

Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study

Ulrike Ravens-Sieberer · Nora Wille · Xavier Badia · Gouke Bonsel ·
Kristina Burström · Gulia Cavrini · Nancy Devlin · Ann-Charlotte Egmar ·
Narcis Gusi · Michael Herdman · Jennifer Jelsma · Paul Kind ·
Pedro R. Olivares · Luciana Scalone · Wolfgang Greiner

Accepted: 21 March 2010 / Published online: 17 April 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract

Purpose To examine the feasibility, reliability, and validity of the newly developed EQ-5D-Y.

Methods The EQ-5D-Y was administered in population samples of children and adolescents in Germany, Italy, South Africa, Spain, and Sweden. Percentages of missing values and reported problems were calculated. Test–retest reliability was determined. Spearman’s rank correlation coefficients with other generic measures of HRQOL were calculated. Known groups’ validity was examined by

comparing groups with a priori expected differences in HRQOL.

Results Between 91 and 100% of the respondents provided valid scorings. Sweden had the lowest proportion of reported problems (1–24.9% across EQ-5D-Y dimensions), with the highest proportions in South Africa (2.8–47.3%) and Italy (4.3–39.0%). Percentages of agreement in test–retest reliability ranged between 69.8 and 99.7% in the EQ-5D-Y dimensions; Kappa coefficients were up to 0.67. Correlation coefficients with other measures of self-rated health indicated convergent validity (up to $r = -0.56$). Differences between groups classified according to

The EQ-5D-Y is a copyrighted instrument. All requests for EQ-5D-Y translations should be sent to the EuroQol Executive Office in Rotterdam, the Netherlands (userinformationservice@euroqol.org).

U. Ravens-Sieberer (✉) · N. Wille
Department of Psychosomatics in Children and Adolescents,
Research Unit Child Public Health, University Medical Center
Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
e-mail: ravens-sieberer@uke.de

X. Badia
Health Economics and Outcomes Research, IMS Health,
Doctor Ferran 25-27, 2, 08034 Barcelona, Spain

G. Bonsel
Institute of Health Policy and Management, and Department
of Prenatal Medicine and Obstetrics (Location Woudenstein,
L3-060), Erasmus Medical Centre, P.O. Box 1738,
3000 DR Rotterdam, The Netherlands

K. Burström
Department of Learning, Informatics, Management and Ethics,
Medical Management Centre, Karolinska Institutet,
Nobels väg 15a, 171 77 Stockholm, Sweden

K. Burström
Department of Public Health Sciences, Division of Social
Medicine, Karolinska Institutet, Nobels väg 15a,
171 77 Stockholm, Sweden

G. Cavrini
Department of Statistics, University of Bologna, Via delle Belle
Arti 41, 40126 Bologna, Italy

N. Devlin
Office of Health Economics and Senior Associate, King’s Fund,
12 Whitehall, London SW1A2DY, UK

A.-C. Egmar
The Red Cross University College, Box 55676,
102 15 Stockholm, Sweden

N. Gusi · P. R. Olivares
Faculty of Sport Sciences, University of Extremadura,
Avda Universidad, 10071 Cáceres, Spain

M. Herdman
CIBER en Epidemiología y Salud Pública (CIBERESP),
Barcelona, Spain

M. Herdman
IMIM-Hospital del Mar, Unitat de Recerca en Serveis Sanitaris,
Parc de Recerca Biomedica de Barcelona, Health Services
Research Unit, Doctor Aiguader, 88, 08003 Barcelona, Spain

presence of chronic conditions, self-rated overall health and psychological problems provided preliminary evidence of known groups' validity.

Conclusions Results provide preliminary evidence of the instrument's feasibility, reliability and validity. Further study is required in clinical samples and for possible future applications in economic analyses.

Keywords Child health · Adolescent health · HRQOL · Measurement · EQ-5D

Abbreviations

HRQOL Health-related quality of life
VAS Visual analogue scale
EQ-5D-Y EQ-5D youth version

Introduction

The assessment of young people's health-related quality of life (HRQOL) is considered to be of increasing importance in public health research and the evaluation of medical and psychosocial treatment [1, 2]. A large number of measures of HRQOL have been developed specifically for children and adolescents (here defined as persons aged 8–11 and 12–18, respectively) taking the special requirements in these age-groups into account [1, 3–6]. However, one disadvantage of those instruments is their lack of correspondence to adult HRQOL instruments. This shortcoming makes it difficult to track changes in HRQOL across the life course in, for example, cohort studies investigating

severe or progressive chronic childhood conditions that last into adulthood. It is therefore desirable to have a modified version of an adult instrument at hand that is also suitable for younger age-groups and can be used in the transition from childhood and adolescence into adulthood.

The generic EQ-5D is a brief and easy to administer instrument that provides scores for different health dimensions as well as an index value which can be used to assess health status and is useful in health economic analyses. Since the EQ-5D has been utilized internationally in many different settings, such as clinical trials and population surveys [7], the instrument was considered a suitable candidate for development of a modified version that could be used in children and adolescents. Within the framework of an international task force on behalf of the EuroQol Group including 13 experts in quality of life research from seven countries (Germany, Italy, South Africa, Spain, Sweden, the Netherlands, United Kingdom), a version for use in respondents from 8 years onwards—the EQ-5D-Y—was developed based on the standard adult EQ-5D. All experts additionally had specific expertise in child psychology, paediatrics, health economics, statistics, sport sciences, or rehabilitation sciences. The methodology of the questionnaire development process of the EQ-5D-Y as well as background information regarding the modifications and their consequences are described elsewhere [8]. In summary, the development process included the revision of the content and wording of EQ-5D to ensure relevance and clarity for young respondents. After translation of the resulting modified version, cognitive interviews were conducted in Germany, Italy, Spain and Sweden to test the instrument's comprehensibility in children and adolescents. Results indicated the adapted EQ-5D-Y was satisfactorily understood by young respondents in different countries and that it might be a useful tool to measure HRQOL in children and adolescents in an age-appropriate manner.

In order to investigate the feasibility, reliability, and validity of the EQ-5D-Y in a multinational, multilingual context, a series of national validation studies were undertaken which were coordinated and methodologically harmonized to ensure the comparability of the findings. The results from the validation studies performed in five countries (Germany, Italy, South Africa, Spain, Sweden) are presented in the current paper.

Since the EQ-5D is widely used for economic evaluation purposes, many questions arise with respect to the EQ-5D-Y regarding the possible development of preference weights in the future. Even though this paper concentrates on the new EQ-5D-Y as a stand-alone outcome measure as it is used in many settings (such as population health surveys, routine health system use, and use in clinical settings), we will address some of these important questions in an outlook at the end.

J. Jelsma

School of Health and Rehabilitation Sciences, Division of Physiotherapy, University of Cape Town, Anzio Road Observatory, 7925 Cape Town, South Africa

P. Kind

Centre for Health Economics, Alcuin College, University of York, York YO10 5DD, UK

L. Scalone

Center of Pharmacoeconomics, University of Milan, Via Vanvitelli 32, 20129 Milan, Italy

L. Scalone

CHARTA Foundation, c/o Department of Medical Pharmacology, University of Milan, Via Vanvitelli 32, 20129 Milan, Italy

W. Greiner

Department for Health Economics and Health Care Management, School of Public Health, University of Bielefeld, P.O. Box 10 01 31, 33501 Bielefeld, Germany

Methods

Study design and sample description

In general, study methodologies used in the different countries were comparable; however, some variations in individual countries were permitted so that local teams could study specific research questions. National representativeness was not a priority in this study as representative population samples are not a requirement of validation studies. The main characteristics of the national studies are presented in Table 1. A minimum number of $n = 200$ EQ-5D-Y respondents per country was required according to previous power calculations [8]. In order to include respondents aged 8 and older from the general population, schools were used for recruitment in all countries, except Sweden (household sample). Only mainstream schools, i.e. not special-needs schools were included in the sample. Pupils who were present on the day of questionnaire administration constituted the study sample. They completed the questionnaires in the classroom on their own and were provided with short written instructions. In Italy, an investigator was present to provide assistance. No restrictions were made regarding the mother tongue of the pupils, although the questionnaires were presented in the language of instruction in the schools. In Italy, only native Italian speakers were included. In Sweden, families received the questionnaire, a letter to the parents explaining the study, and a similar letter for the child by post. A reminder was sent after 2 weeks, which included the questionnaire and the two letters. The letters emphasized that the child should complete the questionnaire on his or her own.

Test–retest procedures were conducted in Italy and in Spain in order to investigate reliability of EQ-5D-Y, where a third of the study sample received the questionnaire again 7–10 days after the first examination.

In all countries, informed consent from parents or guardians was a precondition for children and adolescents

to be able to participate in the study. Depending on national regulations, permission to collect data was obtained from the data protection commissioner in charge (Germany, Italy, and Spain) or the appropriate ethics committee (Sweden: Karolinska Institutet Number 2006/1534-31/2; South Africa: Medical Research Ethics Committee of the University of Cape Town and South African Department of Education).

Instruments and variables

To examine convergent and known group validity of the EQ-5D-Y, a core set of internationally standardized instruments and variables was administered alongside the EQ-5D-Y and questions regarding basic socio-demographic information (age, gender, level of education, migration status) in all national studies. Since these instruments needed to be available in a variety of languages and to have been shown to be valid for use in cross-cultural comparison studies, many instruments from international studies [9, 10] were employed here. The core set included measures of HRQOL and subjective health to analyse convergent validity as well as indicators of mental and somatic health problems, e.g. a screener for emotional and behavioural problems.

EQ-5D-Y

The EQ-5D-Y was developed from the EQ-5D by adapting the original questionnaire to the requirements of measuring HRQOL in children and adolescents from 8 years onwards [8]. As in the adult version, it consists of a descriptive system that comprises five items referring to mobility ('walking about'), self-care ('looking after myself'), usual activities ('doing usual activities'), pain and discomfort ('having pain or discomfort'), and anxiety and depression ('feeling worried, sad or unhappy'). Each item has three levels of problems reported (no problems, some problems,

Table 1 Characteristics of the different national samples

	Germany	Italy	South Africa	Spain	Sweden
Data collection	School (pupils)	School (pupils)	School (pupils)	School (pupils)	Postal survey (from general population)
Study design	Cross-sectional	Cross-sectional & retest	Cross-sectional	Cross-sectional & retest	Cross-sectional
Mode of recruitment	Random sample	Convenience sample	Convenience sample	Convenience sample	Random sample
Method of application	Paper–pencil	Paper–pencil	Paper–pencil	Paper–pencil	Paper–pencil
<i>N</i>	756	415	258	973	407
Male	385	199	130	494	208
Female	371	216	128	479	199
Age-range	10–18	8–15	13–19	8–18	8–16
Mean age (SD)	13.8 (1.9)	11.8 (2.2)	15.5 (1.3)	13.0 (2.7)	13.2 (2.7)

a lot of problems). The EQ-5D-Y also includes an easily understandable modification of the vertical, graduated Visual Analogue Scale (VAS) of EQ-5D, where the respondent rates his or her overall health status on a scale from 0 and 100 with 0 representing the worst and 100 the best health state he or she can imagine. All items refer to the health state ‘today’.

KIDSCREEN-27

The generic KIDSCREEN-27 was administered as a cross-cultural measure to assess HRQOL in children and adolescents aged 8–18. Its five Rasch-scaled dimensions provide detailed profile information on physical well-being, psychological well-being, autonomy & parents, peers & social support, and school environment within the last week. The instrument has been shown to have good psychometric properties with internal consistencies of the subscales ranging between 0.80 and 0.84 [10]. In addition, the KIDSCREEN-10 Index score provides an overall measure of global HRQOL using 10 of the KIDSCREEN-27 items [11].

PedsQL

The PedsQL™ Quality of Life generic core Scales were administered in Italy. The instrument consists of 23 items that can be grouped in 4 multidimensional scales (physical functioning, emotional functioning, social functioning, school functioning) and 3 summary scales [12]. The PedsQL refers to the last month and is suitable for self-completion by respondents aged 8–18.

Self-rated health

The general health item asks the respondent how he or she would describe his or her health in general and was used in all countries as a measure of perceived health status. Response options were ‘*excellent*’, ‘*very good*’, ‘*good*’, ‘*fair*’ and ‘*poor*’. This question has been used in large international health surveys in children and adolescents and has been shown to be a valid measure of subjective health [13].

Cantril-ladder

The adapted version of Cantril’s ‘life-satisfaction-ladder’ [14] used in WHO surveys in children and adolescents was included to measure general subjective life satisfaction. Respondents were presented with the picture of a ladder with steps ranging from 0 to 10 and asked to indicate where on the ladder they ‘feel they are standing at the moment’ with the top of the ladder (10) representing the best

possible life and the bottom (0) representing the worst possible life.

Strength and Difficulties Questionnaire (SDQ)

To test for known group validity differences between responders with and without emotional and behavioural problems, the SDQ [15, 16] was administered in Germany and Spain. The SDQ is a short behavioural screening instrument that asks the respondent for 20 symptoms of mental health problems within the last 6 months (regarding behaviour, emotions, hyperactivity-inattention and peer problems). A total difficulties score can be calculated and is recoded into three categories (normal, borderline, abnormal mental health problems). For the present study, the borderline and abnormal categories were collapsed.

Chronic condition

Responders were asked ‘Do you have a long-term chronic condition or disability which had been diagnosed by a health professional?’ and ‘Do you take medicine for your long-term illness, disability or medical condition?’ to establish whether respondents had a longstanding illness, disability or medical condition. The question about medication should indicate the severity of the condition reported [17].

Statistical analysis

The feasibility and acceptability of EQ-5D-Y was investigated by calculating the percentage of missing values and inappropriate responses on the descriptive system and VAS. A missing value was defined as a respondent completely leaving out an item or the VAS. Responses on the VAS that did not indicate unambiguously one score (e.g. by drawing a circle that included more than one score) were also defined as missing values. Responses that did not follow the instruction of drawing a line from the box to the chosen VAS score but which provided an unambiguously interpretable score (e.g. when a pupil used the VAS like a thermometer, drawing a line from the bottom to one score) were defined as inappropriate responses. Frequencies of reported problems were calculated for all samples from the five participating countries.

In order to investigate reliability, the percentage of agreement and kappa coefficients [17] was calculated to estimate concordance between test and retest responses (‘no problem’ versus ‘any problems’) in each profile domain. For the VAS, the intraclass correlation coefficient (ICC) [18] was computed. Kappa values were interpreted according to Landis and Koch’s guidelines [19] with kappa < 0.2 indicating poor agreement, 0.21–0.40 indicating fair agreement,

0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and kappa > 0.81 indicating almost perfect agreement. An ICC > 0.7 is generally considered as acceptable for test–retest reliability. When applicable, a P -value < 0.05 (two-tailed test) was considered as statistically significant.

Convergent validity was investigated by determining the correlations between EQ-5D-Y dimensions and VAS and previously validated measures of child HRQOL using Spearman's rank and Pearson's correlation coefficients, respectively [20]. In line with the guidelines provided by Cohen et al. [21], coefficients from 0.1 to 0.29 were deemed to be low, 0.3 to 0.49 moderate and correlations of 0.5 or above as high. It was hypothesized that the mobility and the pain and discomfort dimension of the EQ-5D-Y would show a moderate correlation with dimensions of physical well-being and other QoL measures. Even though the level of an individual's physical activity, energy and fitness as assessed by the KIDSCREEN-27 physical well-being dimension is not directly related to the experience of pain and discomfort as assessed by the relevant EQ-5D-Y dimension, a moderate relationship could be expected as both aspects refer to physical health and well-being. Further, it was hypothesized that the 'feeling worried, sad or unhappy dimension' would show a moderate to high correlation with dimensions of psychological well-being and that a similar level of correlation would be seen between the VAS and overall scores of QoL measures as well as with general health items and life satisfaction.

The known groups' validity of the EQ-5D-Y was examined by comparing the results on the descriptive system between groups which were a priori expected to show differences in HRQOL. Groups analysed were: (a) those reporting a chronic condition and taking medication for that condition versus those reporting no chronic condition, (b) those with excellent, very good or good self-reported health versus those with fair or poor self-reported health on the General Health Item, and (c) those who had mental health problems based on their SDQ scores versus those who did not. Comparisons were performed using χ^2 -tests, and the categories of 'some' and 'a lot of problems' were collapsed to one category ('any problems').

Results

Feasibility

Complete data was obtained for 91–100% of respondents from the general population samples depending on the country and the part of the EQ-5D-Y (descriptive system vs. VAS). Missing or inappropriate responses on the EQ-5D-Y dimensions ranged from 0% in Spain and Italy to 2%

in South Africa. On the VAS, the level of missing or inappropriate responses ranged from 0% in Italy to 9% in Germany.

Distribution of EQ-5D scores: frequencies of reported problems

Table 2 shows the proportion of respondents reporting problems on EQ-5D-Y dimensions by country. In all countries, the highest proportions of problems (i.e. 'some'/'a lot') were reported on the 'having pain and discomfort' and 'feeling worried, sad or unhappy' dimensions. In all countries, problems were reported least frequently on the 'looking after myself' dimension. High ceiling effects were seen in all dimensions. In general, the proportion of respondents reporting any problem was highest in Italy and South Africa. Swedish respondents reported fewer problems compared to the other countries.

Reliability

Table 3 shows the results on test–retest reliability. In the descriptive system, a test–retest agreement was observed in 69.8–93.8% of Italian youths and in 86.2–99.7% of Spanish respondents. This agreement is generally confirmed by the kappa coefficients. However, the high ceiling effects in the descriptive system of EQ-5D-Y caused some apparent non-confirmation of the results. In Italy, no kappa coefficient could be computed for the self-care domain since all children reported no problems in the retest. Similarly, the kappa coefficients in the mobility dimension are of limited value since nearly all retest responses were in the 'no problems' category. The intraclass correlation coefficients (ICCs) for the VAS were 0.82 in Italy and 0.83 in Spain.

Convergent validity

Regarding convergent validity, Table 4 shows the correlations between the EQ-5D-Y and selected KIDSCREEN-27 dimensions and the KIDSCREEN-10 score by country. Similar patterns of associations between the scores on both instruments in the different countries could be observed. As hypothesized, the two dimensions dealing with psychological well-being in the EQ-5D-Y and the KIDSCREEN-27 showed moderate to high correlations ($r = -0.41$ to -0.52), suggesting convergent validity. In Italy, the Spearman rank coefficient between PedsQL Emotional functioning and 'feeling worried, sad or unhappy' was $\rho = -0.47$ (data not shown in Table 4). Conversely, the mobility dimension of the EQ-5D-Y barely correlated with the psychological well-being dimension of the KIDSCREEN-27 (Table 4) or the PedsQL Emotional functioning Scale ($\rho = -0.10$). Generic questions on well-being and life

Table 2 Percentages of reported problems in the EQ-5D-Y

	Germany (<i>n</i> = 756)		Italy (<i>n</i> = 415)		South Africa (<i>n</i> = 258)		Spain (<i>n</i> = 973)		Sweden (<i>n</i> = 407)	
	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>
Mobility (walking about)										
No	91.9	695	93.5	388	85.3	220	95.3	927	96.3	392
Some	7.5	57	6.3	26	12.0	31	4.6	45	2.7	11
A lot of	0.1	1	0.2	1	0.4	1	0.1	1	0.0	0
Missing values	0.4	3	0	0	2.3	6	0	0	1.0	4
Looking after myself										
No	97.9	740	95.7	397	95.0	245	98.6	959	98.0	399
Some	1.6	12	4.1	17	2.7	7	1.1	11	1.0	4
A lot of	0.3	2	0.2	1	0.4	1	0.3	3	0.0	0
Missing values	0.3	2	0	0	1.9	5	0	0	1.0	4
Doing usual activities										
No	93.3	705	84.3	350	82.6	213	93.7	912	95.8	390
Some	6.2	47	14.9	62	14.7	38	5.7	55	2.7	11
A lot of	0.3	2	0.7	3	0.8	2	0.6	6	0.3	1
Missing values	0.3	2	0	0	1.9	5	0	0	1.2	5
Having pain or discomfort										
No	62.0	469	61.0	253	51.6	133	80.0	778	74.0	301
Some	36.0	272	38.3	159	45.7	118	19.0	185	23.6	96
A lot of	1.3	10	0.7	3	0.8	2	1.0	10	0.7	3
Missing values	0.7	5	0	0	1.9	5	0	0	1.7	7
Feeling worried, sad or unhappy										
Not	59.8	452	61.0	253	60.1	155	76.9	748	80.3	327
A bit	35.8	271	34.7	144	34.1	88	22.0	214	18.2	74
Very	3.8	29	4.3	18	3.9	10	1.1	11	0.5	2
Missing values	0.5	4	0	0	1.9	5	0	0	1.0	4

Table 3 Test–retest reliability in Italy and Spain (retest after 7–10 days)

	Italy		Spain	
	Kappa coefficient	Agreement (%)	Kappa coefficient	Agreement (%)
Mobility (walking about)	0.222	91.5	−0.003*	99.4
Looking after myself	0.000 ^a	93.8	0.665*	99.7
Doing usual activities	0.352*	82.9	0.557*	97.5
Having pain or discomfort	0.350*	69.8	0.435*	86.2
Feeling worried, sad or unhappy	0.549*	78.3	0.468*	87.4

* Significant at $P \leq 0.01$ ^a The responses in the retest are identical

satisfaction such as the self-rated general health item, the KIDSCREEN-10 QoL Index and the Life Satisfaction Ladder showed moderate to high correlations with the EQ-5D-Y VAS (correlation coefficients from 0.33 to 0.56), which also suggests adequate convergent validity. Associations between the KIDSCREEN-27 dimension of physical well-being and the EQ-5D-Y mobility and pain or discomfort dimension were low and did not reach the expected correlation threshold of $\rho = 0.3$. The same was true for the correlations between the PedsQL Physical functioning scale and the EQ-5D-Y

mobility and pain dimensions. The ‘looking after myself’ and ‘usual activities’ dimensions did not correlate with any of the selected established HRQOL dimensions or with the general health item or life satisfaction ladder.

Known groups’ validity

As presented in Table 5, respondents reporting a chronic condition and taking medication for that condition reported significantly more problems on the EQ-5D-Y dimensions of

Table 4 Convergent validity: Spearman rank Correlation between EQ-5D-Y and KIDSCREEN, general health and life satisfaction scores (significant correlations are given in bold)

	Germany (<i>n</i> = 756 ^a)	South Africa (<i>n</i> = 258 ^a)	Spain (<i>n</i> = 973 ^a)	Sweden (<i>n</i> = 407 ^a)
KIDSCREEN-10-HRQOL-index				
Mobility (walking about)	-0.14	-0.27	0.04	-0.17
Looking after myself	0.03	0.08	0.01	-0.13
Doing usual activities	-0.13	-0.32	-0.10	-0.19
Having pain or discomfort	-0.28	-0.39	-0.15	-0.30
Feeling worried, sad or unhappy	-0.45	-0.45	-0.14	-0.46
VAS ^b	0.33	0.54	0.43	0.48
KIDSCREEN-27 physical well-being				
Mobility (walking about)	-0.10	-0.25	-0.08	-0.15
Looking after myself	0.00	0.03	-0.05	-0.08
Doing usual activities	-0.09	-0.21	-0.09	-0.17
Having pain or discomfort	-0.27	-0.26	-0.18	-0.22
Feeling worried, sad or unhappy	-0.27	-0.29	-0.23	-0.34
VAS ^b	0.33	0.55	0.46	0.47
KIDSCREEN-27 psychological well-being				
Mobility (walking about)	-0.10	-0.21	-0.10	-0.10
Looking after myself	0.04	0.06	-0.04	-0.08
Doing usual activities	-0.09	-0.17	-0.14	-0.18
Having pain or discomfort	-0.28	-0.34	-0.23	-0.27
Feeling worried, sad or unhappy	-0.46	-0.52	-0.41	-0.49
VAS ^b	0.33	0.48	0.21	0.48
Self-rated general health item^c				
Mobility (walking about)	0.13	0.02	-0.01	0.15
Looking after myself	-0.04	-0.09	0.01	0.06
Doing usual activities	0.08	0.27	0.14	0.14
Having pain or discomfort	0.36	0.26	0.21	0.25
Feeling worried, sad or unhappy	0.26	0.25	0.20	0.42
VAS ^b	-0.46	-0.56	-0.50	-0.51
Life satisfaction ladder				
Mobility (walking about)	-0.07	-0.14	-0.12	-0.10
Looking after myself	0.01	-0.01	0.01	-0.01
Doing usual activities	-0.05	-0.17	-0.10	-0.10
Having pain or discomfort	-0.20	-0.15	-0.18	-0.19
Feeling worried, sad or unhappy	-0.32	-0.26	-0.31	-0.37
VAS ^b	0.36	0.51	0.54	0.45

^a Missing values were excluded casewise

^b Pearson correlation coefficient

^c In Italy (*n* = 415), Spearman rank correlations of the General Health Item with the EQ-5D-Y descriptive system were for mobility $\rho = 0.15$, 'looking after myself' $\rho = 0.04$, 'doing usual activities' $\rho = 0.18$, 'having pain and discomfort' $\rho = 0.17$, 'feeling worried, sad and unhappy' $\rho = 0.20$; Pearson correlation coefficient with the VAS was $r = -0.52$

mobility (South Africa, Spain, Sweden), 'looking after myself' (Sweden), 'doing usual activities' (Germany, South Africa), 'having pain or discomfort' (Germany, South Africa) and 'feeling worried, sad or unhappy' (Germany, South Africa, Sweden) than those who declared no chronic condition for which they were taking medication.

Respondents with 'fair' or 'poor' self-reported health displayed significantly more problems ($P < 0.05$) on the 'mobility' (Germany, South Africa), 'usual activities' (Italy, South Africa, Spain, Sweden), 'pain and discomfort' (Germany, Spain, Sweden) and 'anxiety and depression'

(Germany, South Africa, Spain, Sweden) dimensions than respondents with 'excellent', 'very good' or 'good' self-reported health (data not shown).

In Germany and Spain, a comparison between respondents with and without mental health problems was conducted (data not shown). Those with borderline and abnormal SDQ scores reported significantly ($P < 0.05$) more problems on four EQ-5D-Y dimensions (mobility, usual activities, pain/discomfort and anxiety/depression), though the largest differences were seen on the 'feeling worried, sad or unhappy' dimension (61.9% reporting

Table 5 Comparison of reported problems on EQ-5D-Y in those with and without self-reported chronic conditions

<i>n</i>	Germany		Italy		South Africa		Spain		Sweden	
	Chronic condition		Chronic condition		Chronic condition		Chronic condition		Chronic condition	
	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
	642	95	382	33	188	42	818	155	378	18
Mobility ^{b,c,d} (walking about)										
No	92.7	90.4	93.7	90.9	89.4	80.5	95.7	92.9	97.6	88.9
Some	7.3	8.5	6.0	9.1	10.1	19.5	4.3	6.5	2.4	11.1
A lot of	0.0	1.1	0.3	0	0.5	0	0	0.6	0.0	0.0
Looking after myself ^d										
No	98.1	98.9	95.5	97.0	96.3	100	98.9	96.8	99.2	94.4
Some	1.7	0.0	4.2	3.0	3.2	0	0.9	2.6	0.8	5.6
A lot of	0.2	1.1	0.3	0	0.5	0	0.2	0.6	0.0	0.0
Doing usual activities ^{a,b}										
No	93.8	91.6	84.6	81.8	87.8	66.7	94.4	90.3	97.1	94.4
Some	5.9	8.4	14.6	18.2	12.2	31.0	5.0	9.0	2.9	0.0
A lot of	0.3	0.0	0.8	0	0	2.3	0.6	0.7	0.0	5.6
Having pain or discomfort ^{a,b}										
No	66.0	41.5	61.5	54.4	58.5	30.9	80.0	80.0	75.4	72.2
Some	32.9	55.3	37.7	45.6	41.0	66.7	19.0	18.7	24.1	22.2
A lot of	1.1	3.2	0.8	0	0.5	2.4	1.0	1.3	0.5	5.6
Feeling worried, sad or unhappy ^{a,b,d}										
Not	60.8	54.3	62.0	48.5	63.8	42.9	77.8	72.3	82.3	61.1
A bit	35.4	42.5	34.1	42.4	34.0	42.9	21.4	25.2	17.2	38.9
Very	3.8	3.2	3.9	9.1	2.1	14.2	0.9	2.6	0.5	0.0

^a Significant differences ($P < 0.05$) between the groups ('no problem' vs. 'any problem') in Germany

^b Significant differences ($P < 0.05$) between the groups ('no problem' vs. 'any problem') in South Africa

^c Significant differences ($P < 0.05$) between the groups ('no problem' vs. 'any problem') in Spain

^d Significant differences ($P < 0.05$) between the groups ('no problem' vs. 'any problem') in Sweden

problems versus 33.4% for those with and without mental health problems, respectively, in Germany and 43.5 vs. 19.8%, respectively, in Spain).

Discussion

This study aimed to examine the feasibility, reliability and validity of the newly developed EQ-5D-Y in four European countries and South Africa.

The results clearly show that the EQ-5D-Y is easy to fill in, has few missing values and is highly feasible for children as a HRQOL measure. The very low proportions of missing values in Italy and Sweden may be due to the fact that an investigator was at hand to help if necessary (Italy), or because some children might have received assistance at home (Sweden). On the whole, however, the overall small proportion of missing or inappropriate responses confirmed the feasibility of the EQ-5D-Y. Furthermore, the fact that there are only small differences regarding non-responses

between the countries suggests the instrument might be viable in a cross-cultural setting. The most frequent problems were observed in filling out the VAS, suggesting that there is potential for further refinement of its presentation and instruction.

In general, only a low prevalence of severe problems was reported in the different dimensions of the EQ-5D-Y, which is typical for general population samples. The highest proportion of problems was reported on the 'having pain or discomfort' and 'feeling worried, sad or unhappy' dimensions. For the other EQ-5D-Y dimensions of mobility, 'looking after myself' and 'doing usual activities', only relative small proportions of respondents reported problems. The very high ceiling effects of up to 99% (especially in the 'looking after myself dimension) are connected to several methodical limitations of the new instrument. The findings indicate that the ability of EQ-5D-Y to detect moderate impairments of HRQOL is limited and that consequently the instrument might not be very capable of discriminating between respondents in the general population.

Furthermore, the large ceiling effects in the test scores cause problems in determining the instruments psychometric properties such as convergent validity and reliability. In this regard, more differentiated response options can be considered to be helpful to improve the EQ-5D-Y in the future. A five level response choice of the EQ-5D is currently in development (EuroQol group, personal communication). On the basis of the data presented here, the development of such a modified measure can be highly recommended.

The EQ-5D-Y shows fair to moderate levels of test-retest reliability, with high percentage of youths reporting the same levels of problems in the profile domains and satisfactory ICC with respect to the VAS. However, as noted above—the examination of reliability was limited by partly high ceiling effects. Reliability should therefore be further tested in a different context—e.g. clinical samples—to reduce these ceiling effects.

Regarding convergent validity, we interpreted correlation coefficients according to the guidelines provided by Cohen et al. [21]. In interpreting validity correlations, it has to be considered that due to measurement errors, a correlation can never reach the maximum of 1 but only the square root of the product of the reliabilities of the instruments involved. Against this background, it can be said that the EQ-5D-Y demonstrated convergent validity and displayed distinct patterns of association with child-specific measures of HRQOL and other comparable scales. As expected, the VAS, as an overall measure of global health, showed the highest correlation with the KIDSCREEN-10 Index of general HRQOL, the General Health Item, and with the Life Satisfaction Ladder. The VAS was also associated with both physical well-being and psychological well-being, suggesting that VAS scores are driven by aspects of both physical and psychological health.

The EQ-5D-Y dimension ‘feeling worried, sad or unhappy’ displayed convergent validity in terms of a strong association with the KIDSCREEN-27 and PedsQL Psychological Well-being dimension, and discriminant validity [22] in terms of low correlation with other health information. The EQ-5D-Y dimensions ‘mobility’ failed to display convergent validity—at least with KIDSCREEN-27 Physical Well-being dimension. However, it can be argued that by looking at the content of the Physical Well-being dimension of the KIDSCREEN, the latter is more focussed on physical well-being/energy level and less on physical functioning than is the case with the EQ-5D mobility dimension. Additionally, again the reduced variation in EQ-5D-Y test scores generally limits the possibilities for correlations with other measures. This might be improved by extending the range of response options, as mentioned above.

In general, due to the lack of objective data on the health of participants, the results on known groups’ validity have

to be interpreted carefully. Overall, the response categories were used in a more differentiated manner by respondents who reported health problems. Even though a number of meaningful differences between the ‘known groups’ could be detected by the EQ-5D-Y, for no health attribute significant differences across all five countries could be observed (irrespective of the indicator such as presence of chronic conditions, impaired self-reported or mental health). In general, the observed ability of the EQ-5D-Y to discriminate between the compared groups supports the validity of all its dimensions but ‘looking after myself’. However, due to the large ceiling effects, only respondents with severe health problems seem to be identified reliably with the instrument. The fact that differences were not seen may also be partly attributable to the types of conditions present. For example, some of the children who report a chronic condition might do so due to an allergy with minor symptoms and thus cannot be expected to differ that much in HRQOL from their ‘healthy’ peers. Furthermore, all children (except for the Swedish household sample) were obviously healthy enough to attend school and thus cannot suffer from a very serious condition.

Even though we observed substantial ceiling effects on most EQ-5D-Y dimensions, these results are consistent with those observed when using the EQ-5D in population health surveys [23]. Although ceiling effects with the adult version have been shown to be higher than those of other measures such as the SF-12 and HUI3, the EQ-5D was nevertheless shown to perform as well or better than those other instruments in terms of discriminant validity [24]. It should also be noted that the EQ-5D-Y actually reduced the ceiling effect on some dimensions in comparison with the EQ-5D [8]. Finally, it should be remembered that these were general population samples, where higher ceiling effects would be expected, and that further testing of the EQ-5D-Y is required in clinical samples, where the ceiling effect would likely be significantly reduced.

This study has several strengths, but also some limitations. All samples included comprised children and adolescents from the general population. Thus, no information on the performance of EQ-5D-Y in specific populations is available. Another limitation is that due to ethical constraints, it was not possible to obtain additional clinical data on respondents’ physical and psychological health status. Instead, several screening instruments were used. However, these additional screeners represent self-report questionnaires as well. Thus, to a certain extent, the association between these additional measures and the EQ-5D-Y might be attributable to the ‘same source of information bias.’ The statistical and psychometric analyses reported in this paper represent a first examination of the EQ-5D-Y psychometric properties. It was beyond the scope of this paper to examine other issues, such as sensitivity to

change, which should be examined in future studies. Similarly, the content validity, i.e. whether the instrument encompasses all aspects of HRQOL that are important in children and adolescents was not examined, though as stated earlier, the intention was to adapt an adult tool for use in children primarily to allow for follow-up and comparisons over a wide range of ages.

Another important topic that could not be appropriately addressed within the scope of this paper is the further possible use of the new instrument in economic evaluation. Since there are differences between the EQ-5D-Y and the standard EQ-5D, the existing social value sets may not be applicable. Furthermore, valuing EQ-5D-Y health states raises some potentially interesting issues. The normative argument using social preference weights in economic evaluation is that it is the preferences of the general public that are relevant—not those of patients themselves—in making resource allocation decisions. This would suggest that preference weights for the EQ-5D-Y should be established by eliciting values from the general public, in much the same manner as for the EQ-5D [25]. Time Trade Off and other improved methods for eliciting preferences [26] are equally applicable to the valuation for EQ-5D-Y. However, it is unclear whether, in asking the general public to value EQ-5D-Y states, they should be informed that the states they are being asked to consider will be potentially experienced by children. Whether participants are informed or not could conceivably make a difference to the values. Similarly, there may be a systematic difference between the values the general public assign to such states and the values young people themselves place on the states, if they were asked to consider EQ-5D-Y states hypothetical to them. This relates to a wider debate about whose values are relevant in economic evaluations and in how far subgroup preferences (such as young people) are useful in economic evaluation [27, 28].

The issue of valuation of EQ-5D-Y states and appropriate means by which social preferences for those states should be elicited is currently under consideration and discussion and the considerable experience of the entire EuroQoL group is guiding this process. The present paper can only provide a basis for this further discussion, since clearly no weights can be developed until the underlying descriptor domains are found to be reliable and valid.

Conclusion

In summary, this first multinational administration of the newly developed EQ-5D-Y indicates that it is a feasible, reliable and valid instrument for the measurement of HRQOL in children and adolescents. However, the EQ-5D-Y needs further testing in population-based and clinical

studies. Population-based studies could help to establish norms for improved interpretation of test scores. Applying the EQ-5D-Y in clinical studies will allow further testing of feasibility and acceptance as well as test score distribution and psychometric properties in more specific populations and over a wide range of settings. Longitudinal studies are required to investigate the measure's responsiveness with regard to change in clinical status and to monitor the effect of medical interventions. The assignment of utility values to the different health profiles described by the EQ-5D-Y descriptive system should also be a priority in the future.

Acknowledgments This work was financially supported by the EuroQoL Foundation. (The EuroQoL Foundation is a non-profit organization.) We thank Professor Magnus Svartengren and PhD Candidate Sun Sun at Karolinska Institutet, Sweden for participation in data collection and analyses. We also thank the Italian team members: Carlo Tomasetto, Maria C Matteucci and Patrizia Selleri, from the Department of Educational Sciences and Serena Broccoli, from the Department of Statistics, University of Bologna; Barbara Pacelli, from the Epidemiology Unit, Local Health Authority, Bologna, Italy; Francesca Borghetti, Centre of Pharmacoeconomics, University of Milan, Milan, Italy.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Ravens-Sieberer, U., Erhart, M., Wille, N., Wetzel, R., Nickel, J., & Bullinger, M. (2006). Generic health-related quality of life assessment in children and adolescents: Methodological considerations. *Pharmacoeconomics*, *24*, 1199–1220.
2. Sullivan, M. (2003). The new subjective medicine: Taking the patient's point of view on health care and health. *Social Sciences & Medicine*, *56*, 1595–1604.
3. Eiser, C., & Morse, R. (2001). A review of measures of quality of life for children with chronic illness. *Archives of Disease in Childhood*, *84*, 205–211.
4. Clarke, S., & Eiser, C. (2004). The measurement of health-related quality of life (QoL) in paediatric clinical trials: A systematic review. *Health & Quality of Life Outcomes*, *2*, 66.
5. Riley, A. W. (2004). Evidence that school-age children can self-report on their health. *Ambulatory Pediatrics*, *4*, 371–376.
6. Harter, S., & Whitesell, N. R. (1989). Developmental changes in children's understanding of single, multiple and blended emotion concepts. In E. Saarni & P. L. Harris (Eds.), *Children's understanding of emotion* (pp. 81–116). New York: Cambridge University Press.
7. Rabin, R., & de Charro, F. (2001). EQ-5D: A measure of health status from the EuroQoL Group. *Annals of Medicine*, *33*, 337–343.
8. Wille, N., Badia, X., Bonsel, G., Burström, K., Cavrini, G., Egmar, A.-C., et al. (2010). Development of the EQ-5D-Y: A child-friendly version of the EQ-5D. *Quality of Life Research*. doi:10.1007/s11136-010-9648-y.
9. Currie, C., Samdal, O., Boyce, W., et al. (Eds.) (2001). Health behaviour in school-aged children: A WHO Cross-National Study (HBSC), research protocol for the 2001/2002 survey. Child and

- Adolescent Health Research Unit (CAHRU), University of Edinburgh.
10. Ravens-Sieberer, U., Auquier, P., Erhart, M., Gosch, A., Rajmil, L., Bruil, J., et al. (2007). The Kidscreen-27 quality of life measure for children and adolescent: Psychometric results from a cross-cultural survey in 13 European countries. *Quality of Life Research*, *16*, 1347–1356.
 11. Ravens-Sieberer, U., & the European KIDSCREEN Group. (2006). *The KIDSCREEN Questionnaires—Quality of life questionnaires for children and adolescents—Handbook*. Lengerich: Papst Science Publisher.
 12. Varni, J. W., Seid, M., & Rode, C. A. (1999). The PedsQL: Measurement model for the Pediatric Quality of Life Inventory. *Medical Care*, *37*, 126–139.
 13. Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, *38*, 21–37.
 14. Cantril, H. (1965). *The pattern of human concern*. New Brunswick, NJ: Rutgers University Press.
 15. Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581–586.
 16. Goodman, R., Meltzer, H., & Bailey, V. (1998). The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry*, *7*, 125–130.
 17. Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*, 213–220.
 18. Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*, 613–619.
 19. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.
 20. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McCraw-Hill.
 21. Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Publishers.
 22. Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.
 23. Kind, P., Dolan, P., Gudex, C., & Williams, A. (1998). Variations in population health status: Results from a United Kingdom national questionnaire survey. *BMJ*, *316*, 736–741.
 24. Macran, S., Weatherly, H., & Kind, P. (2003). Measuring population health: A comparison of three generic health status measures. *Medical Care*, *41*, 218–231.
 25. Szende, A., Oppe, M., & Devlin, N. (2007). *EQ-5D valuation sets: An inventory, comparative review and users' guide*. Rotterdam: EuroQol Foundation, Springer.
 26. Devlin, N., Tsuchiya, A., Buckingham, K., & Tilling, C. (2009). A uniform Time Trade Off method for states better and worse than dead: Feasibility study of the lead time approach. Economics Discussion Paper 09/08, City University. www.city.ac.uk/economics/dps/discussion_papers/0908.pdf.
 27. Sculpher, M., & Gafni, A. (2002). Recognising diversity in public preferences: The use of preference sub-groups in cost-effectiveness analysis. Authors Reply. *Health Economics*, *11*, 653.
 28. Robinson, A., & Parkin, D. (2002). Recognising diversity in public preferences: The use of preference sub-groups in cost-effectiveness analysis. A response to Sculpher and Gafni. *Health Economics*, *11*, 649–651.