	3.28	0.63	3.05	0.60
5	Latvia	323	0.74	0.00	3.44	0.48	3.42	0.50	3.49	0.42	3.51	0.47	3.28	0.46
6	Croatia	361	0.78	0.28	3.72	0.52	3.71	0.48	3.73	0.56	3.85	0.51	3.63	0.51
7	Romania	602	0.81	0.00	3.47	0.56	3.38	0.57	3.56	0.52	3.57	0.51	3.35	0.58
8	Slovakia	400	0.77	6.70	3.56	0.52	3.54	0.54	3.58	0.50	3.59	0.54	3.50	0.48
9	Czech Republic	512	0.84	0.02	3.57	0.62	3.53	0.59	3.61	0.65	3.66	0.62	3.44	0.62
10	Israel	992	0.82	1.01	3.86	0.62	3.84	0.64	3.91	0.60	4.00	0.56	3.66	0.66
	Total	4.849	0.83	1.77	3.68	0.62	3.65	0.63	3.73	0.61	3.70	0.61	3.67	0.64

a: Missing data listwise
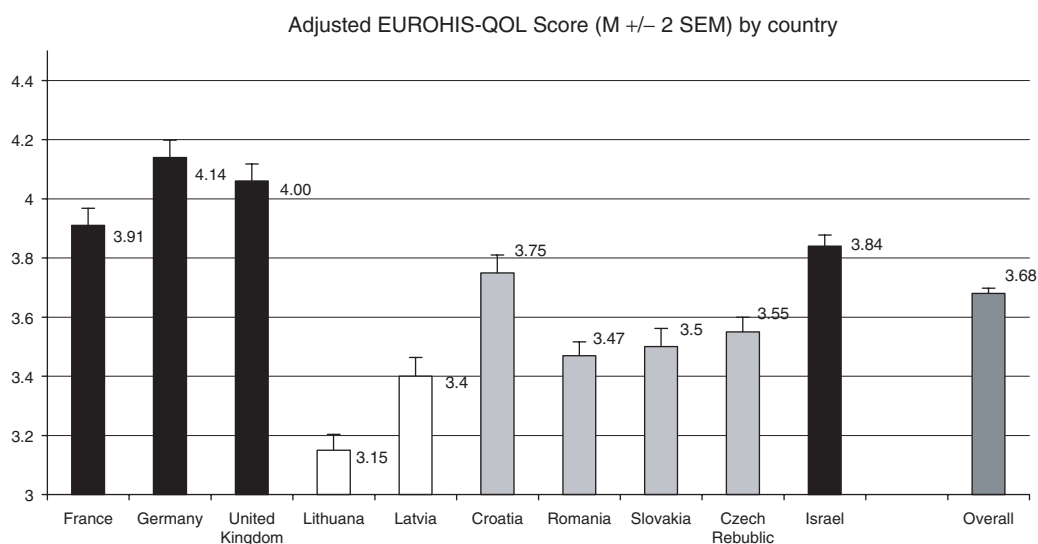
Figure 1 Adjusted means (M) and two standard errors of the mean (SEM) for total score EUROHIS-QOL 8-item index for different countries. Scores were adjusted for gender, age and health status

distinguish between healthy and ill populations (table 4). The variable discriminating between healthy and ill populations was derived from the chronic conditions list including the longstanding condition indicator showed a high convergence with recent diagnosis ($\chi^2 = 999.41$, 1; $P < 0.001$) and being under treatment ($\chi^2 = 154.46$, 1; $P < 0.001$).

Testing the capacity of the EUROHIS-QOL 8-item index in its performance to discriminate between healthy and ill populations, measured dichotomously, a significant discriminative potential for the overall score can be shown across all countries except for the Israel ($P = 0.090$) and Slovakian ($P = 0.111$) subsamples.

Analysis of DIF in the EUROHIS-QOL 8-item index

Applying Rasch analysis in order to estimate the unidimensionality of the measure (table 5), the scale showed very good item fit statistics, and no reversed thresholds. In terms of the residuals, item 7 ('enough money to meet your needs') and item 8 ('satisfied with living place') showed slightly problemtaic scores.

Table 6 shows a summary of the DIF analyses in the country (western vs. eastern), gender, age (younger vs. older by center split) and chronic condition (yes/no) subgroups. Differential Item functioning (= DIF) means that an item performs and measures differently for one subgroup of a population than for another. According to this definition of DIF, there should be no association between the item and country, respectively any other subgroups. Table 6 shows that, on an overall level, only in two items DIF was identified in respect to country identified ('satisfaction with yourself', 'having enough money to meet your needs'), in one item in respect to age ('satisfied with your living place'), and no DIF in gender and condition groups according to the criterion proposed by Bjorner *et al.*[37] This 'conservative' criterion implies that pseudo-R^2-differences are considered to be significant when they are higher than a cut-off point of at least
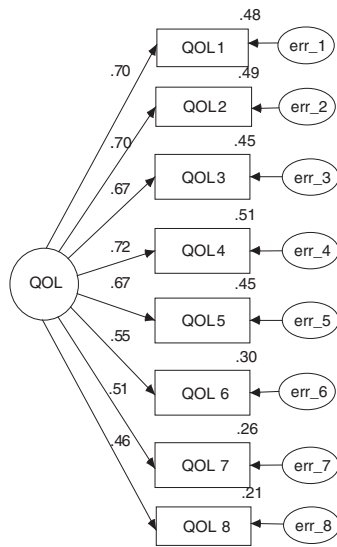
Country	n	Chi square / df (p)	CFI	RMSEA
France	400	46.240 / 20 (.001)	.95	.06
Germany	401	73.321 / 20 (.000)	.94	.08
United Kingdom	402	86.364 / 20 (.000)	.92	.09
Lithuana	455	96.207 / 20 (.000)	.94	.09
Latvia	323	62.056 / 20 (.000)	.90	.08
Croatia	361	97.108 / 20 (.000)	.89	.10
Romania	602	356.040 / 20 (.000)	.78	.17
Slovakia	401	63.790 / 20 (.000)	.92	.08
Czech Republic	512	151.809 / 20 (.000)	.92	.11
Israel	992	287.255 / 20 (.000)	.89	.12
Total	4849	1082.173 / 20 (.000)	.91	.10

Path diagram (QOL latent variable with items QOL1–QOL8):

- QOL1 ← err_1 (.48), loading .70
- QOL2 ← err_2 (.49), loading .70
- QOL3 ← err_3 (.45), loading .67
- QOL4 ← err_4 (.51), loading .72
- QOL5 ← err_5 (.45), loading .67
- QOL6 ← err_6 (.30), loading .55
- QOL7 ← err_7 (.26), loading .51
- QOL8 ← err_8 (.21), loading .46

Figure 2 Selective results (χ^2 statistics, CFI, RMSEA) of a confirmatory factor analyses (CFA) for the eight items of the EUROHIS-QOL index with one latent variable (QOL), overall and for each country ($n = 4849$). CFI = comparative fit index; RMSEA = root mean square estimation approximation

Table 4 Discriminant validity (*t*-tests) of the total score of the EUROHIS-QOL 8-item index ($n = 4849$) for ill versus healthy sample, by country and total sample (adjusted for α)

	Country	Ill sample			Healthy sample			Total	
		Mean	SD	n	Mean	SD	n	t	P
1	France	3.67	0.48	130	3.93	0.43	261	5.58	0.000
2	Germany	3.93	0.49	182	4.21	0.45	214	5.88	0.000
3	UK	3.82	0.66	205	4.19	0.51	192	6.33	0.000
4	Lithuana	2.98	0.61	184	3.34	0.59	263	6.30	0.000
5	Latvia	3.25	0.50	140	3.58	0.40	183	6.54	0.000
6	Croatia	3.61	0.51	266	4.04	0.41	94	7.39	0.000
7	Romania	3.34	0.55	390	3.71	0.488	212	8.19	0.000
8	Slovakia	3.47	0.55	76	3.58	0.51	278	1.59	0.111
9	Czech Republic	3.43	0.64	332	3.83	0.50	179	7.15	0.000
10	Israel	3.94	0.62	154	3.85	0.62	828	−1.69	0.090
	Total	3.52	0.63	2.059	3.81	0.58	2.704	16.27	0.000

The category 'Ill' was defined by a lifetime, longstanding condition

2 % (using Nagelkerke R^2). Of course, this score might vary also with respect to the sample size and is not reported consistently by different authors. A second approach to DIF detection, the ANOVA significance test of the residuals, was employed using RUMM software. Using this approach, more cases of DIF for the country factor were identified. DIF did not occur in respect to gender in the RUMM approach. Almost no DIF occurred in respect to having a health condition with the exception of the EUROHIS health item, which makes sense from a theoretical point of view in chronic diseases. Comparing the different sources of DIF, DIF rather occurred in respect to country and also in respect to age than in respect to condition and gender.

On a descriptive level, we found slight indication of country DIF in different subscales in two items ('satisfaction with yourself', 'having enough money to meet your needs'). DIF did not occur with respect to gender; however, there was slight DIF with respect to the age group factor on the item 8 referring to 'satisfaction with one's own living place'. Almost no DIF occurred in condition with the exception of the health item. This makes sense from a theoretical point of view in chronic conditions. Comparing the different sources of DIF, DIF rather occurred in respect to country and also with respect to age than in respect to condition and gender.

Another, more profound analysis of DIF is to test each translation in comparison to the English draft items of the UK

Table 5 Itemfit and parameter estimation for the EUROHIS-QOL 8-item index (using WINMIRA and RUMM)

	EUROHIS-QOL 8-item index	Q	Zq	P	Location	Residuals	χ^2	Reverse thresholds[a] (RUMM)	Reverse thresholds (WINMIRA)
1	How would you rate your quality of life	0.15	−0.09	0.54	−0.08	−6.14	4.13	–	–
2	How satisfied are you with your health	0.17	−0.23	0.59	0.19	−1.28	33.38	–	–
3	Do you have enough energy for everyday life	0.15	−0.05	0.52	−0.21	−0.94	53.41	–	–
4	How satisfied are you with your ability to perform your daily activities	0.13	−0.38	0.65	−0.14	−4.01	53.13	–	–
5	How satisfied are you with yourself	0.14	0.06	0.48	−0.03	−2.31	11.77	–	–
6	How satisfied are you with your personal relationships	0.13	−0.25	0.60	−0.35	1.51	160.67	–	–
7	Have you enough money to meet your needs	0.13	0.50	0.31	0.87	8.80	1003.19	–	–
8	How satisfied are you with the conditions of your living place	0.13	0.47	0.32	−0.25	8.82	165.87	–	–

a: Reverse thresholds indicate that response scales do not fit Rasch properties
Q = Q index: respresents an unstandardized parameter for an infit of undimensionality according to the Rasch model
Higher scores indicate higher deviations. Zq is a standardized paramter of Q with a threshold of 0.30 suggested to indicate violations of the Rasch model

Table 6 Differential item functioning analyses of the eight items of the EUROHIS-QOL-measure ($n = 4849$)

	EUROHIS-QOL 8-item index	Country[a]		Gender		Age group		Condition	
		R^2-diff	RUMM	R^2-diff	RUMM	R^2-diff	RUMM	R^2-diff	RUMM
1	How would you rate your quality of life	0.005	0.000	0.003	0.121	0.002	0.592	0.001	0.078
2	How satisfied are you with your health	0.012	0.000	0.001	0.186	0.017	0.000	0.020	0.034
3	Do you have enough energy for everyday life	0.008	0.076	0.003	0.954	0.004	0.065	0.003	0.288
4	How satisfied are you with your ability to perform your daily activities	0.016	0.000	0.002	0.272	0.009	0.853	0.003	0.234
5	How satisfied are you with yourself	0.027	0.000	0.001	0.647	0.006	0.039	0.003	0.596
6	How satisfied are you with your personal relationships	0.010	0.072	0.003	0.335	0.003	0.048	0.006	0.082
7	Have you enough money to meet your needs	0.027	0.000	0.001	0.013	0.015	0.005	0.006	0.102
8	How satisfied are you with the conditions of your living place	0.014	0.000	0.004	0.919	0.027	0.977	0.010	0.078

a: Tables for DIF analyses of individual countries against UK draft version can be obtained from the authors. DIF analyses were performed usiong a logistic regression approach (Zumbo[40]) with pseudo-R^2 differences (R^2-diff) for detecting the DIF effect size as well as the RUMM approach using ANOVA residuals

version. When comparing cross-national DIF, one should use populations that are comparable with respect to other subgroups. Possible sex, age and severity biases were controlled within the sampling strategy. Comparing each item in each language with the English original, the following observations can be made. The smallest amount of DIF occurred comparing the Israel with the English data except for item 8 ('money to meet your needs'). This suggests that DIF does not occur as a result of a country effect, but as a translation effect because Israel employed the English draft version. French and Germany, being closer to the British culture, showed fewer DIF than countries from the eastern states and the Baltic states. All countries showed some substantial DIF in comparison with the UK draft version of the EUROHIS-QOL 8-item index; however, nearly the same items displayed DIF. Again, this would suggest translation difficulties. The items with most significant DIF were related to the ability to perform daily living activities, the living conditions of your living place and the amount of energy for everyday life. Some DIF was also observed for gender and age.

Testing the universal applicability and structural validity of the EUROHIS-QOL 8-item index

Testing the universality of the QOL one-factor model of the EUROHIS-QOL 8-item index, a confirmatory factor analysis using structural equation modelling was performed (figure 2). The maximum-likelihood method was selected for parameter estimation. The analyses were conducted across all countries, as well as in each country sample. The model fitted the data well (CFI = 0.91, RMSEA = 0.10) with a satisfactory contribution of the latent factor on each item. The fit in each country was satisfactory, with a better fit in the UK, France and Germany and a lower fit in Slovakia and Israel. With the Romanian data, there was a poor fit for the one factor model (CFI = 0.78, RMSEA = 0.17), both as compared with other country samples and also in

terms of cut-off criteria for the fit of structural equation models.[38,39] However, the fit for the Romanian data was improved by allowing error covariance between item 8 ('living place') and item 7 ('money') to covary with the overall QOL factor (revised model CFI = 0.910, χ^2 = 156.7, RMSEA = 0.053). This improvement in fit demonstrates that the one-factor model fits the Romanian data, but it also demonstrates the importance of money and living place for the general rating of QOL.

Discussion

Recently there has been a trend of constructing and using short versions of QOL measures as an economic tool.[7] For instance, cross-national survey research that includes a range of indicators requires short indices that are easy to complete. Because these measures might be developed at the expense of content validity, there is an urgent need to test the performance of these measures on both a national and an international level. The current study tested the performance of a recently developed short version of the WHOQOL-100[11] in a sample of 4849 adults across 10 European countries. The EUROHIS-QOL 8-item index showed good qualities in terms of internal consistency across all countries, good and equal discriminant power to distinguish between conditions as well as associations with health status in all countries, and a good internal structure, for example, from the loading of all items on the overall QOL factor in each country, with the initial exception of the Romanian sample. The model fitted worse in this sample, both as compared with other country samples and also in terms of cut-off criteria for the fit of structural equation models.[38,39]

In the Romanian data, two items showed a very high impact on the overall QOL factor, which led to an improvement of the structure when covariance terms were included. This finding could result from different mechanisms: it could be assumed that the lower fit results from the fact that there is no direct translation of the term 'quality of life' in Romanian, which has been evident in focus groups (G. Hawthorne, N. Davidson, K. Quinn and THE WHOQOL Group (40 people), manuscript in preparation). Therefore, the two terms between money and living place may have a pervasive impact on overall QOL, so that they reflect more closely the QOL concept in Romania than the other items of the EUROHIS-QOL index. Conceptually, this finding may also illustrate a link between objective and subjective measures of QOL for countries where poverty, for example, is a current concern for the respondent, or may reflect effects on respondents that result from the transition from one political system to another.

With respect to the DIF analyses, two different methods were used to detect DIF: a logistic regression approach[40] and a significance test of the residuals.[35] The size of pseudo-R^2 differences in the logistic regression approach serves as an indicator for the effect-size of DIF. According to the Zumbo,[40] cut-off scores that can be interpreted as meaningful pseudo-R^2 differences, ranged from 0.07 to 0.13. Rumm software was used to perform additional DIF analyses, computing the *P*-value of the ANOVA significance test of the residuals. Both methods were used to identify DIF for country, gender, age and health status. These analyses revealed that, overall, there was only moderate DIF. The *P*-values of the ANOVA significance test of the residuals showed that DIF occurred more often between country groups, suggesting item bias possibly caused by translation or cultural differences. To proceed with a more advanced analysis of DIF for translation versus culturally caused item bias, the logistic regression approach was applied at the cross-national level. That is, each translation was tested for DIF in comparison with the English draft items of the UK version. From these analyses a tendency could be observed, particularly for different country groups, with the closer the culture to the western 'draft language culture', the less the DIF was observed. That is, both the frequency of items affected by DIF as well as the amount of DIF for a single item, as indicated by the pseudo-R^2 differences, vary with respect to cultural setting, related to a geographical shift. To date, analysing DIF in QOL research has mostly been used in order to identify translation mistakes as one special case of culturally originated item bias.[41,42] A different understanding of DIF is suggested when interpreting DIF as an effect of culture. However, this cannot finally answer the question whether either translation difficulties or cultural differences cause these differences.

To summarize, the current study was able to demonstrate that the EUROHIS-QOL index functions well as an overall score assessing QOL across a range of countries. However, it does not provide consistent evidence for the validity of two items representing the original WHOQOL domains. These items are, from a conceptual point of view, more distant to the construct of health-related QOL. The highest association of general health perception occurred with the EUROHIS general health item, as well as with the total QOL score, which supports the clinical validity of the measure.

A serious limitation of the study is related to the fact that the samples were comparatively small and that they varied between countries so that representativeness cannot be assumed. For instance, in the group of older people (age group 66–75 years) some countries (e.g. Lithuania) had very small sample sizes. The fact that the prevalence of a long-standing conditions was so small in the Israel data and that there was no QOL difference between healthy and ill samples might be attributed to these circumstances.

Future studies will have to test the index as well as its parent measure, the WHOQOL-BREF, jointly in representative samples in order to demonstrate whether the index is able to replicate the overall score of the original. Nevertheless, the current study has demonstrated that the index works well as an overall QOL factor across a range of countries and that it can now be used as a short and easy-to-complete index in health surveys in those countries that have undergone cross-cultural validation.

Acknowledgements

Key points

- The cross-cultural performance of a recently developed quality of life (QOL) instrument was tested employing a comprehensive empirical methodology.
- Measures of classical test theory showed a a range of indices for a good reliabilty and validity across a range of countries.
- Differential item functioning (DIF) analyses revealed in general moderate DIF across countries; however, the closer the culture to the western 'draft language culture', the less the DIF that was observed.
- The EUROHIS-QOL 8-item index showed a good internal structure, for example, from the loading of all items on the overall QOL factor in each country, with the initial exception of the one country sample, presumably owing to a divergent cultural concept of QOL.
- The index should now be used in multinational clinical and public health studies in convergence with other measures.

References

1 Young TK. *Population Health—Concepts and Methods*. New York: Oxford University Press, 1998.

2 Bullinger M. International comparability of health interview surveys: An overview of methods and approaches. In Nosikov A, Gudex C, editors. *Developing Common Instruments for Health Surveys*. Amsterdam: IOS Press, 2003:1–11.

3 Andrews FM, Whitey SB. *Social Indicators of Well-being. Americans' Perceptions of Life Quality*. New York: Plenum Press, 1976.

4 Doyal L, Gough I. *A Theory of Human Need*. London: MacMillan, 1991.

5 Jenkinson C, Chandola T, Coulter A, et al. An assessment of the construct validity of the SF-12 summary scores across ethnic groups. *J Public Health Med* 2001;23:187–94.

6 Ware JE. The SF-36 health survey. In Spilker B, editor. *Quality of Life and Phramacoeconomics in Clinical Trials*. 2nd edition. Hagerstown, MD: Lippincott-Raven, 1996.

7 Smith GT, McCarthy DM, Anderson KG. On the sins of short-form development. *Psychol Assess* 2000;12:102–11.

8 WHOQOL Group. The World Health Organization Quality of Life Assessment (WHOQOL): Development and General Psychometric Properties. *Soc Sci Med* 1998;46:1569–85.

9 WHOQOL Group. Development of The World Health Organization WHOQOL-BREF Quality of Life Assessment. *Psychol Med* 1998;28:551–8.

10 Skevington SM, Lotfy M, O'Connell KA. 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. *Qual Life Res* 2004;13:299–310.

11 Power MJ. Development of a common instrument for quality of life. In Nosikov A, Gudex C, editors. *Developing Common Instruments for Health Surveys*. Amsterdam: IOS Press, 2003:145–63.

12 Nosikov A, Gudex C. Overview of the EUROHIS project. In Nosikov A, Gudex C, editors. *Developing Common Instruments for Health Surveys*. Amsterdam: IOS Press, 2003:13–20.

13 Schmidt S, Power M, Bullinger M. Cross-cultural analysis of relationships of health indicators across Europe: First results based on the Eurohis project. 2002. World Health Organisation—Regional Office for Europe. EUR/02/5041391. http://www.euro.who.int/document/e77964.pdf. Accessed 11 August 2005.

14 Bailis DS, Segall A, Chipperfield JG. Two views of self-rated general health status. *Soc Sci Med* 2003;56:203–17.

15 Bailis DS, Segall A, Mahon MJ, et al. Perceived control in relation to socioeconomic and behavioral resources for health. *Soc Sci Med* 2001;52:1661–76.

16 Idler EL, Benyamini Y. Self-rated health and mortality: A review of twenty-seven community studies. *J Health Soc Behav* 1997;38:21–37.

17 Liang J, Bennett J, Whitelaw N, et al. The structure of self-reported physical health among the aged in the United States and Japan. *Med Care* 1991;29:1161–80.

18 Simon GE, Revicki DA, Grothaus L, et al. SF-36 summary score—Are physical and mental health truly distinct? *Med Care* 1998;36:567–72.

19 Buratta V, Frova L, Garguilo EG, et al. Development of a common instrument for chronic physical conditions. In Nosikov A, Gudex C, editors. *Developing Common Instruments for Health Surveys*. Amsterdam: IOS Press, 2003:21–34.

20 Heliövaara M, Aromaa A, Klaukka T, et al. Reliability and validity of interview data on chronic diseases. *J Clin Epidemiol* 1993;46:181–91.

21 Hupkens C. Coverage of health topics by surveys in the European Union (Eurostat Working Paper, Population and Social Conditions). Brussels: European Commission, Statistical Office of the European Communities, 1998.

22 Maggi S, Zucchetto M, Grigoletto F, et al. The Italian Longitudinal Study on Aging (ILSA): Design and methods. *Aging* 1994;6:464–73.

23 Schmidt S, Power M, Bullinger M. The conceptual interrelationship between health determinants and outcomes. *Clin Psychol Psychother* 2005;12:28–50.

24 Meltzer H. Development of a common instrument for mental health. In Nosikov A, Gudex C, editors. *Developing Common Instruments for Health Surveys*. Amsterdam: IOS Press, 2003:35–60.

25 Rumpf HJ, Meyer C, Hapke U, et al. Screening for mental health: validity of the MHI-5 using DSM-IV Axis I psychiatric disorders as gold standard. *Psychiatry Res* 2001;105:243–53.

26 Ware JE, Gandek B, Kosinski MA, et al. The equivalence of SF-36 summary health scores estimated using standard and country-specific agorithms in 10 countries: Results from the IQOLA Project. *Soc Sci Med* 1998;51:1167–70.

27 Ware JE, Kosinski MA, Gandek BL, et al. The factor structure of the SF-36 health survey in 10 countries: Results from the IQOLA Project. International quality of life assessment. *J Clin Epidemiol* 1998;51:1159–65.

28 Jenkinson C, Layte R. Development and testing of the UK SF-12. *J Health Serv Res Policy* 1997;2:14–8.

29 Ware JE, Gandek BL, Keller SD, et al. Evaluating instruments used cross-nationally: Methods from the IQOLA Project. In Spilker B, editor. *Quality of Life and Pharmaeconomics in Clinical Trials*. 2nd edition. Philadelphia: Lippincott-Raven Publishers, 1996:80–7.

30 Brevik JI, Dalgard OS. The Oslo Health Profile Inentory, 1996.

31 Dalgard OS. Community health profile as tool for psychiatric prevention. In Trent DR, Reed C, editors. *Promotion of Mental Health*. Aldershot: Avebury, 1996:681–95.

32 De Bruin A, Beukenhorst D. Development of a common instrument for use of curative services. In Nosikov A, Gudex C, editors. *Developing Common Instruments for Health Surveys*. Amsterdam: IOS Press, 2003:93–119.

33 Arbuckle JL. Amos 4.01. Chicago: SmallWaters Corporation, 1999.

34 Bentler P, Wu E. EQS for Windows User's Guide, 2001.

35 RUMM Laboratory. User Manual, 2000.

36 von Davier M. WINMIRA V1.74 User Manual. A program system for analyses with the Rasch Model, with Latent Class Analysis and with Mixed Rasch Model. Kiel: Institute for Science Education, 1996.

37 Bjorner JB, Kosinski M, Ware JE. Calibration of an item pool for assessing the burden of headaches: An application of item response theory to the Headache Impact Test (HIT). *Qual Life Res* 2003;12:913–33.

38 Hu L, Bentler PM. Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equation Model* 1999;6:1–55.

39 Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods Psychol Res Online* 2003;8:23–74.

40 Zumbo BD. *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense, 1999.

41 Bjorner JB, Kreiner S, Ware JE, et al. Differential Item Functioning in the Danish translation of the SF-36. *J Clin Epidemiol* 1998;51:1189–202.

42 Petersen MA, Groenvold M, Bjorner JB, et al. Use of differential item functioning analysis to validate translations of a health status questionnaire. *Qual Life Res* 2003;12:373–85.