

Reliability and validity of comprehensive health status measures in children: The Child Health Questionnaire in relation to the Health Utilities Index

Hein Raat^{a,*}, Gouke J. Bonsel^b, Marie-Louise Essink-Bot^a,
Jeanne M. Landgraf^c, Reinoud J.B.J. Gemke^d

^a*Department of Public Health, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands*

^b*Department of Social Medicine, Amsterdam Medical Center, Amsterdam, The Netherlands*

^c*HealthAct, Boston, MA, USA*

^d*Department of Pediatrics, Free University Medical Center, Amsterdam, The Netherlands*

Received 3 August 2000; received in revised form 11 May 2001; accepted 20 June 2001

Abstract

This study assesses the feasibility, reliability and validity of the Child Health Questionnaire-Parent Form (CHQ-PF50), consisting of 11 multi-item scales covering the physical, emotional and social well-being of children. The Health Utilities Index mark 2 (HUI2) was selected for comparison. Parents of 467 Dutch schoolchildren (age 5–13) were sent CHQ and HUI questionnaires. A subgroup of 79 parents of children aged 10–11 were sent a retest after 2 weeks. Feasibility: 78% response with few missing/non-unique CHQ-answers (<1%). Internal consistency: Cronbach's alphas of .39–.96 (mean .72). Item-own scale correlations were higher than item-other scale correlations. Test–retest ICCs were statistically significant for all but two CHQ scales (ICCs .31–.84). Test–retest CHQ-scale means did not show statistically significant differences except for one scale (“Behavior”). Validity: the CHQ-scales, with one exception, correlated better with predefined parallel HUI domains (ICCs .26–.53), than with non-parallel domains. Six CHQ scales discriminated clearly between children with and without chronic conditions and three scales discriminated between high and low medical consumption. This was at least equivalent to the discriminative ability of the HUI2 in this study. Additional studies of test–retest reliability and responsiveness to change of the CHQ in varied populations are needed. From the present study and literature data on other (clinical) populations we conclude that psychometric properties of CHQ-based health status measurement justify application in pediatric outcome studies, in addition to clinical measures. © 2002 Elsevier Science Inc. All rights reserved.

Keywords: Health status indicators; Quality of life; Validity; Reproducibility; Child Health Questionnaire (CHQ, CHQ-PF50); Health Utilities Index (HUI, HUI2)

1. Introduction

With the increase of chronic diseases in children, traditional medical indicators of outcome (e.g., survival, symptom status, blood chemistry) no longer suffice. The comprehensive assessment of health status or health-related quality of life, including physical, psychological and social domains, is required to complement clinical measures. Generic measures assess health status in a non-disease-specific way and claim to be applicable in a wide range of populations: in general population samples (e.g., for “burden of disease” descriptive studies and effect studies of community interventions) as well as in varied clinical samples [1–4].

For adults, several validated, multi-language questionnaires are available, such as the SF-36, the Sickness Impact Profile, Nottingham Health Profile, the COOP/WONCA Charts and the EQ-5D [1]. Health status measurement in children, however, is still at a developmental stage [2–10]. The Child Health Questionnaire (CHQ) has been proposed as a broadly applicable health status measure in children covering physical and psychosocial domains [6,11]. The CHQ utilizes the same structure and methodological approach as the SF-36. However, it was developed specifically for children and adolescents and therefore includes salient domains such as self-esteem and family functioning that are not found in tools for adults such as the SF-36. The CHQ has been applied in general populations as well as in studies with chronically ill children in several countries [6,11,12].

* Corresponding author. Tel.: +31 10 4087714; fax +31 10 4089449.
E-mail address: raat@mgz.fgg.eur.nl (H. Raat).

In the present study, the merits of the 50-item proxy version of the CHQ (CHQ-PF50) were assessed. The CHQ-PF50 is suitable for parents of children aged 4 years and older. The Health Utilities Index mark 2 (HUI2), was chosen to evaluate the concurrent validity of the CHQ, as it is a widely applied alternative pediatric health status measure which has been validated cross-culturally [13–15]. The aim of this study was to assess:

1. the feasibility of the Child Health Questionnaire (CHQ-PF50) as a proxy measurement (indicators: response rate, missing/non-unique answers);
2. the reliability of the CHQ-PF scales (internal consistency and test–retest reliability);
3. the validity of the CHQ-PF as judged by specific comparisons with the HUI2 (concurrent validity), and the ability to discriminate between groups with/without chronic conditions and with high/low medical consumption (discriminant validity).

2. Methods

2.1. Study population and data collection

The study population consisted of 467 schoolchildren in grades 3–8 (5–13 years old) at three representative elemen-

tary schools in Rotterdam, the Netherlands. The teachers of each class distributed the health questionnaires to the children, to be handed over by them to their parents (guardians). The children were required to return the forms within 2 weeks. No rewards or other response-increasing policies were applied. After 2 weeks this procedure was repeated in grade 7 (79 children age 10–11), in order to assess test–retest reliability.

Two criteria for eligibility for analysis were: (1) the parents' ability to read and write Dutch, and (2) at least an 80% response to CHQ-items. In the case of a missing answer, the answer to the same question of another, randomly chosen respondent was imputed to the data file for the purpose of the analyses. In the case of non-unique answers (more than one answer per question), a random selection of either answer was imputed during analysis as response.

2.2. Child Health Questionnaire (CHQ)

The CHQ-PF50 consisting of 50 items divided over 11 multi-item scales and 2 single-item questions (Table 1), was translated into Dutch according to international guidelines

Table 1
CHQ-PF50: scales, items per scale and score interpretation^a

Scale	Number of items	Description low score	Description high score
Physical functioning (PF)	6	Child is limited a lot in performing all physical activities, including self-care, due to health	Child performs all types of physical activities, including the most vigorous, without limitations due to health
Role functioning: Emotional/Behavior (REB)	3	Child is limited a lot in school work or activities with friends as a result of emotional or behavior problems	Child has no limitations in schoolwork or activities with friends as a result of emotional or behavior problems
Role functioning: Physical (RF)	2	Child is limited a lot in school work or activities with friends as a result of physical health	Child has no limitations in schoolwork or activities with friends as a result of physical health
Bodily pain (BP)	2	Child has extremely severe, frequent and limiting bodily pain	Child has no pain or limitations due to pain
General behavior (BE)	6	Child very often exhibits aggressive, immature, delinquent behavior	Child never exhibits aggressive, immature, delinquent behavior
Mental health (MH)	5	Child has feelings of anxiety and depression all of the time	Child feels peaceful, happy and calm all of the time
Self-esteem (SE)	6	Child is very dissatisfied with abilities, looks, family/peer relationships and life overall	Child is very satisfied with abilities, looks, family/peer relationships and life overall
General health perceptions (GH)	6	Parent believes child's health is poor and likely to get worse	Parent believes child's health is excellent and will continue to be so
Parental impact: Emotional (PE)	3	Parent experiences a great deal of emotional worry/concern as a result of child's physical and/or psychosocial health	Parent doesn't experience feelings of emotional worry/concern as a result of child's physical and/or psychosocial health
Parental impact: Time (PT)	3	Parent experiences a lot of limitations in time available for personal needs due to child's physical and/or psychosocial health	Parent doesn't experience limitations in time available for personal needs due to child's physical and/or psychosocial health
Family activities (FA)	6	The child's health very often limits and interrupts family activities or is a source of family tension	The child's health never limits or interrupts family activities nor is a source of family tension
Family cohesion (FC)	1	Family's ability to get along is rated "poor"	Family's ability to get along is rated "excellent"
Change in health (CH) ^b	1	Child's health is much worse now than 1 year ago	Child's health is much better now than 1 year ago

^aReproduced with permission [6].

^bThis concept was not fielded in this study.

[11,16,17]¹. Each CHQ scale consists of 3 to 6 items with 4, 5 or 6 response options per item. Following the CHQ standard procedure, the scale item scores (some recoded/recalibrated) were summed and transformed into 0 (worst possible health state) to 100 (best possible health state) scales [6]. “Physical” and “Psychosocial” CHQ summary scores were calculated following the CHQ standard procedure (a procedure based on a factor-analytical model of a U.S. child population sample), analogous to the construction of the summary scores in the SF-36 [18]. Given the underlying assumption of the CHQ developers that child health should be described by several child domains as well as the impact of child health on parental/family functioning, the CHQ-summary scores as proposed by the developers are based the “child-related” scales and the “parental/family-impact” scales (see discussion) [6]. A “Physical” or “Psychosocial” summary score of 50 represents the mean in the general population (here: U.S. children). Ten points above/below 50 reflects one standard deviation difference in either direction [6,18].

2.3. Health Utilities Index mark 2 (HUI2)

The HUI2 has six domains (attributes), each with four or five levels (Fig. 1) [7]². Contrary to the CHQ, the HUI2 does not consist of multi-item scales, but of an ordinal classification system. A HUI2 classification is either the direct result of expert judgement (doctor/nurse) or of a classification algorithm applied to the respondent’s (parent or doctor/nurse) answers on a standardized questionnaire. The Dutch questionnaire for parents was used to retrieve HUI2 data in this study [13]. Health states described by the HUI2 can be “valued” by representatives of a relevant population. These valuations can be used to estimate a multi-attribute utility function for the HUI2. By imputing individual HUI2 score profiles to the Canadian multiattribute utility function, utilities (range 0.0–1.0) indicating overall child health, were calculated. As suggested by Torrance et al. [7] HUI2 domain scores were transformed into “single-attribute utilities,” based on the same Canadian data, because these are more likely to have interval properties.

2.4. Other data

Standard socio-demographic data were collected. “Today’s prevalence” of 11 chronic conditions was measured

by the parent questionnaire. A child’s medical consumption over the past year was measured by asking about the number of visits to the family physician.

2.5. Analysis

Evaluation of feasibility consisted of the assessment of the percentage of missing/non-unique answers at the item level, and the content of written comments. Scale scores were described by conventional descriptive statistics. The CHQ scale-score distributions (as well as scale reliabilities; see below) in the sample of schoolchildren were compared with those in three available reference populations: general U.S. population age 5–18 ($n = 391$); mild-to-moderate asthmatic U.S. children ($n = 177$); U.S. children with attention deficit hyperactivity disorder (ADHD) ($n = 83$) [6]. Cronbach’s alpha was used to evaluate the internal consistency of the CHQ scales [19]. We assessed whether correlations (Pearson’s Product Moment coefficients) between the items and their own scale score (without the item under consideration) were higher than the correlations between these items and any other scale, to evaluate whether the CHQ items were well chosen and if the scales represent separate entities.

Test–retest reliability of the CHQ scales was assessed by test–retest correlations (Intraclass Correlation Coefficients, ICCs) [20], as well as at-group level, by comparison of the mean scores at first/second measurement with Mann-Whitney U-tests, given a non-normal distribution of the data.

2.5.1. Concurrent validity

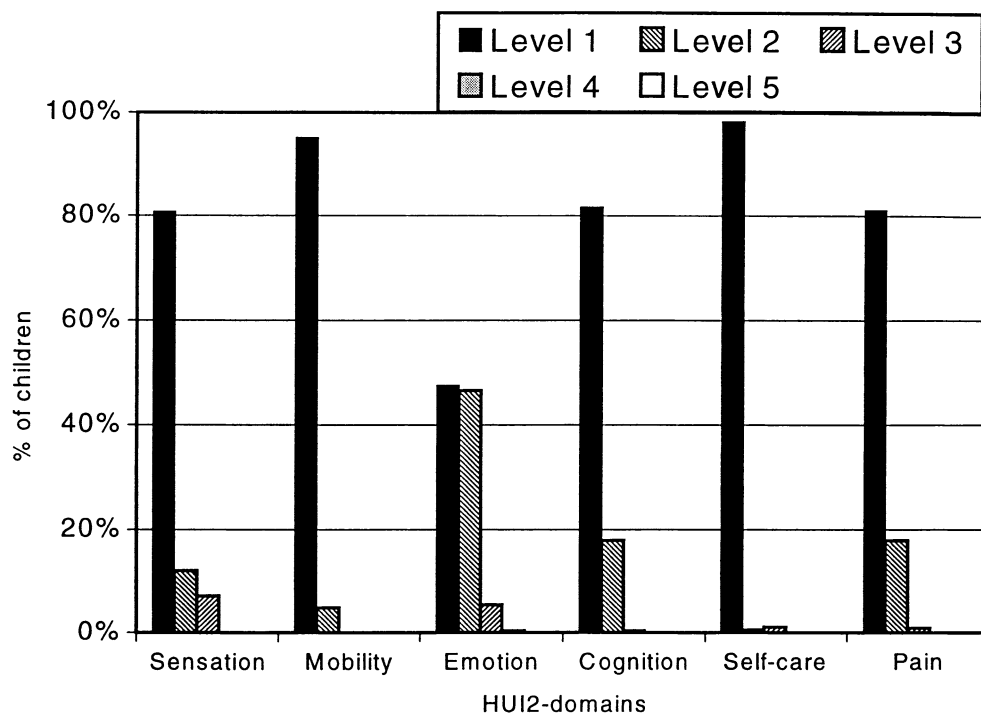
Concurrent validity was tested whether CHQ scales correlated better with their assumed parallel HUI2 domain (or, failing this, the multiattribute utility score) than with any other domain/score, as measured by Spearman’s rank correlation coefficients, considering non-normal distribution of the data. The following CHQ scale/HUI2 domain pairs were defined: “Physical functioning”–“Mobility”; “Role functioning-emotional/behavior”–“Emotion”; “Role functioning-physical”–“Pain”; “Bodily pain”–“Pain”; “General behavior”–“Emotion”; “Mental health”–“Emotion”; “Self-esteem”–“Emotion”.

2.5.2. Discriminant validity

The ability of the CHQ and HUI2 to discriminate between a group of children without any chronic condition and a group with at least two (parent-reported) chronic conditions was assessed. A similar assessment was subsequently made of a group of children with zero medical consumption versus one reporting at least three visits to a physician during the past year. One-sided Mann-Whitney U-tests were applied as it was hypothesized that mean scores would be lower in groups with reported chronic conditions/physician visits. Effect size estimations (d) were calculated which relate the difference in mean scores to the dispersion of the scores: $[\text{Mean}(a) - \text{Mean}(b)] / \text{standard}$

¹ Three independent forward translations were discussed in a consensus meeting with the translators, researchers and CHQ developer in order to decide on a single Dutch version. After double backward translation, the CHQ developer agreed upon the final Dutch version. In this study the full-length 94-item CHQ-PF was fielded, from which the CHQ-PF50 items were taken for analysis. The means of separate role-emotional and role-behavior items, as well as the means of two general-behavior items were used as an approximation of rewording in the standardized CHQ-PF50 [6]. The “Change-in-health” item was not fielded in this study.

² A seventh optional domain of HUI2 is “Fertility,” for specific purposes such as evaluation of juvenile cancer therapies. Complementary to the HUI2 a HUI3 exists with eight domains with five to six levels each [15].



DESCRIPTION OF HEALTH UTILITIES INDEX mark 2 (HUI2) LEVELS [7]		
Attribute	Level	Description
Sensation	1	Able to see, hear and speak normally for age
	2	Requires equipment to see or hear or speak
	3	Sees, hears or speaks with limitations even with equipment
	4	Blind, deaf or mute
Mobility	1	Able to walk, bend, lift, jump, and run normally for age
	2	Walks, bends, lifts, jumps, or runs with some limitations but does not require help
	3	Requires mechanical equipment (such as canes, crutches, braces, or wheelchair) to walk or get around independently
	4	Requires the help of another person to walk or get around and requires mechanical equipment as well
	5	Unable to control or use arms and legs
Emotion	1	Generally happy and free from worry
	2	Occasionally fretful, angry, irritable, anxious, depressed, or suffering "night terrors"
	3	Often fretful, angry, irritable, anxious, depressed, or suffering "night terrors"
	4	Almost always fretful, angry, irritable, anxious, or depressed
	5	Extremely fretful, angry, irritable, or depressed, usually requiring hospitalization or psychiatric institutional care
Cognition	1	Learns and remembers school work normally for age
	2	Learns and remembers school work more slowly than classmates as judged by parents and/or teachers
	3	Learns and remembers very slowly and usually requires special educational assistance
	4	Unable to learn or remember
Self-care	1	Eats, bathes, dresses, and uses the toilet normally for age
	2	Eats, bathes, dresses, and uses the toilet independently with difficulty
	3	Requires mechanical equipment to eat, bathe, dress, or use toilet independently
	4	Requires the help of another person to eat, bathe, dress, or use the toilet
Pain	1	Free of pain and discomfort
	2	Occasional pain. Discomfort relieved by nonprescription drugs or self-control activity without disruption of normal activities
	3	Frequent pain. Discomfort relieved by oral medicines with occasional disruption of normal activities
	4	Frequent pain; frequent disruption of normal activities. Discomfort requires prescription narcotics for relief
	5	Severe pain. Pain not relieved by drugs and constantly disrupts normal activities

Fig. 1. Frequencies of levels within HUI2-domains ($n = 353$).

Table 2

Score distributions and reliability of CHQ-PF50 scales in 4 child populations (Dutch schoolchildren $n = 353$; US general population children $n = 391^a$; mild-to-moderate asthmatic children $n = 178^a$; children with attention deficit hyperactive disorder (ADHD) $n = 83^a$)

CHQ-PF50 scales	Population	Mean (SD)	Range	% min ^b	% max ^c	25th %tile	50th ^d %tile	75th %tile	Cronbach's alpha ^e
Physical functioning	Dutch pupils	99.1 (4.3)	39–100	0	91	100	100	100	.70
	US-children	96.1 (13.9)	0–100	1	85	100	100	100	.94
	Asthma	85.5 (10.3)	27–100	0	24	78	89	94	.74
	ADHD	96.7 (12.2)	0–100	1	81	100	100	100	.60
Role fct.-Emo/Behav.	Dutch pupils	97.9 (7.2)	33–100	0	86	100	100	100	.87
	US-children	92.5 (18.6)	0–100	2	77	100	100	100	.88
	Asthma	91.3 (15.1)	22–100	0	65	89	100	100	.73
	ADHD	68.7 (30.0)	0–100	5	25	44	78	95	.86
Role funct.-Physical	Dutch pupils	95.8 (15.6)	0–100	1	91	100	100	100	.96
	US-children	93.6 (18.6)	0–100	2	85	100	100	100	.92
	Asthma	86.5 (19.4)	0–100	1	59	67	100	100	.80
	ADHD	96.8 (14.7)	16–100	0	95	100	100	100	f
Bodily pain	Dutch pupils	85.7 (17.2)	10–100	0	46	80	90	100	.86
	US-children	81.7 (19.0)	0–100	1	40	70	80	100	.89
	Asthma	75.5 (20.3)	10–100	0	26	60	80	100	.90
	ADHD	85.1 (19.4)	20–100	0	54	70	100	100	.81
General behavior	Dutch pupils	78.5 (13.1)	25–100	0	5	70	80	90	.64
	US-children	75.6 (16.7)	0–100	0	4	68	79	88	.81
	Asthma	72.4 (15.7)	20–100	0	2	65	75	85	.73
	ADHD	54.5 (17.3)	10–90	0	0	45	56	65	.72
Mental health	Dutch pupils	81.4 (12.1)	40–100	0	7	75	85	90	.68
	US-children	78.5 (13.2)	20–100	0	4	70	80	90	.75
	Asthma	78.1 (12.3)	30–100	0	3	75	80	85	.72
	ADHD	66.8 (15.6)	35–95	0	0	55	65	80	.80
Self-esteem	Dutch pupils	79.2 (11.0)	46–100	0	6	75	75	88	.77
	US-children	79.8 (17.5)	0–100	0	14	71	83	92	.84
	Asthma	82.4 (14.6)	33–100	0	18	75	83	96	.73
	ADHD	62.6 (19.5)	16–100	0	1	50	63	79	.79
General health	Dutch pupils	82.9 (13.4)	22–100	0	12	73	85	93	.39
	US-children	73.0 (17.3)	8–100	0	4	64	77	85	.66
	Asthma	56.3 (17.7)	0–100	1	1	44	56	68	.67
	ADHD	81.7 (19.0)	13–100	0	28	71	84	100	.64
Parental-Emotional	Dutch pupils	86.3 (15.2)	25–100	0	34	83	92	100	.59
	US-children	80.3 (19.1)	0–100	0	20	71	83	96	.70
	Asthma	71.1 (21.7)	4–100	0	7	58	75	88	.78
	ADHD	58.5 (18.8)	20–92	0	0	42	58	75	.56
Parental-Time	Dutch pupils	94.0 (13.0)	0–100	0	71	89	100	100	.69
	US-children	87.8 (19.9)	0–100	1	53	83	100	100	.80
	Asthma	81.6 (19.1)	22–100	0	31	72	89	100	.75
	ADHD	72.9 (21.8)	16–100	0	16	61	78	92	.69
Family activities	Dutch pupils	91.5 (11.9)	30–100	0	48	90	95	100	.74
	US-children	89.7 (18.6)	0–100	1	59	85	100	100	.93
	Asthma	—	—	—	—	—	—	—	—
	ADHD	62.1 (24.1)	0–100	2	2	50	67	80	.89
Family cohesion	Dutch pupils	72.2 (19.4)	30–100	0	16	60	60	85	na
	US-children	72.3 (21.6)	0–100	1	14	60	85	85	na
	Asthma	—	—	—	—	—	—	—	na
	ADHD	—	—	—	—	—	—	—	na
Physical summary ^g	Dutch pupils	56.4 (5.7)	29–65	na	na	55	58	60	.86
	US-children	53.0 (8.8)	—1–64	na	na	52	55	58	.93
	Asthma	45.7 (8.4)	19–61	na	na	41	48	52	.88
	ADHD	57.6 (6.2)	23–66	na	na	56	59	61	.87
Psychos. summary ^g	Dutch pupils	53.2 (6.4)	26–65	na	na	50	54	58	.87
	US-children	51.2 (9.1)	6–64	na	na	47	53	57	.93
	Asthma	51.2 (8.6)	23–66	na	na	47	53	58	.91
	ADHD	36.9 (10.9)	10–58	na	na	30	37	45	.92

^aReproduced with permission [6].

^{b,c} = % of respondents with worst/best possible score (ceiling/floor).

^d = median

^eaverage alpha of the 11 scales is .72.

na = not applicable, f = no variability across items, reliability could not be computed [6].

— = scale not fielded in this sample [6].

^gPhysical and Psychosocial CHQ summary scores based on a factor-analytical model on U.S. population samples. A score of 50 represents the mean in the general U.S. population; scores above/below 50 are above/below the average in the general U.S. population [6].

Table 3

Average inter-item, item-own scale and item-other scale correlations of CHQ-PF50 scales ($n = 353$)

CHQ-PF50 scales	Average inter-item correlation	Average item-own scale correlation ^a	Average item-other scale correlation
Physical functioning	.38	.50	.07
Role funct.-Emotional/behavior	.73	.77	.23
Role funct.-Physical	.93	.93	.14
Bodily pain	.75	.75	.20
General behavior	.24	.39	.15
Mental health	.30	.44	.21
Self-esteem	.37	.52	.18
General health	.14	.22	.13
Parental-Emotional	.32	.41	.28
Parental-Time	.42	.52	.25
Family activities	.39	.53	.25

^aEach item was correlated with the applicable scale excluding the item under consideration from the scale score.

deviation.³ Following Cohen's suggested guidelines, $d = .2$ was taken to indicate a small effect size, $d = .5$ a moderate effect size, and $d = .8$ a large effect size [21].

3. Results

3.1. Response and characteristics of parents and children

Of the 467 questionnaires distributed, 364 were returned (a 78% response). Thirteen were not eligible for analysis (language difficulty $n = 9$; incomplete $n = 4$). The retest in grade 7 ($n = 79$) had a response rate of 63%. In the absence of a first response, two retest questionnaires were treated as a first measurement. One retest questionnaire was incomplete.

The mean age of the parents was 37 years (range 27–56; SD 5); 88% were mothers; 9% were of non-Dutch ethnic origin. The educational background of the parents showed that 28% had completed higher vocational education or university and 43% had at least secondary education. Most parents (54%) were employed. The children ranged from 5 to 13 years of age (mean 8.8); 52% were girls. The age in grade 7 (retest population) ranged from 10 to 11.

3.2. Feasibility

Forty-four of 50 CHQ items had less than 1% missing answers. The highest percentage found was 2% ("Satisfaction with athletic ability"). Forty-five of 50 CHQ-items had less than 1% non-unique answers; the single-item scale "Family cohesion" had the highest percentage (4%). The scale "General health perceptions" evoked written comments such as "How should I know" (regarding expectations for the future). Nevertheless, these items showed few non-unique and missing answers.

³ Given unequal score variance between groups, the denominator used in calculating the d statistic was the square root of: $[(N_a - 1)S_a^2 + (N_b - 1)S_b^2]/[(N_a - 1) + (N_b - 1)]$.

Table 4

Test-retest reliability of the CHQ-PF50 in a subgroup children 10–11 years old ($n = 47$)

CHQ-PF scales	Test $n = 47$ X (SD)	Retest $n = 47$ X (SD)	P-value (MWU) ^a	Intra Class correlation Test-retest
Physical functioning	100 (1)	100 (1)	.56	-.03
Role funct.- Emotional/behavior	98 (5)	98 (4)	.51	.31*
Role funct.-Physical	96 (13)	98 (7)	.66	-.08
Bodily pain	81 (16)	85 (13)	.38	.40**
General behavior	79 (12)	85 (12)	.01	.65**
Mental health	80 (13)	83 (13)	.35	.84**
Self-esteem	77 (9)	79 (11)	.46	.53**
General health	85 (13)	86 (12)	.81	.76**
Parental-Emotional	85 (14)	90 (10)	.22	.63**
Parental-Time	94 (10)	96 (8)	.19	.51**
Family activities	92 (11)	95 (8)	.37	.49**
Family cohesion	68 (19)	71 (17)	.44	.75**
Physical summary ^b	57 (5)	57 (3)	.99	.27*
Psychosoc. summary ^b	53 (6)	55 (6)	.11	.79**

^aTest for differences between the average scale scores at the test and at the retest: two-sided Mann-Whitney U test.

^bPhysical and Psychosocial CHQ summary scores based on a factor-analytical model on U.S. population samples. A score of 50 represents the mean in the general U.S. population; scores above/below 50 are above/below the average in the general U.S. population [6].

* $P < .05$, ** $P < .01$.

3.3. Scoring distributions and internal consistencies

Mean CHQ scale scores in the physical and role-functioning domains were high and showed very skewed distributions, while "General behavior," "Mental health," "Self-esteem" and "General health perceptions" showed somewhat lower scores in the Dutch "schoolchildren" sample as well as the U.S. "general population" sample (Table 2). In the "Asthma" and "ADHD" samples mean scale scores were only convincingly lower in specific "affected" domains. The "Asthma" sample showed relatively low "General health perceptions" scores and a low "Physical summary" score. The "ADHD" sample showed relatively low scores on "Role functioning-emotional/behavior," "General behavior," "Mental health," "Self-esteem," "Parental impact-emotional," "Family activities," and the "Psychosocial summary" (Table 2). The Cronbach's alphas of the CHQ scales averaged .72 (.39–.96) in the Dutch sample, which compares to the Cronbach's alphas in all three (both clinical and non-clinical) reference populations. Only the "General health perceptions" scale had a very low .39 alpha in the Dutch sample (Table 2). The item-own scale correlations of the CHQ-scales were considerably higher than most of the corresponding item-other scale correlations (Table 3).

3.4. Test-retest reliability

Forty-seven test-retest pairs were analyzed. At the individual level, the test-retest correlations (ICCs) of the CHQ scales were statistically significant (.31–.84) for all but two scales (Table 4). Also, the test-retest CHQ "Psychosocial

Table 5

Concurrent validity of the CHQ-PF50 assessed by Spearman correlations between CHQ-PF50 scales/summary scores and predefined related (“parallel”) vs. unrelated (“non-parallel”) HUI2-domains/multiattribute utility ($n = 353$)^a

HUI2-domains ^b	Sensation	Mobility	Emotion	Cognition	Self-care	Pain	HUI2-multiattribute utility
CHQ-PF50 scales							
Physical Functioning	.07	.31**	.11*	.03	.12*	.19**	.20**
Role funct.-Emot./Behavior	.08	.19**	.28**	.29**	.07	−.06	.29**
Role funct.-Physical	.05	.12	.08	.06	.04	.42**	.18**
Bodily pain	.12*	.07	.24**	.02	.05	.26**	.29**
General behavior	.09	.04	.38**	.28**	.07	−.05	.34**
Mental health	.07	.07	.49**	.23**	.09	−.07	.42**
Self-esteem	.05	.15**	.36**	.20**	.10*	−.01	.34**
General health	.08	.08	.25**	.10	.00	.16**	.29**
Parental-Emotional	.17**	.13*	.43**	.31**	.08	.04	.48**
Parental-Time	.07	.18**	.31**	.31**	.18**	.08	.39**
Family activities	.12*	.06	.31**	.14**	.12*	.04	.34**
Physical summary ^c	.09	.17**	.08	.00	.01	.34**	.19**
Psychosoc. summary ^c	.11*	.10	.53**	.33**	.14**	−.11*	.49**

^aCorrelations with predefined related (‘parallel’) HUI2-domains (or HUI2-multiattribute utility) are in bold italics; other (spurious) correlations are in standard font.

* $P < .05$, ** $P < .01$.

^bHUI2-singleattribute utilities are applied, as well as the HUI2-multiattribute utility based on the Canadian utility formula’s [7].

^cPhysical and Psychosocial CHQ summary measures are based on a factor-analytical model on U.S. population samples [6].

summary” scores correlated well (.79), while the “Physical summary” scores did not (.27). The ICCs between first and second measurement showed no level effect. At group level, no significant difference emerged between the test- and re-test means of all but one (“General behavior”) of the scales, and the summary scores.

3.5. Concurrent validity

The distributions of HUI2 domain scores are presented in Fig. 1: near to perfect scores (level 1 = best possible health) were achieved in the domains “Mobility” and “Self-care,” whereas the lowest scores showed up in the domain “Emotion.” Of the parent-reported chronic conditions, the prevalence of allergies (11%), eczema (9%), asthma (5%) and “problems with hearing” (3%) were highest. Other conditions (vision problems unable to be corrected by glasses/contact lenses; diabetes; back problems; joint problems; seizures; migraine; depression) were rare (<2%).

All CHQ scales, except “Role functioning-emotional/behavior,” correlated better with the assumed parallel HUI2-domains (.26–.49), than with other domains (Table 5). Also, the CHQ summary scores correlated with appropriate HUI2-domains.

3.6. Discriminant validity

Table 6 shows the intermediate-to-strong ability of six CHQ scales and the “CHQ-Physical summary score” to discriminate between the subgroup of children without any parent-reported chronic condition and those with two or more conditions. Four CHQ scales and the “CHQ-Physical summary score” discriminated well between a subgroup of children who had not visited the family physician at all last year and a subgroup consulting a physician at least three

times. This was at least equivalent to the discriminative ability of HUI2 in this study.

4. Discussion

In our population of predominantly healthy schoolchildren some aspects of the performance of the CHQ were excellent (feasibility), others were good (concurrent and discriminant validity), and still others need further study (test–retest reliability). The performance of the CHQ in this study was at least equivalent to that of the HUI2.

The CHQ was well accepted by parents under non-supported conditions, as judged by the high response and few missing/non-unique answers. The appreciation of the CHQ and the HUI2 by respondents with regard to aspects such as readability and perceived relevance has not been studied in this project. Evaluation of the respondents’ perspective might be recommended for future studies.

A shorter CHQ-PF form than the one applied in this study is currently available, notably a 28-item form [6,11,12]. In general, abridged versions, which omit the less informative items, have been found not to limit group comparisons [22]. However, decreased sensitivity may play a role in individual use, repeated-measures designs and where assessments of rather specific conditions are concerned.

A unique feature of the CHQ, which distinguishes it from the SF-36 and from other pediatric questionnaires, are the three family impact scales. They showed good psychometric properties, especially in the ability to discriminate between the presence/absence of chronic conditions. This extension of the health concept is appropriate from a clinical and social perspective. In evaluative applications, these scores must be handled with care. In particular, if changes in the “family impact” domains would be incongruent with

Table 6

Discriminant ability of the CHQ-PF50 and the HUI2 between groups differing in the number of parent-reported chronic conditions of the child, and the number of visits of the child to the family physician last year

	Number of chronic conditions per child				Number of visits to physician last year			
	0 <i>n</i> = 251 X (SD)	≥2 <i>n</i> = 25 X (SD)	<i>P</i> -value (MWU) ^a	Effect size <i>d</i> ^b	0 <i>n</i> = 149 X (SD)	≥3 <i>n</i> = 49 X (SD)	<i>P</i> -value (MWU) ^a	Effect size <i>d</i> ^b
CHQ-PF50 scales (range 0–100)								
Physical functioning	99 (4)	98 (5)	<.01	.41	99 (5)	98 (5)	<.05	.18
Role funct.-Emo/Behav.	98 (7)	98 (5)	.08	.10	98 (7)	98 (8)	.45	.03
Role funct.-Physical	97 (13)	89 (28)	<.01	.56*	97 (15)	87 (26)	<.01	.51
Bodily pain	88 (15)	75 (19)	<.01	.84**	89 (15)	74 (23)	<.01	.81**
General behavior	80 (12)	77 (15)	.21	.25	78 (13)	77 (13)	.18	.12
Mental health	82 (12)	77 (15)	<.05	.41	82 (11)	78 (13)	<.05	.34
Self-esteem	80 (11)	75 (11)	<.05	.46	79 (11)	78 (11)	.14	.17
General health	85 (12)	70 (17)	<.01	1.24**	85 (13)	76 (15)	<.01	.68*
Parental-Emotional	89 (13)	75 (20)	<.01	.98**	88 (15)	80 (16)	<.01	.54*
Parental-Time	95 (11)	88 (15)	<.01	.66*	94 (14)	91 (14)	<.05	.26
Family activities	93 (10)	82 (17)	<.01	1.05**	93 (11)	86 (16)	<.05	.53*
Physical summary ^c	57 (5)	51 (9)	<.01	1.12**	57 (5)	52 (8)	<.01	.83**
Psychosoc. summary ^c	54 (6)	51 (8)	<.05	.48	53 (6)	52 (7)	.18	.20
HUI2-dimensions (range 0–1) ^d								
Sensation	.97 (.08)	.91 (.13)	<.01	.70	.96 (.10)	.95 (.11)	.17	.13
Mobility	1.00 (.01)	1.00 (.02)	.42	.04	1.00 (.02)	.99 (.02)	.18	.15
Emotion	.92 (.10)	.85 (.14)	<.01	.69*	.91 (.11)	.91 (.09)	.23	.04
Cognition	.98 (.06)	.98 (.05)	na	na	.97 (.06)	.97 (.06)	.27	.08
Self care	.99 (.09)	.96 (.20)	.13	.30	.99 (.08)	1.00 (.00)	na	na
Pain	.99 (.02)	.97 (.07)	<.01	.95**	1.00 (.02)	.98 (.04)	<.01	.62*
Multiaattribute utility	.93 (.07)	.86 (.12)	<.01	1.03**	.92 (.08)	.91 (.07)	<.05	.23

^aOne-sided Mann-Whitney U-test as it is hypothesised that mean values are lower in groups with conditions/physician visits.

na = not applicable, as mean (b) ≥ mean (a).

^bInterpretation: *d* = .2: small effect; *d* = .5: medium effect; *d* = .8 large effect [21].

^cPhysical and Psychosocial CHQ summary scores based on a factor-analytical model on U.S. population samples. A score of 50 represents the mean in the general U.S. population; scores above/below 50 are above/below the average in the general U.S. population [6].

^dHUI2-singleattribute utilities are applied, as well as the HUI2-multiaattribute utility based on the Canadian utility formulas [7].

*.5 < *d* < 8; ***d* > 8.

changes in the domains that directly concern the functioning of the child him/herself, it may be difficult to judge which outcomes are the most relevant in a specific study.

To facilitate comparisons of CHQ outcomes between groups, “Physical” and “Psychosocial” CHQ summary measures have been developed based on a factor-analytical model in a reference population. The psychometric properties of these CHQ summary measures were satisfactory, adequately reflecting the underlying scales within the limitations of this study. The current CHQ summary formulas, based on a U.S. study only, should be checked in international reference populations [18]. For descriptive purposes the current CHQ summary scores incorporating child as well as parent/family effects can be defended, but if CHQ summary scores are intended for evaluations of medical decisions, the breakdown in “child components” and “parent/family components” seems mandatory.

In this population considerable ceiling effects were present, notably on the “Physical functioning” and “Role functioning” scales. This phenomenon is inevitable—in the HUI it is equally apparent—but limits, however, the use of these instruments to detect changes in a generally healthy

population, or to describe excellent health beyond the average. The psychosocial scales produced less skewed data, as expected [8,9,11,13]. In specific clinical populations (asthma and ADHD in Table 2, as well as in other published studies), CHQ scale scores were lower only in those specific domains that can be expected to be affected by the specific condition. This finding can be interpreted as support for the validity of the CHQ [6,15,23].

Our study, to our knowledge for the first time, evaluated test-retest reliability of the CHQ. At group level, average scale scores were similar on two occasions, but only 8 of 14 of the scales/summary measures showed test-retest ICCs above .50. The rather small sample size of the retest, and, in particular, the obvious ceiling effects in some scales go a long way toward explaining this finding. The moderate test-retest performance of “Bodily pain” (ICC = .40) may in part be attributable to additional factors, as is also the case in its adult counterpart in the SF-36 [24]. Here, “real” variation may be indistinguishable from measurement error. The present study did not assess “responsiveness to change” of the CHQ, which would require a repeated-measures design in more varied samples [25]. Future studies should cover

this issue, preferably after reconfirmation of conventional retest characteristics.

4.1. Concurrent and discriminant validity

Direct comparison of the CHQ scales with relevant single HUI2 domains or the multiattribute utility value supported the validity for the large majority of scales. Higher correspondence in this population should not be expected, as content and structure of both instruments show distinct differences. Similar tests in defined patient groups should determine whether validity can be reproduced for the full range of scale scores.

The CHQ scale “General health perceptions”—although showing the lowest Cronbach’s alpha and evoking many comments from respondents—and the “Physical summary measure” in particular discriminated very well between groups with and without parent-reported chronic conditions. The CHQ was at least equal to the HUI2 in this respect. Medical consumption is a difficult external anchor in this regard, as factors other than health status may also determine utilization. This explains why, in this study, the presence/absence of chronic conditions was better able to be discriminated than high/low medical consumption by health status measures (both CHQ and HUI2). The further establishment of validity of the CHQ will require a similar field study among defined patient groups.

4.2. Generalizability

The application of the CHQ in younger age groups (below 4 years of age) requires adaptation rather than leaving out non-relevant items. The CHQ-Toddler Form (2 months–4 years of age) is currently being developed [12]. From about 10 years of age, the self-report CHQ-Child Form⁴ is indicated, whenever possible combined with the Parent Form [6,26,27]. Another relevant issue is the use of the CHQ outside the country of origin. Procedures for translation, adaptation and subsequent validation of instruments such as the HUI, the SF36 and the CHQ are available and indeed have been applied in this and other studies [11,12,16,17]. The coverage of the CHQ is currently being expanded to 21 languages in 32 countries, which will facilitate international projects [28]. Our application in the Netherlands yielded no socio-cultural difficulties.

4.3. Choice of instruments

Both CHQ and HUI2 may be regarded as suitable generic health status measures. Depending on the study objectives either of the two may be chosen, or where possible, the complementary use of both may be opted for. The HUI2, as a classification system is less detailed than the CHQ, focuses on functional/clinical aspects of health, but can be used for economic evaluation [7]. Moreover the HUI2 can

be applied in pediatric as well as adult populations, whereas the CHQ should be complemented by the SF-36 for the adult component of study populations. On the other hand, the CHQ provides a feasible and detailed child health profile balanced for physical and psychosocial aspects of health and impact of ill health on family life. The CHQ has as specific advantage its availability in many languages and similarity to the scale structure of the SF-36. Before reaching a final judgement, additional studies are needed to assess test–retest stability as well as the responsiveness of the CHQ to change in varied samples. From this study and the presented data from other (clinical) studies, we conclude, at this stage, that psychometric properties of CHQ-based health status assessment satisfy the requirements for application in clinical as well as public health outcome research, in addition to disease-specific health status measures and clinical measures.

Acknowledgments

This project was supported from the beginning by prof. Ferd Sturmans, M.D., Ph.D., director of the Rotterdam Municipal Health Service in his endeavor to stimulate ongoing outcome research in health care. Prof. Paul van der Maas, M.D., Ph.D. and Prof. Johan Mackenbach, M.D., Ph.D., of the Department of Public Health of the Erasmus University rendered assistance with stimulating ideas as well as research facilities. We are grateful to the school physicians, nurses and doctor’s assistants of the department of Youth Care of the Rotterdam Municipal Health Service, especially Fifi Pouw, M.D., for facilitating this project in collaboration with related schools. We thank Rieneke Dekker, M.Sc., Ingrid Walda, M.Sc., Wilma Jansen, M.Sc. and Jan Bleeker, Ph.D. for help with the organization of this project, translation of the CHQ, expert data collection and statistical support.

References

- [1] Essink-Bot ML, Krabbe PF, Bonsel GJ, Aaronson NK. An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. *Med Care* 1997;35(5):522–37.
- [2] Pal DK. Quality of life assessment in children: a review of conceptual and methodological issues in multidimensional health status measures. *J Epidemiol Commun Health* 1996;50:391–6.
- [3] Jenney ME, Campbell S. Measuring quality of life. *Arch Dis Child* 1997;77(4):347–50.
- [4] Morrow RH, Hyder AA, Murray CJ, Lopez AD. Measuring the burden of disease. *Lancet* 1998;352(9143):1859–61.
- [5] Stein RE, Jessop DJ. Functional status II(R). A measure of child health status. *Med Care* 1990;28(11):1041–55.
- [6] Landgraf JM, Abetz L, Ware JE. The CHQ user’s manual. Boston: The Health Institute, New England Medical Center, 1996.
- [7] Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care* 1996;34(7):702–22.
- [8] Ravens-Sieberer U, Bullinger M. Assessing health-related quality of

⁴ The CHQ-CF has been assessed in a project parallel to the current study (manuscript in preparation).

- life in chronically ill children with the German KINDL: first psychometric and content analytical results. *Qual Life Res* 1998;7(5):399–407.
- [9] Vogels T, Verrips GH, Verloove-Vanhorick SP, Fekkes M, Kamphuis RP, Koopman HM, et al. Measuring health-related quality of life in children: the development of the TACQOL parent form. *Qual Life Res* 1998;7(5):457–65.
 - [10] Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the pediatric quality of life inventory. *Med Care* 1999;37(2):126–39.
 - [11] Landgraf JM, Maunsell E, Speechley KN, Bullinger M, Campbell S, Abetz L, et al. Canadian-French, German and UK versions of the Child Health Questionnaire: methodology and preliminary item scaling results. *Qual Life Res* 1998;7(5):433–45.
 - [12] Landgraf JM. Measuring pediatric outcomes in applied clinical settings: an update about the Child Health Questionnaire (CHQ). *Qual Life Newsletter MAPI Research Institute* 1999;23:5–6.
 - [13] Gemke RJ, Bonsel GJ. Reliability and validity of a comprehensive health status measure in a heterogeneous population of children admitted to intensive care. *J Clin Epidemiol* 1996;49(3):327–33.
 - [14] Trudel JG, Rivard M, Dobkin PL, Leclerc JM, Robaey P. Psychometric properties of the Health Utilities Index Mark 2 system in paediatric oncology patients. *Qual Life Res* 1998;7(5):421–32.
 - [15] Nixon Speechley K, Maunsell E, Desmeules M, Schanzer D, Landgraf JM, Feeny DH, et al. Mutual concurrent validity of the child health questionnaire and the health utilities index: an exploratory analysis using survivors of childhood cancer. *Int J Cancer Suppl* 1999;12:95–105.
 - [16] Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46(12):1417–32.
 - [17] Ware JE, Keller SD, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health status questionnaires. Methods from the IQOLA project. *International Quality of Life Assessment. Int J Technol Assess Health Care* 1995;11(3):525–51.
 - [18] Ware JE, Gandek B, Kosinski M, Aaronson NK, Apolone G, Brazier J, et al. The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. *International Quality of Life Assessment. J Clin Epidemiol* 1998;51(11):1167–70.
 - [19] Bland JM, Altman DG. Cronbach's alpha. *BMJ* 1997;314(7080):572.
 - [20] Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials* 1991;12(Suppl 4):142S–58.
 - [21] Cohen J. *Statistical power analysis for the behavioral sciences*. New York: Academic Press, 1977.
 - [22] Coste J, Guillemin F, Pouchot J, Fermanian J. Methodological approaches to shortening composite measurement scales. *J Clin Epidemiol* 1997;50(3):247–52.
 - [23] Kurtin PS, Landgraf JM, Abetz L. Patient-based health status measurements in pediatric dialysis: expanding the assessment of outcome. *Am J Kidney Dis* 1994;24(2):376–82.
 - [24] Dougherty CM, Dewhurst T, Nichol WP, Spertus J. Comparison of three quality of life instruments in stable angina pectoris: Seattle Angina Questionnaire, Short Form Health Survey (SF-36), and Quality of Life Index-Cardiac Version III. *J Clin Epidemiol* 1998;51(7):569–75.
 - [25] Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol* 1997;50(3):239–46.
 - [26] Landgraf JM, Abetz LN. Measuring health outcomes in pediatric populations: issues in psychometrics and application. In: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. Philadelphia/New York: Lippincott-Raven Publishers, 1996. p. 793–802.
 - [27] Theunissen NC, Vogels TG, Koopman HM, Verrips GH, Zwinderman KA, Verloove-Vanhorick SP, et al. The proxy problem: child report versus parent report in health-related quality of life research. *Qual Life Res* 1998;7(5):387–97.
 - [28] Ruperto N, Martini A. A European network for randomised actively controlled clinical trials in paediatric rheumatic diseases: parenteral methotrexate in medium versus higher doses in juvenile chronic arthritis. In: XIV European League Against Rheumatism Congress “Rheumatic Diseases—from Genes to Therapies”; 1999; Scotland, UK, 6–11 June 1999; 1999.