VALUE IN HEALTH 18 (2015) 334–345

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

A Systematic Review of Generic Multidimensional Patient-Reported Outcome Measures for Children, Part II: Evaluation of Psychometric Performance of English-Language Versions in a General Population

Astrid Janssens, PhD[1,], Morwenna Rogers, MSc[1], Jo Thompson Coon, PhD[1], Karen Allen, MSc[1], Colin Green, PhD[1], Crispin Jenkinson, DPhil[2], Alan Tennant, PhD[3], Stuart Logan, MSc, F.R.C.P.C.H.[1], Christopher Morris, DPhil[1]*

[1]*University of Exeter Medical School, University of Exeter, Exeter, UK;* [2]*Nuffield Department of Population Health, University of Oxford, Oxford, UK;* [3]*Department of Rehabilitation Medicine, University of Leeds, Leeds, UK*

ABSTRACT

Objectives: The objectives of this systematic review were 1) to identify studies that assess the psychometric performance of the English-language version of 35 generic multidimensional patient-reported outcome measures (PROMs) for children and young people in general populations and evaluate their quality and 2) to summarize the psychometric properties of each PROM. **Methods:** MEDLINE, EMBASE, and PsycINFO were searched. The methodological quality of the articles was assessed using the COnsensus-based Standards for selection of health Measurement INstruments checklist. For each PROM, extracted evidence of content validity, construct validity, internal consistency, test-retest reliability, proxy reliability, responsiveness, and precision was judged against standardized reference criteria. **Results:** We found no evidence for 14 PROMs. For the remaining 21 PROMs, 90 studies were identified. The methodological quality of most studies was fair. Quality was generally rated higher in more recent studies. Not reporting how missing data were handled was the most common reason for downgrading the quality. None of the 21 PROMs has had all psychometric properties evaluated; data on construct validity and internal consistency were most frequently reported. **Conclusions:** Overall, consistent positive findings for at least five psychometric properties were found for Child Health and Illness Profile, Healthy Pathways, KIDSCREEN, and Multi-dimensional Student Life Satisfaction Scale. None of the PROMs had been evaluated for responsiveness to detect change in general populations. Further well-designed studies with transparent reporting of methods and results are required.

Keywords: children and young people, measurement properties, patient-reported outcomes, review.

Introduction

Patient-reported outcome measures (PROMs in the United Kingdom and patient-reported outcomes in the United States) are increasingly advocated for use in clinical trials [1,2] and as key performance indicators for evaluating health systems [3]. PROMs can be domain-specific, and focus on particular aspects of health (e.g., mental health or physical functioning), or be multidimensional instruments with subscales that assess different aspects of health. Some PROMs are condition-specific, designed for use by people with a particular diagnosis; other PROMs are generic and appropriate for anyone to report their health. Generic PROMs can be used across people with a range of health conditions, which is particularly useful when no condition-specific measure is available, or when comparisons are made between the health of subgroups of people and findings from general population surveys [4].

When selecting PROMs for a specific purpose, it is necessary to examine both what is being assessed and how robust (valid and reliable) is the measurement. Language and cultural issues can affect how people interpret and/or respond to questions; hence, one cannot simply assume that PROMs perform consistently across languages and cultures [5,6]. Therefore, for example, the Food and Drug Administration guidance on PROMs recommends that evidence be provided of the process used to test measurement properties across different languages and cultures [1].

This article reports the results of a systematic review and critical evaluation of the literature on the measurement properties of PROMs for children and young people up to 18 years old. We focused on evaluations of English-language versions of

Conflicts of interest: The authors have no conflicts of interest to disclose.

* *Address correspondence to:* Astrid Janssens, PenCRU, Institute of Health Research, University of Exeter, Salmon Pool Lane, EX24SG Exeter, UK.

E-mail: a.janssens@exeter.ac.uk.

1098-3015$36.00 – see front matter Copyright © 2015, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2015.01.004

generic multidimensional PROMs for children to take account of methodological developments and any evidence published since previous reviews [7–9]. A new quality evaluation tool, the COnsensus-based Standards for the Selection of health status Measurement INstruments (COSMIN) system, has been developed to standardize the assessment of methodological quality of measurement studies [10–12]. In a related article, we have documented a systematic search and descriptive review of generic multidimensional PROMs for children, identifying 35 PROMs. In this study, we sought to identify and critically appraise studies that have assessed the psychometric performance of these PROMs, and to describe available evidence for the psychometric properties of each PROM.

Methods

Search Strategy

A separate search strategy was created for each of the 35 PROMs. MEDLINE, EMBASE, and PsycINFO were searched (via OvidSP) between July 18 and September 5, 2012, using three groups of terms: 1) name(s) and standard acronym of the PROM, 2) terms to describe children and young people, and 3) psychometric terms. No language or date limits were applied to the search. An illustration of the search strategy as used in EMBASE for one PROM (EuroQol 5D Youth [EQ-5D-Y]) can be seen in Data 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2015.01.004. Individual search strategies for the remaining PROMs can be supplied on request.

Backwards citation chasing (one generation) was carried out using all reference lists from articles included in the review. Forward citation chasing was carried out between January 28 and February 6, 2013, using Science Citation Index and Social Science Citation Index (via Web of Knowledge) for the key reference(s) of each of the selected PROMs. Developers of PROMs for which no published peer-reviewed articles were found were contacted to verify that we had not missed any eligible articles.

Inclusion and Exclusion Criteria

Articles were selected when written in English and reporting on a study that 1) was specifically designed to evaluate the psychometric properties of a selected PROM using an English-language version of the questionnaire, 2) was conducted in a general population of children up to 18 years old, and 3) published in a peer-reviewed journal. Articles were excluded if 1) the PROM was used as a criterion standard to test another instrument, 2) less than 10% of the study population was younger than 18 years, and 3) the study targeted children and young people with a specific condition or illness.

Study Selection

Titles and abstracts of records were screened against the eligibility criteria by one reviewer (A.J.); 10% were checked by a second reviewer (C.M.), with disagreements resolved by discussion with a third (C.J.) where necessary. The full text of any potentially relevant article was retrieved and screened using the same procedure.

Assessment of Methodological Quality of Included Articles

For each article, the methodological quality of the study and the completeness of the report were assessed using the COSMIN checklist (Table 1) [12]. This checklist consists of nine boxes with methodological standards for how each measurement property should be assessed [13]. Each item is rated on a four-point scale (poor, fair, good, or excellent); an overall score for each methodological quality is determined by a "worst-score counts" procedure.

Table 1 – Appraisal of psychometric properties and indicative criteria.

Psychometric property	Indicative criteria
Content validity	Clear conceptual framework consistent with stated purpose of measurement Qualitative research with potential respondents
Construct validity	Structural validity from factor analysis Post hoc tests of unidimensionality by Rasch analysis Hypothesis testing, with a priori hypotheses about direction and magnitude of expected effect sizes Tests for differential item and scale functioning between sex, age groups, and different diagnoses
Reproducibility	Test-retest reliability: ICC >0.7 adequate, >0.9 excellent. Proxy reliability: Child and parent-reported reliability ICC > 0.7
Internal consistency	Cronbach α coefficient >0.7 and <0.9
Precision	Assessment of measurement error; floor or ceiling effects <15%; evidence provided by Rasch analysis and/or interval-level scaling
Responsiveness	Longitudinal data about change in scores with reference to hypotheses, measurement error, minimal important difference

ICC, intraclass correlation coefficient.

The checklist was administered by one reviewer (C.M./A.T.), and a 10% sample was rated by a second (A.J./C.M.). Any discrepancies were resolved by discussion, or with the involvement of a third reviewer (C.J.), where necessary.

Data Extraction

For each article describing a study evaluating the psychometric performance of an eligible PROM, the following descriptive data were extracted: instrument version, first author name, publication year, study aim, study population, number of participants, age range, mean age, and setting or country where the study was conducted. Data were extracted by one reviewer (K.A.), and 50% were checked by a second (A.J.), with disagreements resolved by discussion with a third (C.M.), where necessary.

For each version of a PROM, evidence of the following psychometric properties was extracted: content validity (theoretical framework and/or qualitative research), construct validity (structural validity and hypothesis testing), internal consistency, test-retest reliability, proxy reliability, responsiveness, and precision. Data were extracted by one reviewer (K.A./A.J./A.T.) and checked by a second (A.J./K.A./A.J.), with disagreements resolved by discussion with a third (C.M.), where necessary.

Appraisal and Summary of Evidence for Psychometric Performance

Evidence of performance was summarized by psychometric property and judged using standardized reference criteria and thresholds (Table 1). We included an appraisal of validity, reliability, responsiveness, and precision [4]. These data were

summarized in a single rating for each measurement property following methods commonly used for the presentation of findings against the COSMIN criteria [14,15]. Our summary judgment took into account the following elements: 1) data extracted from included studies, with reference to standard criteria; 2) the methodological quality of studies and number of studies; and 3) the thoroughness of testing, giving further weight to any studies that appeared not to have been conducted by the original developers (Table 2) [16]. Two reviewers (A.J./C.M.) made the judgment through discussion based on available evidence.

Results

In the following text, we use the word PROM to refer to the group of questionnaires (different versions according to age group, length, or responder) of a certain instrument; we use the word questionnaire to refer to a specific version of an instrument (Table 3).

Thirty-five PROMs were identified for children and young people, as previously described. Here, the combined search strategies for these 35 PROMs resulted in 2750 records after duplicates were removed (Fig. 1). From these searches we included 77 articles in this review. Reference tracking resulted in the identification of 13 additional articles. For 14 PROMs, that is taking into account all versions, we were unable to identify a published study examining a psychometric property in an English-speaking general population: Auto Questionnaire Enfant Image - Child Pictured Self-Report, Child Quality of Life Questionnaire, Duke Health Profile - Adolescent version, Health and Life Functioning Scale, How Are You, Infant Toddler Quality of Life Questionnaire, Nordic Quality of Life Questionnaire for children, Quality of Life Questionnaire for Adolescents/Quality of Life Questionnaire for Adolescents (Taiwanese version), Quality of Life Questionnaire for Children, TNO-AZL questionnaires, Vécu et Santé Perçue de l'Adolescent, 16 Dimensional, Assessment of Quality of Life Mark 2, 6D adolescents, and Comprehensive Health Status Classification System – Preschool [17–30]. The 14 PROMs, accounting for 22 questionnaires, for which no evidence was found are excluded from the Results section. Four PROMs had evidence lacking of one or more versions: Child Health Questionnaire (CHQ) Parent Short Form; Functional Status II Revised (FSIIR), Long version, infants; FSIIR, Long version, toddlers; FSIIR, Long version, pre-scholars; FSIIR, Long version, school-age children; and FSIIR 7- item version; Kiddy-KINDL; and Personal Wellbeing Index School Children.

We present a description of the study population of the 90 studies reporting evidence on the psychometric performance of the 21 remaining PROMs (see Supplementary Data 2 in Supplemental Material found at http://dx.doi.org/10.1016/j.jval.2015.01.004). The methodological quality of the studies was found to be variable, but appears to have improved with more recent studies (see Data 3 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2015.01.004). Not reporting how missing data were handled was the most common reason of methodological weakness.

Table 2 – Indices for summarizing appraising psychometric properties of patient-reported outcome measures.

Rating	Definition	Description
0	Not reported	No studies found that evaluate this measurement property
?	Not clearly determined	Studies were rated poor methodological quality; results not considered robust
–	Evidence not in favor	Studies were rated good or excellent methodological quality; results did not meet standard criteria for this property
+/–	Conflicting evidence	Studies were rated fair, good, or excellent methodological quality; results did not consistently meet standard criteria for this property, e.g., not for all domain scales
+	Some evidence in favor	Studies were rated fair or good methodological quality; standard criteria were met for the property
++	Some good evidence in favor	Studies were rated good or excellent methodological quality; standard criteria were met or exceeded
+++	Good evidence in favor	Studies were rated good or excellent methodological quality; standard criteria were exceeded, results have been replicated

Measurement Properties for Each PROM

A summary appraisal of the evidence of the psychometric performance of each generic PROM in a general population is given in Table 4.

Content Validity

Stronger evidence for content validity is available for the questionnaires Child Health and Illness Profile- Child Edition (CHIP-CE) [31], Child Health and Illness Profile-Adolescent Edition (CHIP-AE) [32], KIDSCREEN-52 [33], and Child Health Utility 9D (CHU-9D) [34], with extensive qualitative research having been used to generate the items. For the latter, no information supporting content validity was found in an adolescent population, even though use of the instrument has been generalized to this older age group. The studies reporting on the content validity of Exeter Quality of Life Measure and FSIIR 14-item version are unclear whether and how children have been involved in the item development [35,36]. For KIDSCREEN and Multi-dimensional Student Life Satisfaction Scale (MSLSS), content validity was assessed for the development of one version only, KIDSCREEN-52 and MSLSS, respectively; content validity was not reassessed for consecutive versions [37–40]. Seven PROMs (accounting for 12 questionnaires) had no studies reporting on content validity: Child Health Assessment Questionnaire, CHQ (Child Health Questionnaire Parent Long Form [CHQ-CF50] and Child Health Questionnaire Short Form 80/87 version [CHQ-PF80/87]), ComQOL (ComQOL and Personal Wellbeing Index School Children), EQ-5D-Y, Health Utilities Index (HUI2 and HUI3), KINDL (KINDL-Kid and KINDL-Kiddo), and Pediatric Quality Of Life Inventory Trade Mark 4.0 (PedsQL) (Pediatric Quality Of Life Inventory Trade Mark 4.0 Generic Core Scales [PedsQL 4.0] and Pediatric Quality Of Life Inventory Trade Mark 4.0 Short Form 15 [PedsQL-SF15]) [41–47]. Content validity of KINDL was not studied for the English version; evidence is available for the original (German) version [48]. The PedsQL 4.0 is one of the most extensively studied PROMs in the list. Item generation and reduction is described for the Pediatric Cancer Quality of Life Inventory, from which the PedsQL 4.0 is derived [49,50], but not for the PedsQL 4.0.

Table 3 – Index table of PROMs and questionnaires: acronyms, names, and reference citations.

PROM group and citation number	Questionnaire acronym	Questionnaire full name	Author (year)	Questionnaire citation number
AUQUEI* [19]	QUALIN	Infant's quality of life	Manificat et al. (1999)	[99]
	AUQUEI Ours	Auto Questionnaire Enfant Imagé - Child Pictured Self-Report	Manificat and Dazord (1998)	[19]
	AUQUEI Soleil	Auto Questionnaire Enfant Image - Child Pictured Self-Report	Gayral-Taminh et al. (2005)	[100]
	OK.ado questionnaire	Adolescent quality of life questionnaire	Manificat and Dazord (2002)	[101]
CHAQ [41]	CHAQ	Child Health Assessment Questionnaire	Singh et al. (1994)	[41]
CHIP [31]	CHIP-CE CRF	Child Health and Illness Profile-Child Edition Child-report form	Riley et al. (2004)	[74]
	CHIP-CE PRF45	Child Health and Illness Profile-Child Edition Parent-report Form 45	Riley et al. (2004)	[31]
	CHIP-CE PRF76	Child Health and Illness Profile-Child Edition Parent-report Form 76	Riley et al. (2004)	[31]
	CHIP-AE	Child Health and Illness Profile-Adolescent Edition	Riley et al. (1998)	[32]
CHQ [42]	CHQ-PF28	Child Health Questionnaire Parent Short Form	Kurtin et al. (1994)	[102]
	CHQ-PF50	Child Health Questionnaire Parent Long Form	Landgraf et al. (1998)	[63]
	CHQ-CF87	Child Health Questionnaire Self-Report (87 version)	Landgraf and Abetz (1997)	[42]
CHRS [77]	CHRS	Children's Health Ratings Scale	Maylath (1990)	[77]
CHSCS [78]	CHSCS	Child's Health Self-Concept Scale	Hester (1984)	[78]
COOP [61]	COOP/WONCA Charts	Dartmouth Primary Care Cooperative Information Project	Nelson et al. (1987)	[61]
CQoL* [27]	CQoL	Child Quality of Life Questionnaire	Graham et al. (1997)	[27]
DHP* [21]	DHP-A	Duke Health Profile - Adolescent version	Parkerson et al. (1990)	[21]
ExQoL [35]	ExQoL	Exeter Quality of Life Measure	Eiser et al. (2000)	[35]
FSIIR [36]	FSIIR	Functional Status II Revised, Long version, infants	Stein and Jessop (1990)	[36]
	FSIIR	Functional Status II Revised, Long version, toddlers	Stein and Jessop (1990)	[36]
	FSIIR	Functional Status II Revised, Long version, preschoolers	Stein and Jessop (1990)	[36]
	FSIIR	Functional Status II Revised, Long version, school-age children	Stein and Jessop (1990)	[36]
	FSIIR-7	Functional Status II Revised 7 item	Stein and Jessop (1990)	[36]
	FSIIR-14	Functional Status II Revised 14 item	Stein and Jessop (1990)	[36]
GCQ [57]	GCQ	Generic Children's Quality of Life Measure	Collier (1997)	[57]
HALFS* [22]	HALFS	Health and Life Functioning Scale	Bastiaens and Dello Stritto (2004)	[22]
HAY* [23]	HAY	How Are You	Le Coq et al. (2000)	[23]
HP [23]	Healthy Pathways (SR)	Healthy Pathways Self Report	Bevans et al. (2010)	[51]
	Healthy Pathways (PR)	Healthy Pathways Parent Report	Bevans et al. (2012)	[53]
ITQOL* [24]	ITQOL	Infant Toddler Quality of Life Questionnaire (long version)	Klassen et al. (2003)	[24]
	ITQOL SF47	Infant Toddler Quality of Life Questionnaire (short version)	Landgraf et al. (2013)	[103]
KIDSCREEN [33]	KIDSCREEN-52	KIDSCREEN-52	Ravens-Sieberer et al. (2005)	[33]
	KIDSCREEN-27	KIDSCREEN-27	Ravens-Sieberer et al. (2007)	[39]
	KIDSCREEN-10	KIDSCREEN-10	Ravens-Sieberer et al. (2010)	[38]
KINDL [46]	Kiddy-KINDLR	Kiddy-Fragebogen zur Erfassung der gesundheitsbezogenen Lebensqualität bei Kindern und Jugendlichen Revidierte Form	Ravens-Sieberer and Bullinger (1998)	[46]
	Kid-KINDLR	Kid-Fragebogen zur Erfassung der gesundheitsbezogenen Lebensqualität bei Kindern und Jugendlichen Revidierte Form	Ravens-Sieberer and Bullinger (1998)	[46]
	Kiddo-KINDLR	Kiddo-Fragebogen zur Erfassung der gesundheitsbezogenen Lebensqualität bei Kindern und Jugendlichen Revidierte Form	Ravens-Sieberer and Bullinger (1998)	[65]
	CAT-SCREEN	A computer-assisted version		

continued on next page

Table 3 – *continued*

PROM group and citation number	Questionnaire acronym	Questionnaire full name	Author (year)	Questionnaire citation number
Nordic QOL-Q* [104]	Nordic QOL-Q for children	Nordic Quality of Life Questionnaire for children	Lindström and Köhler (1991), Lindström and Eriksson (1993)	[104,105]
PedsQL [47]	PedsQL Infant Scales	Pediatric Quality Of Life Inventory Trade Mark 4.0 Infant Scales	Varni et al. (2011)	[82]
	PedsQL 4.0	Pediatric Quality Of Life Inventory Trade Mark 4.0 Generic Core Scales	Varni et al. (1999)	[47]
	PedsQL-SF15 Generic Core Scales	Pediatric Quality Of Life Inventory Trade Mark 4.0 Short Form 15	Chan et al. (2005)	[59]
ComQOL [43]	ComQOL-S5	Comprehensive Quality of Life Scale-School version, Fifth edition	Cummins (1997)	[43]
	PWI-PS	Personal Wellbeing Index Pre-school	Cummins and Lau (2005)	[106]
	PWI-SC	Personal Wellbeing Index School Children	Cummins and Lau (2005)	[83]
QOLQA* [28]	QOLQA/TQOLQA	Quality of Life Questionnaire for Adolescents (Chinese, Japanese, and Taiwanese versions)	Wang et al. (2000)	[28]
	TQOLQA (Short version)	Quality of Life Questionnaire for Adolescents (Taiwanese version)	Fuh et al. (2005)	[107]
QoLQC* [26]	QoLQC	Quality of Life Questionnaire for Children	Bouman et al. (1999)	[26]
QoLP-AV [84]	QoLP-AV	Quality of Life Profile: Adolescent Version	Raphael et al. (1996)	[84]
SLSS [52]	SLSS	Student Life Satisfaction Scale	Huebner (1991)	[52]
	MSLSS	Multi-dimensional Student Life Satisfaction Scale	Huebner (1994)	[62]
	BMSLSS	Brief Multi-dimensional Student Life Satisfaction Scale	Seligson et al. (2003)	[37]
	MSLSS-A	Multi-dimensional Student Life Satisfaction Scale-Adolescent version	Gilligan and Huebner (2007)	[40]
TNO-AZL questionnaires* [29]	TAPQOL	TNO-AZL Questionnaire for Preschool Children's Health-Related Quality of Life	Fekkes et al. (2000)	[29]
	TACQOL	TNO-AZL Questionnaire for Children's Health-Related Quality of Life	Vogels et al. (1998)	[108]
	TAAQOL	TNO-AZL Questionnaire for Adult Health-Related Quality of Life	Bruil (2001)	[109]
VSP-A* [30]	VSP-A	Vécu et Santé Perçue de l'Adolescent	Simeoni et al. (2000)	[30]
WCHMP [60]	WCHMP	Warwick Child Health and Morbidity Profile	Spencer and Coe (1996)	[60]
YQoL [89]	YQoL-S	Youth Quality of Life instrument-Surveillance version	Edwards et al. (2002)	[89]
	YQoL-R	Youth Quality of Life instrument-Research version	Patrick et al. (2002)	[90]
Preference-based measures				
16D* [17]	16D	16 Dimensional	Apajasalo et al. (1996)	[17]
	17D	17 Dimensional	Apajasalo et al. (1996)	[110]
AQoL-6D* [18]	AQoL-6D	Assessment of Quality of Life Mark 2, 6D adolescents	Moodie et al. (2010)	[18]
CHU-9D [34]	CHU-9D	Child Health Utility 9D	Stevens (2009)	[34]
EQ-5D-Y [44]	EQ-5D-Y	EuroQol 5D Youth	Wille et al. (2010)	[44]
HUI [45]	HUI2	Health Utilities Index 2	Torrance et al. (1996)	[45]
	HUI3	Health Utilities Index 3	Feeny et al. (2002)	[58]
CHSCS-PS* [20]	CHSCS-PS	Comprehensive Health Status Classification System - Preschool	Saigal et al. (2005)	[20]

PROM, patient-reported outcome measure.
* These PROMs were excluded from the Results section because there was no eligible evidence.

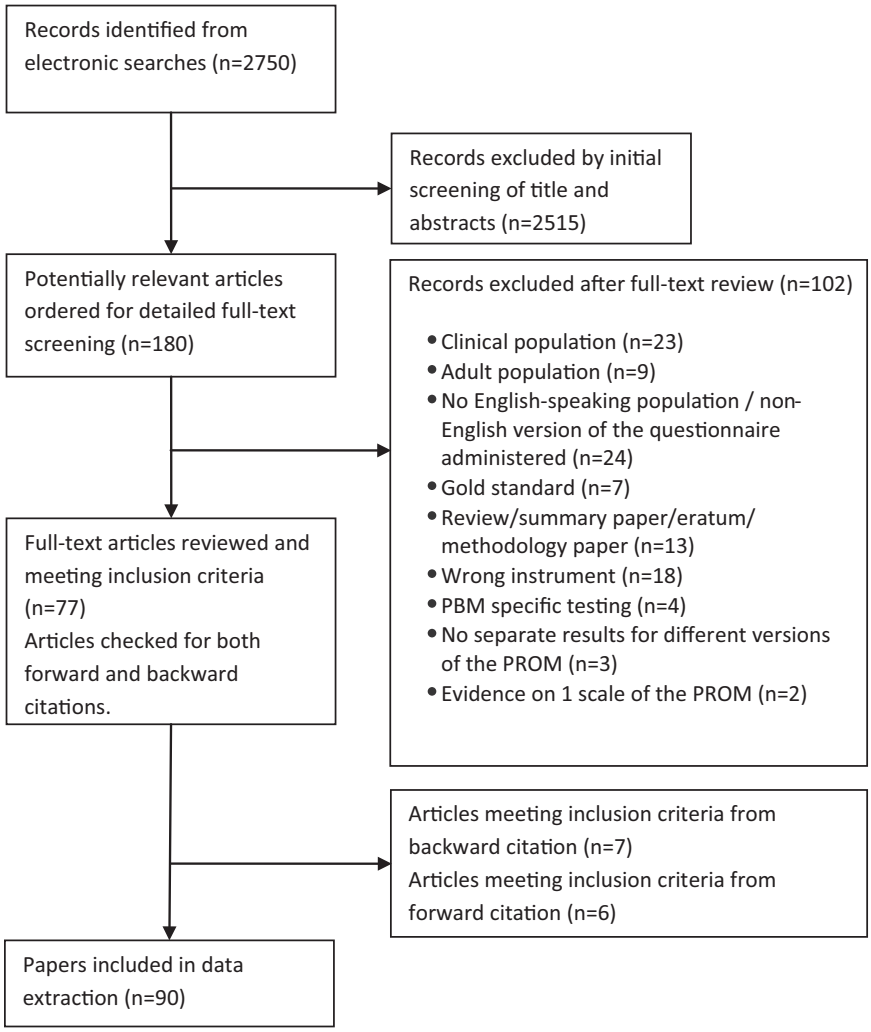

Fig. 1 – PRISMA flowchart describing identification and selection of studies evaluating psychometric performance of PROMs in the general population. PBM, preference-based measure; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PROM, patient-reported outcome measure.

Structural Validity

There is stronger evidence for structural validity using factor analysis for six questionnaires (at least some good evidence in favor): Student Life Satisfaction Scale (SLSS), Brief MSLSS (BMSLSS), KIDSCREEN-27, KIDSCREEN-52, Healthy Pathways, and PedsQL 4.0 [33,37,39,47,51,52]. The most robust evidence is available for Healthy Pathways [51,53] and KIDCREEN-52 [54], which are both supported by good-quality studies using Rasch analysis in addition to factor analysis. The factor structure of the Singapore English version of KINDL-Kid did not reflect that of the German version, with items loading on eight factors instead of the hypothesized six factors [55]. Four studies examined the factor structure of the CHQ-PF50, one reporting extra factors not accounted for in the scales; final models were acceptable to strong, with factor loadings varying from 0.34 to 0.86 [56]. Although the overall methodological quality of the study reporting on the development of the MSLSS-Adolescent version was rated as fair, mainly because of lack of information on how missing items were handled, a thorough exploratory and confirmatory factor analysis was performed to establish the six-factor structure [40]. We found no studies reporting on the structural validity of 13 questionnaires [32,34,35,41–45,57–61].

Construct Validity

There is stronger evidence for the construct validity of five questionnaires (at least some good evidence in favor): SLSS, MSLSS, BMSLSS, KIDSCREEN-10, and KIDSCREEN-52 [33,37,38,52,62]. Conflicting evidence was found for seven questionnaires [42,46,47,59,63–65]. One study compared Dartmouth Primary Care Cooperative Information Project scores of "at-risk" young people with those of their peers; only the Health Habits scale was able to differentiate between the two groups [64]. Data on construct validity for the CHQ-PF50 is available in one study reporting that "change in health" and "family cohesion" scales did not show significant differences between a clinical and a general population [66]. One study reports on hypothesis testing of the CHQ-CF87 comparing mean scores of a school sample with attention deficit/hyperactivity disorder and a clinical sample reporting mixed results; in addition, no data are provided to support the findings [42]. The construct validity of KINDL-Kid and KINDL-Kiddo was tested in one study, comparing mean scores of diabetic children with those of healthy children; the clinical group scored higher, indicating better quality of life, on a few scales of both questionnaires [67]. Seven articles reported on the construct validity of the PedsQL 4.0 and although most mean scale scores vary with health conditions, conflicting findings were found for the social

Table 4 – Summary appraisal of measurement properties of PROMs in a general population.

Instrument version	Content validity	Structural validity	Construct validity	Internal consistency	Test-retest reliability	Proxy reliability	Precision	Responsiveness
CHAQ	0	0	0	+	+	–	–	0
CHIP-CE	++	+	+	++	+/–	0	+	0
CHIP-AE	++	0	+	+	+/–	0	+	0
CHQ-PF50	0	+/–	+/–	+	+/–	0	–	0
CHQ-SF80/87	0	0	+/-	+	0	0	–	0
CHRS	+	+	+	+/–	0	0	0	0
CHSCS	+	+	+	–	–	0	0	0
CHU-9D	++	0	+	0	+/–	0	+/–	0
COOP	+	0	+/–	?	+	0	0	0
EQ-5D-Y	0	0	+	0	+/–	0	–	0
ExQoL	?	0	?	+	0	0	0	0
FSIIR-14	?	+	+	+	0	0	0	0
GCQ	+	0	+	+	0	0	0	0
HUI2	0	0	0	0	0	0	+/-	0
HUI3	0	0	0	0	0	0	+/-	0
Healthy Pathways	+	++	+	+	0	–	++	0
KIDSCREEN-52	++	+++	++	+	+/–	–	+	0
KIDSCREEN-27	0	++	+	+	+/–	0	+	0
KIDSCREEN-10	0	+	++	+	+	–	+	0
KINDL-Kid	0	–	+/–	+/–	0	0	+	0
KINDL-Kiddo	0	+/–	+/–	+/–	0	0	+	0
PedsQL 4.0	0	++	+/–	+	+/-	–	+/–	?
PedsQL Infant Scales	+	+	+	+	0	0	+/–	0
PedsQL-SF15	0	0	+/–	+/–	0	0	–	0
ComQoL-S5	0	0	+	+	+/–	0	0	0
PWI-SC	0	+	+	+	0	0	0	0
QoLP-AV	+	+	+	+	0	0	0	0
SLSS	+	++	++	++	+	0	0	0
MSLSS	+	+	++	+	+	0	0	0
BMSLSS	0	++	++	++	+	0	0	0
MSLSS-A	0	+	+	+	+	0	0	0
WCHMP	+	0	0	+	+/–	0	0	0
YQoL	+	+	+	+	+	0	0	0

BMSLSS, Brief Multi-dimensional Student Life Satisfaction Scale; CHAQ, Child Health Assessment Questionnaire; CHIP-AE, Child Health and Illness Profile- Adolescent Edition; CHIP-CE, Child Health and Illness Profile - Child Edition; CHQ-PF50, Child Health Questionnaire Parent Long Form; CHQ-PF80/87, Child Health Questionnaire Short Form (80/87 version); CHRS, Children's Health Ratings Scale; CHSCS, Child's Health Self-Concept Scale; CHU-9D, Child Health Utility 9D; ComQoL-S5, Comprehensive Quality Of Life Scale School version Fifth Edition; COOP, Dartmouth Primary Care Cooperative Information Project; EQ-5D-Y, EuroQol 5D Youth; PROM, patient-reported outcome measure; EQ-5D-Y, EuroQol 5D Youth; ExQoL, Exeter Quality of Life Measure; FSIIR-14, Functional Status II Revised 14 item; GCQ, Generic Children's Quality of Life Measure; HUI2, Health Utilities Index 2; HUI3, Health Utilities Index 3; MSLSS, Multi-dimensional Student Life Satisfaction Scale; MSLSS-A, Multi-dimensional Student Life Satisfaction Scale-Adolescent version; PedsQL 4.0, Pediatric Quality Of Life Inventory Trade Mark 4.0 Generic Core Scales; PedsQL Infant Scales, Pediatric Quality Of Life Inventory Trade Mark 4.0 - Infant Scales; PedsQL-SF15, Pediatric Quality Of Life Inventory Trade Mark 4.0 Short Form 15; PWI-SC, Personal Wellbeing Index School Children; QoLP-AV, Quality of Life Profile: Adolescent Version; SLSS, Student Life Satisfaction Scale; WCHMP, Warwick Child Health and Morbidity Profile; YqoL, Youth Quality of Life instrument.

domain scale [68,69] and emotional functioning [69–72]. A good-quality study examined the shortened version of the PedsQL and reported that the PedsQL-SF15 is able to discriminate between groups of different clinical status but is less sensitive to group differences than is the original [59].

Construct validity for the Exeter Quality of Life Measure was assessed in one study comparing the discrepancy between actual and ideal selves for children with and without asthma. Although a higher discrepancy was reported for children with asthma, statistical significance of the result was not tested [35]. No evidence of construct validity was found for four questionnaires [41,45,58,60].

Internal Consistency

The internal consistency of SLSS and MSLSS was examined in eight and seven studies, respectively (see Appendix Table 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2015.01.004), consistently reporting positive findings, with Cronbach alpha values of 0.7 and higher and good item-total correlations. Positive evidence for internal consistency of CHIP-CE, both parent-report version [31,73] and self-report version [74], was reported in three articles; one study reported only overall internal consistency [73]. The two studies assessing internal consistency of the adolescent version (CHIP-AE) reported marginally positive findings but showed some methodological flaws [75,76]. The article describing the development and initial testing of the Children's Health Ratings Scale reported that the 17-item scale was internally consistent; however, the authors did not assess internal consistency for the five factors identified in the factor analysis [77]. For the Child's Health Self-Concept Scale, an overall Cronbach alpha value of 0.7 was reported in the developmental article [78], but Hoyt coefficients were below 0.7 for four of the five subscales identified in the factor analysis. In the only study examining PedsQL-SF15, internal consistency was good (above 0.70) for all scales except "physical health" (0.60) [59]. Internal consistency of KINDL-Kid and Kiddo was studied in two studies reporting conflicting Cronbach alpha scale scores for both questionnaires [55,67]. The assessment of the internal consistency of Dartmouth Primary Care Cooperative Information Project is of poor quality and prevented us to appraise the findings [64]. Internal consistency was not tested for four questionnaires: EQ-5D-Y, HUI2, HUI3, and CHU-9D [34,44,45,58].

Test-Retest Reliability

This psychometric property was seldom assessed, and if evaluated, results were inconclusive. Two studies of good quality reported varying test-retest results for the Child Health and Illness Profile Child Edition Parent Report Form Version 76, including low intraclass correlation coefficients (ICCs) for the subscales physical comfort (0.63) and restricted activity (0.36) [31,73]. The test-retest reliability of the CHIP-CE child-report version fell to below 0.35 in younger children [74]. The CHIP-AE test-retest reliability was assessed in one study of good quality reporting mixed results, with 19 out of 20 scales having an ICC of 0.60 or higher, and one subdomain "home safety and health" with an ICC of 0.48 [76]. Test-retest reliability assessment of good to excellent quality was done for the CHQ-PF50, reporting generally moderate to high ICCs; however, the retest reliability dropped below 0.10 for physical functioning and role/social functioning after 6 weeks [56,66].

One study conducted a morning/afternoon test-retest of the EuroQol five-dimensional questionnaire and the CHU-9D on 24 children; percentage agreement was above 0.70 for all items, whereas weighted kappa coefficients were fair to moderate and slightly higher for CHU-9D dimensions [79]. A 2-week test-retest reliability was performed for all three KIDSCREEN versions: Kidscreen-52 ICCs varied from 0.56 to 0.77 [54], Kidscreen-27 ICCs ranged from 0.61 to 0.74 [39], and an overall ICC of 0.70 was reported for Kidscreen-10 [38]. PedsQL 4.0 test-retest reliability was assessed in one study, reporting good ICCs for the child version, but poor to moderate ICCs (0.34-0.79) for the proxy version [80]. Hester [78] tested the stability of the Child's Health Self-Concept Scale by readministering the test after 4 weeks. Although all findings were significant, correlations ranged from 0.44 (Healthiness) to 0.58 (Physical health).

For the Comprehensive Quality Of Life Scale School version Fifth Edition, 1-week temporal consistency was tested using multivariate analysis of variance, indicating a time effect for the Health domain for all three ratings (objective, subjective, and importance) [81]. Warwick Child Health and Morbidity Profile test-retest reliability has been assessed in one study, reporting weighted kappas between 0.50 and 0.86; however, they did not specify the retest period [60]. We found no evidence for test-retest reliability of 14 questionnaires [35,36,42,45,46,51,57–59,65,77,82–84].

Proxy Reliability

Proxy reliability has been studied for five questionnaires only: Child Health Assessment Questionnaire, Healthy Pathways, KIDSCREEN-10, KIDSCREEN-52, and PedsQL 4.0 [33,38,41,47,51]. All studies reported poor reliability (ICC $<$ 0.70) between self-reported and proxy-reported scores of most scales.

Precision

In six studies Rasch analysis was used to provide evidence for the precision of scores across the spectrum of measurement scales of Healthy Pathways [51,53] and Kidscreen-52 [54,85,86], Kidscreen-27 [39], and Kidscreen-10 [86]. The appraisal of this psychometric property for all other instruments was based on reported floor and/or ceiling effects. Three articles studying the CHQ-PF50 consistently reported high (up to 85%) ceiling effects for physical functioning and role/social limitations-emotional/behavioral [56,63,87]. Two studies indicated that the CHQ-CF87 suffers the same problem, with reported ceiling effects of 15% and higher (up to 89%) for five domains [87,88]. Likewise, substantial ceiling effects (12.4%–47%) were reported for PedsQL-SF15 [59]. One study reporting on the precision of the EQ-5D-Y mentioned over 70% of the respondents reporting top level for four EQ-5D-Y dimensions and shows that the distribution of the EQ-5D-Y values is concerning [79]. Equivocal findings were found for five questionnaires: CHU-9D, HUI2, HUI3, PedsQL 4.0, and PedsQL Infant Scales [34,45,47,58,82]. For each of these questionnaires we found at least one study reporting ceiling effects of 15% or higher. For fifteen questionnaires we found no studies reporting on floor or ceiling effects (Table 4).

Responsiveness

No studies were found reporting on responsiveness in a general population for any of the PROMs.

Discussion

For 14 of the 35 previously identified generic PROMs, we found no evidence of psychometric performance using English-language versions with children and young people. Evidence of psychometric properties was assessed in only a single study for a further nine questionnaires [35,36,57,64,77,78,84,89,90].

Five questionnaires had undergone testing for six or more properties [33,38,47,51,74]. Positive findings for at least five psychometric properties were found for four PROMs: CHIP,

Healthy Pathways, KIDSCREEN, and MSLSS; some versions of these PROMs showed good performance on four properties (CHIP-AE, KIDSCREEN-10, and BMSLSS and MSLSS-Adolescent version).

None of the eligible questionnaires has had all properties assessed. In this review, no evidence was found for any of the questionnaires to support responsiveness to detect meaningful change. This was, however, less surprising because we excluded clinical populations commonly used to test responsiveness. For seven PROMs no work has been undertaken to ensure content validity for any of the versions [41–47]. Data on precision and test-retest reliability were least available. With evidence lacking for only four questionnaires, construct validity and internal consistency were the most assessed properties.

Those questionnaires lacking evidence on internal consistency were all preference-based measures (PBMs) [34,44,45,58], which incorporate a weighting of scores based on a reference valuation of health states into a single index score. Not all the standard criteria for appraising PROMs are proposed to be appropriate for evaluating PBMs; for instance, the requirement for internal consistency may conflict with the underlying theory [91]. Nevertheless, the criteria for face, content, and construct validity and test-retest and proxy reliability remain apposite. We found no evidence, however, that these properties of PBMs had been tested with children and young people.

The findings regarding test-retest reliability are noteworthy partly because the property appears relatively unassessed and/or underreported and partly because when reported the retest reliability of one or more scales was often below the standard criteria (ICC $>$ 0.7 for use with groups). The implication of a scale with poor test-retest reliability is the likelihood of measurement error that is incurred as a consequence. Evaluating and quantifying test-retest reliability is of fundamental importance to understand whether changes in scores over time or following treatment are robust or simply due to random variation. Most evaluations of the property appear to be done at a 2-week interval, which is reasonable so that respondents are less likely to remember their precise answers, but only if some evidence or theory is presented that no change in status has occurred.

We found no good evidence of the reliability between reports by children and proxy reports by parents. Our findings are similar to those reported by Eiser and Morse [92] that reliability is often better for physical functioning and poorer for emotional and social domains. The evidence suggests poor proxy reliability for one or more domains of all candidate PROMs when this property has been assessed. It would be misleading to recommend a measure for which only some domains are reliable. Proxy reports may still have a use; for instance, they may be the only way to assess very young children. PROMs that specifically target children younger than 5 years include Comprehensive Health Status Classification System - Preschool, ComQOL (Personal Wellbeing Index Pre-school), FSIIR (infants, toddlers, pre-schoolers), KINDL (Kiddy), PedsQL (PedsQL Infant Scales), and Warwick Child Health and Morbidity Profile.

The quality of the studies set up to examine the psychometric performance of PROMs was highly variable. Although we included only peer-reviewed articles, some studies showed significant methodological limitations. In addition, the methodology of developing and evaluating PROMs has progressed over recent years [93,94]. Aside from KIDSCREEN and Healthy Pathways, however, it seems that little use has been made of newer methods of evaluation, such as Rasch analysis.

In contrast to previous reviews [8,9,92,95], our work provides an overall appraisal of measurement performance of PROMs for children and young people up to 18 years. We used the approach advocated by the COSMIN checklist to assess the methodological quality of studies reporting evaluations of psychometric performance [10]. There is undoubted benefit from identifying and considering the methodological quality of studies evaluating psychometric properties of PROMs. In our quality assurance checks with a second reviewer, however, we found the consistency of how those making the ratings interpret some parts of the COSMIN checklist to be an issue. The most difficult COSMIN item to code consistently was "how missing items are dealt with," and this item has a strong influence on the overall quality rating for most psychometric properties. The procedures for handling missing data may not have been reported in all articles that were included in the review, but may have been detailed in other articles or in the manual of the PROM. In addition, the aim and purpose of this exercise should be carefully considered in future systematic reviews because it is a time-consuming task.

Our inclusion criteria were restricted to published peer-reviewed studies that were specifically designed to evaluate measurement properties of PROMs in an English-speaking general population. Hence, we may have excluded articles such as trials and observational studies that present incidental evidence of psychometric performance. In addition, we may have excluded information contained in manuals that has not been published in peer-reviewed journals. Our justification for this is that peer review remains the scientific standard to ensure that methods have been scrutinized and findings can be considered robust and reliable. In addition, excluding studies that tested the performance of eligible PROMs in specific conditions might have overlooked studies reporting on responsiveness; change might be more typically expected in clinical populations and thus provide a more typical context to test the property of ability to detect change. One property included in the COSMIN checklist was not included in the review, cross-cultural validity. Psychometric performance cannot be assumed across languages and cultures [96]; therefore, in our view, limiting the review to evaluations of English-language versions is the strength of the review. Those wishing to use other language versions should appraise evidence in that language and with reference to recommended methods for establishing cross-cultural validity.

Implications for Policy, Practice, and Research

There are a number of research opportunities arising directly from this work, and implications for those using PROMs in research and/or interpreting research incorporating data emanating from PROM questionnaires. With none of the PROMs showing positive evidence for self-proxy reliability, using reports of parents and carers as proxies for outcomes designed to be measured from the perspectives of children and young people appears unsatisfactory. There will always be, however, children and young people who do not have the developmental cognitive capacity to self-report, and it is usually parents and carers who seek health care for their children. Therefore, parent report may be appropriate and may provide important insights. We advocate that the appropriate content of a parent questionnaire should differ for the children and young people's version. Parent questionnaires should assess items and concepts that are important to parents, and in ways that parents feel they can respond accurately. The potential for a primary carer measure but based on the domains of more importance to parents would seem a promising line of enquiry for research.

Adoption of more up-to-date methods for developing and evaluating PROMs is warranted [93]. Most notably there has been increasing use of Rasch analysis to evaluate the structural validity and provide evidence for the precision of scores across the spectrum of measurement scales. Rasch analysis can also be used to test for any evidence of invariance of how items perform across age groups and sex; it can also examine item invariance between different diagnoses, which would be warranted with generic PROMs. Evaluation of these aspects of generic PROMs for

children and young people appears to have been relatively unexplored. In addition, careful attention to the details of study design and transparent reporting of the methods and results is necessary [97].

Finally, we believe that this review is useful as a foundation for any systematic search for evidence regarding how a PROM might perform with any specific clinical population. Repeating the exercise of searching for and evaluating studies that set out to test PROMs for children with that condition will complete the picture for that purpose. For instance, we have completed such a process for appraising evaluations of candidate measures for children with neurodisability that we report in detail in the full project report [98].

Acknowledgments

This article is part of a report published by the National Institute for Health Research Library (Project 10/2002/16). Where possible, we have made substantial changes to avoid direct duplication; however, some content remains the same.

Source of financial support: This study was part of research funded by the National Institute for Health Research (NIHR) Health Services and Delivery Research programme (Project 10/2002/16 http://www.nets.nihr.ac.uk/projects/hsdr/10200216). The work also benefited from support from the NIHR Collaboration for Leadership in Applied Health Research and Care of the South West Peninsula (PenCLAHRC) and the charity Cerebra. The views and opinions expressed in this article are those of the authors and not necessarily those of the National Health Service, the NIHR, the Department of Health, or Cerebra.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2015.01.004 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] US Food and Drug Administration. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Rockville, MD: Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, 2009.
[2] National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal 2013. Available from: http://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf. [Accessed September 23, 2014].
[3] Fitzpatrick R. Patient-reported outcome measures and performance measurement. In: Smith PC, Mossialos E, Papanicolas I, et al., eds., Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects. Cambridge, UK: Cambridge University Press, 2009.
[4] Fitzpatrick R, Davey C, Buxton MJ, et al. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2:i–iv,1–74.
[5] Wild D, Eremenco S, Mear I, et al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. Value Health 2009;12:430–40.
[6] Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8:94–104.
[7] Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. Health Technol Assess 2001;5:1–157.
[8] Solans M, Pane S, Estrada MD, et al. Health-related quality of life measurement in children and adolescents: a systematic review of generic and disease-specific instruments. Value Health 2008;11:742–64.
[9] Schmidt LJ, Garratt AM, Fitzpatrick R. Child/parent-assessed population health outcome measures: a structured review. Child Care Health Dev 2002;28:227–37.
[10] Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10:22.
[11] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737–45.
[12] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19:539–49.
[13] Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651–7.
[14] Uijen AA, Heinst CW, Schellevis FG, et al. Measurement properties of questionnaires measuring continuity of care: a systematic review. PLoS One 2012;7:e42256.
[15] Davies N, Mackintosh A, Gibbons E, et al. A Structured Review of Patient-Reported Outcome Measures for Women with Breast Cancer (Appendix Bii). Oxford, UK: Oxford PROM Group, University of Oxford, 2009.
[16] McDowell I, Newell C. Measuring Health: A Guide to Rating Scales and Questionnaires. (2nd ed.. New York: Oxford University Press, 1996.
[17] Apajasalo M, Sintonen H, Holmberg C, et al. Quality of life in early adolescence: a sixteen-dimensional health-related measure (16D). Qual Life Res 1996;5:205–11.
[18] Moodie M, Richardson J, Rankin B, et al. Predicting time trade-off health state valuations of adolescents in four Pacific countries using the Assessment of Quality-of-Life (AQoL-6D) instrument. Value Health 2010;13:1014–27.
[19] Manificat S, Dazord A. Children's quality of life assessment: preliminary results obtained with the AUQUEI questionnaire. Qual Life Newslett 1998;19:2–3.
[20] Saigal S, Rosenbaum P, Stoskopf B, et al. Development, reliability and validity of a new measure of overall health for pre-school children. Qual Life Res 2005;14:243–57.
[21] Parkerson GR Jr, Broadhead WE, Tse CK. The Duke Health Profile: a 17-item measure of health and dysfunction. Med Care 1990;28:1056–72.
[22] Bastiaens L, Dello Stritto C. Validity and reliability of the Health and Life Functioning Scale. Presented at: 51th Annual Meeting of the American Academy of Child and Adolescent Psychiatry. Washington, DC, October 19–24, 2004.
[23] Le Coq EM, Colland VT, Boeke AJP, et al. Reproducibility, construct validity, and responsiveness of the 'How Are You?' (HAY), a self-report quality of life questionnaire for children with asthma. J Asthma 2000;37:43–58.
[24] Klassen AF, Landgraf JM, Lee SK, et al. Health related quality of life in 3 and 4 year old children and their parents: preliminary findings about a new questionnaire. Health Qual Life Outcomes 2003;1:81.
[25] Lindstrom B, Eriksson B. Quality of life among children in the Nordic countries. Qual Life Res 1993;2:23–32.
[26] Bouman NH, Koot HM, Van Gils AP, et al. Development of a health-related quality of life instrument for children: the Quality of Life Questionnaire for Children. Psychol Health 1999;14:829–46.
[27] Graham P, Stevenson J, Flynn D. A new measure of health-related quality of life for children: preliminary findings. Psychol Health 1997;12:655–65.
[28] Wang X, Matsuda N, Ma H, et al. Comparative study of quality of life between the Chinese and Japanese adolescent populations. Psychiatry Clin Neurosci 2000;54:147–52.
[29] Fekkes M, Theunissen NCM, Brugman E, et al. Development and psychometric evaluation of the TAPQOL: a health-related quality of life instrument for 1-5-year-old children. Quality Life Res 2000;9:961–72.
[30] Simeoni MC, Auquier P, Antoniotti S, et al. Validation of a French health-related quality of life instrument for adolescents: the VSP-A. Qual Life Res 2000;9:393–403.
[31] Riley AW, Forrest CB, Starfield B, et al. The Parent Report Form of the CHIP-Child Edition: reliability and validity. Med Care 2004;42:210–20.
[32] Riley AW, Forrest CB, Starfield B, et al. Reliability and validity of the adolescent health profile-types. Med Care 1998;36:1237–48.
[33] Ravens-Sieberer U, Gosch A, Rajmil L, et al. KIDSCREEN-52 quality-of-life measure for children and adolescents. Expert Rev Pharmacoecon Outcomes Res 2005;5:353–64.

[34] Stevens KJ. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. Qual Life Res 2009;18:1105–13.
[35] Eiser C, Vance Y. Seamark D. The development of a theoretically driven generic measure of quality of life for children aged 6-12 years: a preliminary report. Child Care Health Dev 2000;26:445–56.
[36] Stein RE, Jessop DJ. Functional status II(R): a measure of child health status. Med Care 1990;28:1041–55.
[37] Seligson JL, Huebner E, Valois RF. Preliminary validation of the Brief Multidimensional Students' Life Satisfaction Scale (BMSLSS). Soc Indic Res 2003;61:121–45.
[38] Ravens-Sieberer U, Erhart M, Rajmil L, et al. Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. Qual Life Res 2010;19:1487–500.
[39] Ravens-Sieberer U, Auquier P, Erhart M, et al. The KIDSCREEN-27 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. Qual Life Res 2007;16:1347–56.
[40] Gilligan TD, Huebner S. Initial development and validation of the multidimensional students' life satisfaction scale-adolescent version. App Res Qual Life 2007;2:1–16.
[41] Singh G, Athreya B, Fries J, et al. Measurement of health status in children with juvenile rheumatoid arthritis. Arthritis Rheum 1994;37:1761–9.
[42] Landgraf JM, Abetz LN. Functional status and well-being of children representing three cultural groups: initial self-reports using the CHQ-CF87. Psychol Health 1997;12:839–54.
[43] Cummins RA. The Comprehensive Quality of Life Scale - Adult (5th Edition). Melbourne, Australia: Deakin University, 1997.
[44] Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Qual Life Res 2010;19:875–86.
[45] Torrance GW, Feeny D, Furlong W, et al. Multiattribute utility function for a comprehensive health status classification system: Health Utilities Index Mark 2. Med Care 1996;34:702–22.
[46] Ravens-Sieberer U, Bullinger M. Assessing health-related quality of life in chronically ill children with the German KINDL: first psychometric and content analytical results. Qual Life Res 1998;7:399–407.
[47] Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the pediatric quality of life inventory. Med Care 1999;37:126–39.
[48] Bullinger M. KINDL – a questionnaire for health-related quality of life assessment in children. Zeitschrift für Gesundheitspsychologie 1994;1:64–77.
[49] Varni JW, Katz ER, Seid M, et al. The pediatric cancer quality of life inventory-32 (PCQL-32), I: reliability and validity. Cancer 1998;82:1184–96.
[50] Varni JW, Katz ER, Seid M, et al. The Pediatric Cancer Quality of Life Inventory (PCQL), I: instrument development, descriptive statistics, and cross-informant variance. J Behav Med 1998;21:179–204.
[51] Bevans KB, Riley AW, Forrest CB. Development of the healthy pathways child-report scales. Qual Life Res 2010;19:1195–214.
[52] Huebner E. Initial development of the Student's Life Satisfaction Scale. Sch Psychol Int 1991;12:231–40.
[53] Bevans KB, Riley AW, Forrest CB. Development of the Healthy Pathways Parent-Report Scales. Qual Life Res 2012;21:1755–70.
[54] Ravens-Sieberer U, Gosch A, Rajmil L, et al. The KIDSCREEN-52 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. Value Health 2008;11:645–58.
[55] Wee H-L, Ravens-Sieberer U, Erhart M, et al. Factor structure of the Singapore English version of the KINDL children quality of life questionnaire. Health Qual Life Outcomes 2007;5:4.
[56] Waters E, Salmon L, Wake M. The parent-form Child Health Questionnaire in Australia: comparison of reliability, validity, structure, and norms. J Pediatr Psychol 2000;25:381–91.
[57] Collier J. Developing a generic child quality of life questionnaire. Health Psychol Update 1997;28:12–6.
[58] Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care 2002;40:113–28.
[59] Chan KS, Mangione-Smith R, Burwinkle TM, et al. The PedsQL: reliability and validity of the short-form generic core scales and Asthma Module. Med Care 2005;43:256–65.
[60] Spencer NJ, Coe C. The development and validation of a measure of parent-reported child health and morbidity: the Warwick Child Health and Morbidity Profile. Child Care Health Dev 1996;22:367–79.
[61] Nelson E, Wasson J, Kirk J, et al. Assessment of function in routine clinical practice: description of the COOP chart method and preliminary findings. J Chronic Dis 1987;40:558–638.
[62] Huebner ES. Preliminary development and validation of a multidimensional life satisfaction scale for children. Psychol Assess 1994;6:149–58.
[63] Landgraf JM, Maunsell E, Speechley KN, et al. Canadian-French, German and UK versions of the Child Health Questionnaire: methodology and preliminary item scaling results. Qual Life Res 1998;7:433–45.
[64] Wasson JH, Kairys SW, Nelson EC, et al. A short survey for assessing health and social problems of adolescents. Dartmouth Primary Care Cooperative Information Project (The COOP). J Fam Pract 1994;38:489–94.
[65] Ravens-Sieberer U, Bullinger M. News from the KINDL-Questionnaire – a new version for adolescents. Qual Life Res 1998;7:653.
[66] Nugent J, Ruperto N, Grainger J, et al. The British version of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ). Clin Exp Rheumatol 2001;19(Suppl.): S163–7.
[67] Wee H-L, Lee WWR, Ravens-Sieberer U, et al. Validation of the English version of the KINDL (R) generic children's health-related quality of life instrument for an Asian population - results from a pilot test. Qual Life Res 2005;14:1193–200.
[68] Varni JW, Limbers CA, Burwinkle TM. Parent proxy-report of their children's health-related quality of life: an analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL 4.0 Generic Core Scales. Health Qual Life Outcomes 2007;5:2.
[69] Huang IC, Thompson LA, Chi YY, et al. The linkage between pediatric quality of life and health conditions: establishing clinically meaningful cutoff scores for the PedsQL. Value Health 2009;12:773–81.
[70] Varni JW, Limbers CA, Burwinkle TM. How young can children reliably and validly self-report their health-related quality of life? An analysis of 8,591 children across age subgroups with the PedsQL 4.0 Generic Core Scales. Health Qual Life Outcomes 2007;5:1.
[71] Davis E, Shelly A, Waters E, et al. Measuring the quality of life of children with cerebral palsy: comparing the conceptual differences and psychometric properties of three instruments. Dev Med Child Neurol 2010;52:174–80.
[72] Varni JW, Seid M, Knight TS, et al. The PedsQL 4.0 Generic Core Scales: sensitivity, responsiveness, and impact on clinical decision-making. J Behav Med 2002;25:175–93.
[73] Riley AW, Chan KS, Prasad S, et al. A global measure of child health-related quality of life: reliability and validity of the Child Health and Illness Profile - Child Edition (CHIP-CE) global score. J Med Econ 2007;10:91–106.
[74] Riley AW, Forrest CB, Rebok GW, et al. The Child Report Form of the CHIP-Child Edition: reliability and validity. Med Care 2004;42:221–31.
[75] Starfield B, Bergner M, Ensminger M, et al. Adolescent health status measurement: development of the Child Health and Illness Profile. Pediatrics 1993;91:430–5.
[76] Starfield B, Riley AW, Green BF, et al. The adolescent child health and illness profile: a population-based measure of health. Med Care 1995;33:553–66.
[77] Maylath NS. Development of the Children's Health Ratings Scale. Health Educ Q 1990;17:89–97.
[78] Hester NO. Child's Health Self-Concept Scale: its development and psychometric properties. ANS Adv Nurs Sci 1984;7:45–55.
[79] Canaway A, Frew E. Measuring preference-based quality of life in children aged 6–7 years: a comparison of the performance of the CHU-9D and EQ-5D-Y—the WAVES Pilot Study. Qual Life Res 2013;22:173–83.
[80] Iannaccone ST, Hynan LS, Morton A, et al. The PedsQL in pediatric patients with spinal muscular atrophy: feasibility, reliability, and validity of the Pediatric Quality of Life Inventory Generic Core Scales and Neuromuscular Module. Neuromuscul Disord 2009;19:805–12.
[81] Gullone E, Cummins RA. The comprehensive quality of life scale: a psychometric evaluation with an adolescent sample. Behav Change 1999;16:127–39.
[82] Varni JW, Limbers CA, Neighbors K, et al. The PedsQL Infant Scales: feasibility, internal consistency reliability, and validity in healthy and ill infants. Qual Life Res 2011;20:45–55.
[83] Cummins RA, Lau ALD. Personal Wellbeing Index - School Children (PWI-SC) (3rd ed.). Melbourne, Australia: Deakin University, 2005.
[84] Raphael D, Rukholm E, Brown I, et al. The quality of life profile - Adolescent version: background, description, and initial validation. J Adolesc Health 1996;19:366–75.
[85] Robitail S, Simeoni M-C, Erhart M, et al. Validation of the European proxy KIDSCREEN-52 pilot test health-related quality of life questionnaire: first results. J Adolesc Health 2006;39:596, e1–10.
[86] Erhart M, Ottova V, Gaspar T, et al. Measuring mental health and well-being of school-children in 15 European countries using the KIDSCREEN-10 Index. Int J Public Health 2009;54(Suppl. 2):160–6.
[87] Waters E, Wright M, Wake M, et al. Measuring the health and well-being of children and adolescents: a preliminary comparative evaluation of the Child Health Questionnaire in Australia. Ambulatory Child Health 1999;5:131–41.
[88] Waters EB, Salmon LA, Wake M, et al. The health and well-being of adolescents: a school-based population study of the self-report Child Health Questionnaire. J Adolesc Health 2001;29:140–9.

[89] Edwards TC, Huebner CE, Connell FA, et al. Adolescent quality of life, part I: conceptual and measurement model. J Adolesc 2002;25:275–86.
[90] Patrick DL, Edwards TC, Topolski TD. Adolescent quality of life, part II: initial validation of a new instrument. J Adolesc 2002;25:287–300.
[91] Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. Health Econ 1999;8:41–51.
[92] Eiser C, Morse R. The measurement of quality of life in children: past and future perspectives. J Dev Behav Pediatr 2001;22:248–56.
[93] Huang IC, Revicki DA, Schwartz CE. Measuring pediatric-patient-reported outcomes: good progress but a long way to go. Qual Life Res 2014;23:747–50.
[94] Matza LS, Patrick DL, Riley AW, et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO Good Research Practices for the Assessment of Children and Adolescents Task Force. Value Health 2013;16:461–79.
[95] Ravens-Sieberer U, Erhart M, Wille N, et al. Generic health-related quality-of-life assessment in children and adolescents: methodological considerations. Pharmacoeconomics 2006;24:1199–220.
[96] Saxena S, Carlson D, Billington R. The WHO quality of life assessment instrument (WHOQOL-Bref): the importance of its items for cross-cultural research. Qual Life Res 2001;10:711–21.
[97] Matza LS, Swensen AR, Flood EM, et al. Assessment of health-related quality of life in children: a review of conceptual, methodological, and regulatory issues. Value Health 2004;7:79–92.
[98] Morris C, Janssens A, Allard A, et al. Informing the NHS Outcomes Framework: Evaluating Meaningful Health Outcomes for CHIldren with Neurodisability Using Multiple Methods Including Systematic Review, Qualitative Research, Delphi Survey and Consensus Meeting. Southampton, UK: NIHR Journals Library 2014.
[99] Manificat S, Dazord A, Langue J, et al. A new instrument to evaluate infant quality of life. Qual Life Newslett 1999;23:7–8.
[100] Gayral-Taminh M, Bravi C, Depond M, et al. Auto-évaluation de la qualité de vie d'enfants de 6 à 12 ans: Analyse du concept et élaboration d'un outil prototype. Santé Publique 2005;7:35–45.
[101] Manificat S, Dazord A. Assessing adolescent's quality of life: validation of a new questionnaire. Qual Life Newslett 2002;28:2–3.
[102] Kurtin PS, Landgraf JM, Abetz L. Patient-based health status measurements in pediatric dialysis: expanding the assessment of outcome. Am J Kidney Dis 1994;24:376–82.
[103] Landgraf J, Vogel I, Oostenbrink R, et al. Parent-reported health outcomes in infants/toddlers: measurement properties and clinical validity of the ITQOL-SF47. Qual Life Res 2013;22:635–46.
[104] Lindström B, Köhler L. Youth, disability and quality of life. Pediatrician 1991;18:121–8.
[105] Lindström B, Eriksson B. Quality of life among children in the Nordic countries. Qual Life Res 1993;2:23–32.
[106] Cummins RA, Lau ALD. Personal Wellbeing Index - Pre-School (PWI-PS) (3rd ed.). Melbourne, Australia: Deakin University, 2005.
[107] Fuh JL, Wang SJ, Lu SR, et al. Assessing quality of life for adolescents in Taiwan. Psychiatry Clin Neurosci 2005;59:11–8.
[108] Vogels T, Verrips GHW, Verloove-Vanhorick SP, et al. Measuring health-related quality of life in children: the development of the TACQOL parent form. Qual Life Res 1998;7:457–65.
[109] Bruil J, Fekkes M, Vogels T. The validity and reliability of the TAAQOL: a health-related quality of life instrument comprising health status weighted by the impact of problems on well being. Qual Life Res 2001;10:257.
[110] Apajasalo M, Rautonen J, Holmberg C, et al. Quality of life in pre-adolescence: a 17-dimensional health-related measure (17D). Qual Life Res 1996;5:532–8.