
Article:
Rowen, D. orcid.org/0000-0003-3018-5109, Keetharuth, A., Poku, E. et al. (3 more authors) (2021) A review of the psychometric performance of selected child and adolescent preference-based measures used to produce utilities for child and adolescent health. Value in Health, 24 (3). pp. 443-460. ISSN 1098-3015

https://doi.org/10.1016/j.jval.2020.09.012

ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

A Review of the Psychometric Performance of Selected Child and Adolescent Preference-Based Measures Used to Produce Utilities for Child and Adolescent Health

Donna Rowen, BA, MSc, PhD,* Anju D. Keetharuth, BA, MA, MBA, PhD, Edith Poku, BSc, MBChB, MSc, FHEA, Ruth Wong, BSc, MRes, MSc, PhD, Becky Pennington, BSc, MSc, Allan Wailoo, BSc, MA, MSc, PhD

ABSTRACT

Objective: This review examined the psychometric performance of 4 generic child- and adolescent-specific preference-based measures that can be used to produce utilities for child and adolescent health.

Methods: A systematic search was undertaken to identify studies reporting the psychometric performance of the Child Health Utility (CHU9D), EQ-5D-Y (3L or 5L), and Health Utilities Index Mark 2 (HUI2) or Mark 3 (HUI3) in children and/or adolescents. Data were extracted to assess known-group validity, convergent validity, responsiveness, reliability, acceptability, and feasibility. Data were extracted separately for the dimensions and utility index where this was reported.

Results: The review included 76 studies (CHU9D n = 12, EQ-5D-Y-3L n = 20, HUI2 n = 26,HUI3 n = 43), which varied considerably across conditions and sample size. EQ-5D-Y-3L had the largest amount of evidence of good psychometric performance in proportion to the number of studies examining performance. The majority of the evidence related to EQ-5D-Y-3L was based on dimensions. CHU9D was assessed in fewer studies, but the majority of studies found evidence of good psychometric performance. Evidence for HUI2 and HUI3 was more mixed, but the studies were more limited in sample size and statistical power, which was likely to have affected performance.

Conclusions: The heterogeneity of published studies means that the evidence is based on studies across a range of countries, populations and conditions, using different study designs, different languages, different value sets and different statistical techniques. Evidence for CHU9D in particular is based on a limited number of studies. The findings raise concerns about the comparability of self-report and proxy-report responses to generate utility values for children and adolescents.

Keywords: adolescents, children, CHU9D, EQ-5D-Y-3L, HUI2, HUI3, paediatric, QALYs, utilities.

VALUE HEALTH. 2020; ■(■):■–■

Introduction

Health technology assessment (HTA) can be used as a tool for informing resource allocation decisions by assessing the cost-effectiveness of interventions and enabling comparisons of relative cost-effectiveness across a range of interventions for different conditions and populations. The quality-adjusted life-year (QALY) is commonly used to capture the benefit of interventions for use in HTA. QALY is calculated by quality-weighting survival using a quality adjustment, which is often generated using a generic preference-based measure. A preference-based measure consists of a classification system and a value set that is used to score responses to the classification system. The classification system contains dimensions with severity levels. Responses to the classification system are used to assign people to a health state. A value set is then used to score the relative value of the health state to generate a utility value (index score), on the 1-to-0 full health–dead scale, with values less than zero indicating that the health state is worse than being dead. There are many different preference-based measures available, and these can be condition-specific or generic, and population-specific (eg, adults or children) or suitable across many populations.

Measures for estimating adult health utilities are often assessed by reference to the psychometric performance of measures, such as assessing their validity and responsiveness in particular populations. The psychometric performance of the main generic preference-based measures has been assessed widely in the published literature, and there is a published review of reviews about their performance.[1] This means that both researchers and decision makers have knowledge about the appropriateness of the utility values generated by these measures across a range of conditions and also about whether the measure would be

* Address correspondence to: Donna Rowen, PhD, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, England, United Kingdom, S1 4DA. Email: d.rowen@sheffield.ac.uk

expected to identify a statistically significant change in that population. This can provide valuable information about the confidence in the utility estimates and interpretation of changes in utility values; however, to our knowledge there is no review of the published literature examining the psychometric performance of the child and adolescent preference-based measures.

One existing review examined the development and application of generic preference-based measures available for use in pediatric populations[2] and found 9 measures (Adolescent Health Utility Measure, Assessment of Quality of Life [AQoL-6D], Child Health Utility [CHU9D], EQ-5D-Y-3L, Health Utilities Index Mark 2 (HUI2) and Mark 3 (HUI3), 16D, 17D, and the Quality of Wellbeing Scale [QWB]); researchers concluded that further empirical analyses are required to examine the relative performance of these measures. One review examined the valuation methods used to generate the values sets of these preference-based measures.[3] Another review found that 7 of these preference-based measures had been commonly used internationally in pediatric populations: AQoL-6D, EQ-5D-Y-3L, CHU9D, HUI2, HUI3, 15D/16D/17D, and QWB.[4] This review[4] provided a fully comprehensive source of published utility values from the existing literature across a range of conditions but did not assess the psychometric performance of the measures used to generate the utility values.

The present review examined the psychometric performance of a selection of generic child and adolescent-specific preference-based measures that can be used to generate utility values for children and adolescents: CHU9D, EQ-5D-Y-3L, HUI2, and HUI3.

Methods

Description of Measures

The measures selected for inclusion in the review were CHU9D, EQ-5D-Y-3L, HUI2, and HUI3 because the authors, after consultation with National Institute for Health and Care Excellence staff, considered these to be the measures most appropriate to inform UK policy using criteria about intended and worded appropriately for use in children and adolescents, applicability across conditions using a generic classification system of dimensions and levels, development (or validation) with an English-speaking population, and potential availability in datasets used to inform UK policy. Following is a summary of each of the child- and adolescent-specific preference-based measures examined in the review.

CHU9D

The CHU9D has 9 dimensions each with 5 severity levels: worry, sadness, pain, tiredness, annoyance, school, sleep, daily routine, and activities. The dimensions and severity levels were developed using qualitative research with children aged 7 to 11 years and hence were designed for this age group; however, the measure can be completed via parent/guardian proxy for children aged 4 to 7 years and has been used in adolescents aged 12 to 18 years. Value sets exist for the UK,[5] Australia,[6–9] The Netherlands,[10] and China.[11]

EQ-5D-Y-3L and EQ-5D-Y-5L

The EQ-5D-Y is the youth version of the EQ-5D. The EQ-5D-Y was generated by adapting the adult EQ-5D to ensure relevance and clarity for children and adolescents.[12–14] The same 5 dimensions of the adult EQ-5D are each retained in the EQ-5D-Y, but the dimensions and levels were reworded to be appropriate for children by an international team of collaborators. The EQ-5D-Y-3L has 5 dimensions each with 3 levels of severity: mobility; looking after myself; doing usual activities; having pain or discomfort; and feeling worried, sad or unhappy. There is also a more recent 5-level youth version of the EQ-5D called the EQ-5D-Y-5L.[15] There is no officially accepted value set for the EQ-5D-Y-3L or EQ-5D-Y-5L, though there is a published value set for the EQ-5D-Y-3L in the United States.[16] Recent research has found that current EQ-5D value sets cannot be appropriately used to value EQ-5D-Y health states.[17,18] The EuroQol Group is currently developing an international valuation protocol for the development of country-specific EQ-5D-Y value sets.[19]

HUI2

The HUI2 has 7 dimensions: sensation, mobility, emotion, cognition, self-care, pain, and fertility.[20] Each dimension has between 3 and 5 levels. The measure was originally developed in Canada for use in childhood cancer but is widely used as a generic measure, although the fertility dimension is rarely used. The HUI2 has a UK value set[21] and a Canadian value set[20] and can be used to measure health of children and adults aged 5 and over. HUI2 and HUI3 are typically administered using a single set of 15 self-administered questions, which are then used to generate both HUI2 and HUI3 utilities. Interviewer administration of the set of items used to generate both HUI2 and HUI3 utilities involves between 13 and 39 questions.

HUI3

The HUI3 was developed in Canada with the purpose of resolving some issues faced with HUI2 and for applicability to clinical and general populations. HUI3 has 8 dimensions: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain.[22] Each dimension has between 5 and 6 levels. The HUI3 has only a Canadian value set.[22] The HUI3 can be used to measure health of children and adults aged 5 and older.

Search Strategy and Data Identification

The objective of the literature search was to identify relevant published studies reporting evidence of the psychometric performance of CHU9D, EQ-5D-Y (3L or 5L), HUI2, and HUI3 in children and adolescents. A systematic search was conducted in Medline (Ovid), PsycINFO (Ovid), and the Web of Science Core Collection Science Citation Index Expanded (Clarivate Analytics) in March and September 2019 to identify studies reporting the psychometric performance of EQ-5D-Y (3L or 5L), CHU9D, HUI2, AQoL-6D, and HUI3 in children and adolescents. Terms for the measure (eg, "EQ-5D-Y," "HUI," "CHU9D") were combined with "child" population terms derived from a recently published systematic review of child utilities (which did not assess psychometric performance of measures).[4] The search strategy was translated across each database, and limits for human studies and English language were applied. No study type or publication date limits were applied.

Supplementary gray literature searches included the conference abstract websites in the last 3 years (The International Society for Pharmacoeconomics and Outcomes Research and International Society for Quality of Life Research), Web of Science Cited Reference Search, keyword searching using Google Scholar search engine, and examination of reference lists of included studies.

The criteria for inclusion were an assessment of known-group validity, convergent validity, responsiveness, reliability, acceptability or feasibility of EQ-5D-Y (3L or 5L), CHU9D, HUI2, or HUI3; obtained from pediatric populations or relevant parents/caregivers acting as proxies for children; with analyses reported separately for participants aged younger than 18 years. The search included the AQoL-6D and EQ-5D-Y-5L measures but identified

only 1 relevant study each for AQoL-6D[23] and EQ-5D-Y-5L,[24] meaning that AQoL-6D and EQ-5D-Y-5L were post hoc exclusions from the review.

Retrieved records were screened by 1 of 3 reviewers, with a 10% randomly selected sample of titles and abstracts double checked by a second reviewer. Full-text papers of potentially relevant records were obtained and examined in detail. Study selection was based on predefined eligibility criteria. Summary data relating to study and population characteristics was extracted by 1 of 2 reviewers and checked by a second reviewer for all papers. Subsequently, 2 reviewers independently double extracted the psychometric analyses for 3 randomly selected papers and, after comparing extractions and amended using consensus (differences occurred only in the amount of detail extracted), undertook single extraction of the psychometric data of the remaining papers.

Analytic Strategy

Assessments of the psychometric performance of patient-reported outcome measures (PROMs) focus on validity, reliability and responsiveness,[25] though not all aspects that are relevant for PROMs are relevant for preference-based measures. For example, internal consistency reliability examines whether items within a measure are measuring the same construct, which is important for a PROM but not for a preference-based measure. We focused on the aspects of reliability, validity, and responsiveness that are important for assessing the performance of preference-based measures[1,26,27] and assessments of feasibility and acceptability. We included both the performance of the utility index and dimensions where this information was available. Data were extracted about the following:

- *Measure(s):* preference-based measure(s) used; language version; whether the measure was self-reported and/or proxy reported by parents/caregivers or both; whether the paper assessed the index (ie, the utility scores generated by the measure), dimensions, or both index and dimensions; country preference weights applied (where applicable); other health-related quality of life (HRQL) measures or clinical measures used.
- *Study sample:* age of participants (mean age and age range); proportion of female participants; whether the sample consisted of members of the general population, patients, or both; clinical area (where applicable); sample size.
- *Known-group validity:* assesses the ability to differentiate between groups of different severity[28] or a less rigorous test of case-control construct validity that examines the ability to differentiate between people with and without the condition; determined using the ability to identify a reported statistically significance difference at the 5% level across known groups, along with whether the direction of the difference is in accordance with clinical expectation (eg, general population with higher utilities than patients). Where studies assess dimensions, it is not typically expected that all dimensions will necessarily capture known-group differences because not all conditions affect all dimensions.
- *Convergent validity:* assesses the strength of association between the measure of interest and other measures of HRQL (generic or condition-specific) or disease severity using either correlation coefficients[29] (a more conventional technique) or statistical significance in regression analyses. Evidence of convergent validity is determined by whether moderate (0.41-0.60) or good (0.61-0.80) (or higher and almost perfect) agreement has been identified. These are arbitrary cut-offs but are often reported in the papers included in the review (and based on established criteria; see, eg, Landis and Koch [1977]).[30] Evidence of convergent validity focuses on expected correlations where these are motivated in theory (eg, pain dimensions in 2 measures).
- *Responsiveness:* assesses the ability to capture change over time, where change is expected (eg, as a result of treatment effects).[31] Evidence of responsiveness is determined by the ability to determine a statistically significance change at the 5% level over time. It also considers whether the direction of the change is in accordance with clinical expectation (eg, higher index scores at the end of treatment than at baseline). Where dimensions are assessed, it is not necessarily expected that all dimensions will be responsive because not all conditions or treatments affect all dimensions.
- *Reliability:* assesses the degree of change where no change in health is identified using other HRQL or clinical measures. Evidence of reliability is determined by whether the measure is able to reproduce the same value on 2 separate administrations when there has been no change in health, where this can be over time (test-retest reliability),[32] between methods of administration (intermodal reliability) or between raters (ie, self-report and parent proxy report [interrater reliability]).[33] Reliability is sometimes identified for most but not all dimensions, but if reliability is not identified for some dimensions, this raises issues about reliability of the entire measure.
- *Acceptability and feasibility:* assess the practicality of a measure for administration in a specific group of people, covering aspects such as burden of completion and whether the person completing the measure can meaningfully respond to the questions being asked. Evidence of acceptability and feasibility is indicated where the study reports, for example, low missing data or high levels of understanding. A lack of evidence for acceptability and feasibility is concluded where the study reports, for example, high levels of missing data or low levels of understanding. For child and adolescent measures this includes whether the child and adolescent or the proxy can meaningfully complete the measure, because there may be problems of understanding for younger people and problems of knowing the required information (eg, how the child feels emotionally) for proxy report.
- Other psychometric analyses reported.

Where reported, data were extracted for each of the psychometric assessments about brief summary of analysis undertaken, whether the results were in accordance with clinical expectation (where relevant), and whether the findings were statistically significant.

Results

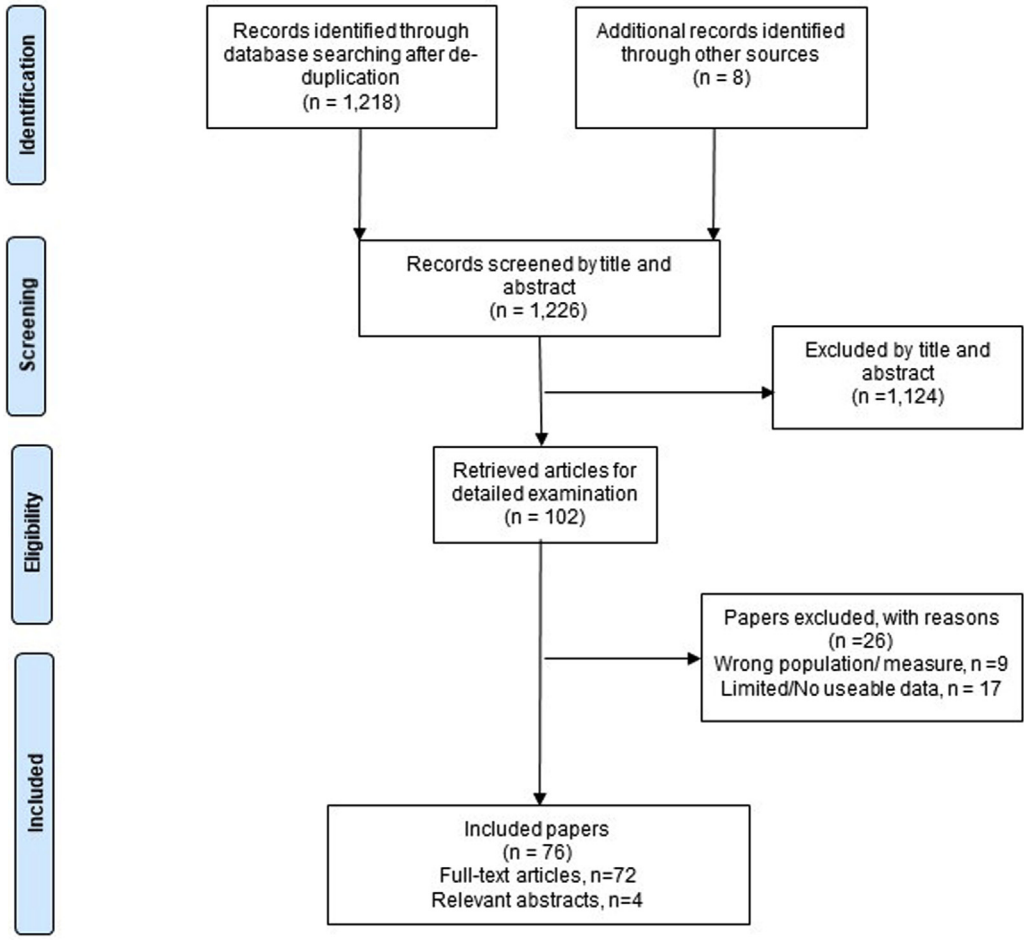
Search Results

A total of 1218 unique records were identified, with 8 additional records identified from reference lists. Of these, 102 records were examined in detail. After the exclusion of 26 papers (see Supplemental Table found at https://doi.org/10.1016/j.jval.2020.09.012), 76 papers, including 72 full-text articles and 4 conference abstracts,[34-37] were considered suitable for providing evidence for the psychometric assessment of EQ-5D-Y-3L, CHU9D, HUI2, and HUI3. A PRISMA flow chart of the study selection process is shown in Figure 1.

Included Studies

Characteristics of included studies are summarized in Table 1, and further details are provided in the Supplemental Table (found

Figure 1. PRISMA flow chart.



at https://doi.org/10.1016/j.jval.2020.09.012). Out of the 76 studies, 53 studies assess only 1 of the child- and adolescent-specific preference-based measures assessed here. Nineteen studies assess both HUI2 and HUI3,[34,38-54] 2 studies assess CHU9D and EQ-5D-Y-3L,[55,56] 1 assesses EQ-5D-Y-3L and HUI2,[57] and 1 assesses CHU9D and HUI2.[7,58] Forty-two studies assess HUI3, 26 studies assess HUI2, 20 studies assess EQ-5D-Y-3L, and 12 studies assess CHU9D. One study[24] compares EQ-5D-Y-3L and EQ-5D-Y-5L.

Value Sets

CHU9D studies use the UK value set (n = 9), Australian adolescent value set (n = 4), Australian adult value set (n = 2). EQ-5D-Y-3L has no accepted value set, meaning that 15 studies do not generate utility scores. The remaining 5 studies assessing EQ-5D-Y-3L scored health states using the UK, Australian, French, and Spanish value sets for the EQ-5D and an unofficial US EQ-5D-Y-3L value set (1 study each). Twenty studies used the Canadian HUI2 value set, 2 used the UK value set, 3 did not use a value set, and 1 did not report the value set used. The HUI3 only has a Canadian value set, though 4 studies reported results for dimensions only.

Country and Language

The data assessed in the studies are from a variety of countries, with Canada (n = 16), the UK (n = 12), the United States (n = 9), and Australia (n = 8) having the largest number of included studies, followed by The Netherlands (n = 4), Sweden (n = 4), Spain (n = 3), China (n = 2), Germany (n = 2), South Africa (n = 2), and many countries with 1 study (France, Hong Kong, Italy, New Zealand, South Korea, Taiwan, Thailand, and Turkey), 2 multinational studies (each included Germany, Italy, South Africa, Spain, and Sweden), 1 study in Australia and New Zealand, 1 study in the UK and Ireland, 1 study in the UK and the United States, and 1 study in which the country was not reported. A large number of studies used the English language version of the measures: CHU9D (n = 11), EQ-5D-Y-3L (n = 6), HUI2 (n = 22), HUI3 (n = 34).

Study Population and Condition

The majority of studies assessed a clinical population (n = 48), though some studies assessed the measure using only a general population sample (n = 15) and other studies compared general population and clinical population samples (n = 13). A wide range of conditions were covered in the studies. Conditions assessed in at least one study include asthma (n = 3), cancer (n = 5), cerebral palsy (n = 3), children born with extremely low birth weight (n = 3), chronic illness (n = 3), chronic kidney disease (n = 2), deafness (n = 2) and permanent hearing loss (n = 1), depression (n = 2), Hodgkin disease (n = 2), overweight and obesity (n = 2), stutter (n = 2), and type 1 diabetes mellitus (n = 2).

Table 1. Characteristics of included studies.

Study reference	Country	Value set	General population*	Condition, where relevant	Self-report	Proxy report	Age range of children (yr)	N
CHU9D								
Canaway, 2013[55]	UK	UK	Yes	No	Yes	No	6-7	160
Chen, 2015[56]	Australia	Australia adult	Yes	No	Yes	No	11-17	2020
Foster Page, 2015[67]	New Zealand	UK	No	Dental caries, carious surfaces, restored surfaces or missing teeth	Yes	No	6-9	87
Frew, 2015[68]	UK	UK	Yes	Underweight, healthy weight, overweight or obese	Yes	No	5-6	1344
Furber, 2015[69]	Australia	UK and Australian adolescent	No	Receiving mental health services	No	Yes	5-17	200
Oluboyede, 2019[70]	UK	UK	Yes	No	Yes	No	11-18	975
Petersen, 2018[71]	Australia	Australia adolescent	Yes	No	Yes	No	15-17	775
Ratcliffe, 2012a[58]	Australia	UK	Yes	No	Yes	No	11-17	500
Ratcliffe, 2012b[23]	Australia	Australia adolescent and adult	Yes	No	Yes	No	11-17	500
Sach, 2017[36]	UK	UK	No	Eczema	N/R	N/R	≥5	137
Stevens, 2012a[72]	Australia	UK	Yes	No	Yes	No	11-17	961
Xu, 2014[73]	China	UK and Australian adolescent	Yes	No	Yes	No	9-19	815
EQ-5D-Y-3L								
Åström, 2018[74]	Sweden	No	Yes	No	Yes	No	13-18	6574
Bergfors, 2015[64]	Sweden	No	No	Asthma	Yes	No	8-16	94
Burstrom, 2014[75]	Sweden	No	Yes	Functional motor, orthopedic and medical disabilities	Yes	No	Clinical population 7-17, general population 8-16	478
Canaway, 2013[55]	UK	UK EQ-5D	Yes	No	Yes	No	6-7	160
Chen, 2015[56]	Australia	Australia EQ-5D	Yes	No	Yes	No	11-17	2020
Eidt-Koch, 2009[76]	Germany	No	No	Cystic fibrosis	Yes	Yes	8-17	96
Hernandez, 2015[35]	Spain	France EQ-5D	No	Asthma	Yes	No	6-11	69
Hsu, 2018[77]	Taiwan	No	No	Chronic kidney disease	Yes	No	7-18	68
Jelsma, 2010[78]	South Africa	No	Yes	No	Yes	No	N/R	522
Kim, 2018[63]	South Korea	No	No	Allergic conditions	Yes	No	7-13	9949
Loof, 2019[79]	Sweden	No	Yes	Idiopathic clubfoot	Yes	Yes	8-10	215
Mayoral, 2017[37]	Unknown	Spain EQ-5D	No	Type 1 diabetes mellitus	N/R	N/R	N/R	136
Oluboyede, 2013[57]	UK	No	Yes	No	Yes	No	11-18	49
Perez-Sousa, 2018[80]	Spain	No	No	Overweight and obese	Yes	Yes	6-14	151

continued on next page

Table 1. Continued

Study reference	Country	Value set	General population*	Condition, where relevant	Self-report	Proxy report	Age range of children (yr)	N
Ravens-Sieberer, 2010[13]	Germany, Italy, South Africa, Spain, Sweden	No	Yes	No	Yes	No	≥8	2809
Robles, 2015[81]	Spain	No	Yes	No	Yes	No	8-18	923
Scalone, 2011[82]	Italy	No	Yes	Acute lymphoblastic leukemia	Yes	No	8-15	440
Scott, 2017[83]	South Africa	USA	Yes	Acutely ill or chronic health condition/ disability	Yes	No	8-12	329
Wille, 2010[14]	Germany, Italy, South Africa, Spain, Sweden	No	Yes	No	Yes	No	8-18	1987
Wong, 2019[24]	China	No	No	Adolescent or juvenile idiopathic scoliosis	Yes	No	8-17	129
HUI2								
Banks, 2008[38]	Canada	Canada	No	Cancer—undergoing chemotherapy	Yes (≥10)	Yes	2-18	29
Barr, 1997[39]	Canada	Canada	No	Cancer	Yes†	Yes	N/R	18
Belfort, 2011[40]	USA	Canada	No	Attending well-child appointments or obesity clinic	Yes	Yes	5-18	76
Boran, 2011[41]	Turkey	Canada	No	Cancer during neutropenia (adverse effect associated with cytotoxic therapy)	No	Yes	11 mo–14 yr	50
Dickerson, 2018[84]	USA	Canada	No	Depression	Yes	No	13-17	392
Feeny, 2004[42]	Canada	Canada	Yes	Extremely low weight at birth	Yes	No	12-16	275
Furlong, 2012[43]	Canada	Canada	Yes	Acute lymphoblastic leukemia	Yes	Yes	5-18	196
Glaser, 1999[85]	UK	Canada	No	Central nervous system tumor survivors	Yes	Yes	6-16	30
Kennedy, 1999[86]	UK	Canada	No	Childhood brain tumor survivors	Yes (≥16)	Yes (<16)	2-11	32
Klaassen, 2010a[87]	Canada	Canada	No	Hodgkin disease	Yes	No	8-17	51
Klaassen, 2010b[44]	Canada	Canada	No	Hodgkin disease	Yes	Yes	8.9-18	49
Kulpeng, 2013[45]	Thailand	Canada	No	Meningitis, bacteremia, pneumonia, acute otitis media, hearing loss, chronic lung disease, epilepsy, mild mental retardation	Yes (≥7 and able)	Yes	5-14	173
Le Gales, 1999[46]	France	N/A	No	Medulloblastoma and ependymoma	Yes	Yes	5-19	43
Lynch, 2016[47]	USA	Canada	Yes	Depression	Yes	No	13-17	392
Mok, 2014[48]	Hong Kong	Canada	No	Down syndrome	No	Yes	5-18	30
Morrow, 2012[49]	Australia	N/A	No	Chronic illness	Yes (≥12 and able)	Yes	5-18	131

continued on next page

Table 1. Continued

Study reference	Country	Value set	General population*	Condition, where relevant	Self-report	Proxy report	Age range of children (yr)	N
Nixon Speechley, 1999[50]	Canada	Canada	No	Childhood cancer survivors	No	Yes	7-16	250
Oluboyede, 2013[57]	UK	No	Yes	No	Yes	No	11-18	49
Petrou, 2013[51]	UK and Republic of Ireland	UK	Yes	Neurologic disability and preterm births	No	Yes	Clinical population 10 yr 1 mo–11 yr 1 mo General population 9 yr 9 mo–12 yr 3 mo	331
Ratcliffe, 2012a[58]	Australia	UK	Yes	No	Yes	No	11-17	500
Stevens, 2012b[34]	UK	UK	No	Intensive care	Yes (>11)	Yes	≥5	685
Sung, 2003[52]	Canada	Canada	No	Cancer	No	Yes	1-18	36
Sung, 2004[53]	Canada	Canada	No	Chronic illness	Yes	Yes	12-17	19
Trevino, 2013[54]	USA	Canada	Yes	Obesity	Yes	No	10-12	4979
Trudel, 1998[88]	Canada	Canada	No	Cancer	No	Yes	4-18	61
Ungar, 2012[65]	Canada	Canada	No	Asthma	Yes, solo then dyad	Yes, solo then dyad	8-17	91
HUI3								
Banks, 2008[38]	Canada	Canada	No	Cancer—undergoing chemotherapy	Yes (≥10)	Yes	2-18	29
Barr, 1997[39]	Canada	Canada	No	Cancer	Yes‡	Yes	N/R	18
Belfort, 2011[40]	USA	Canada	No	Attending well-child appointments or obesity clinic	Yes	Yes	5-18	76
Boran, 2011[41]	Turkey	Canada	No	Cancer during neutropenia (adverse effect associated with cytotoxic therapy)	No	Yes	11 mo–14 yr	50
Boulton, 2006[89]	England	Canada	No	Vision impairment or blindness	Yes	Yes	3-8	79
Cheng, 2000[90]	USA	Canada	No	Deafness	Yes	Yes	N/R	22
de Sonneville-Koedoot, 2014[91]	The Netherlands	Canada	Yes	Stutter	Yes	Yes	3-6.3	197
de Sonneville-Koedoot, 2015[92]	The Netherlands	Canada	No	Stutter	Yes	Yes	3-6.3	198
Dickerson, 2018[84]	USA	Canada	No	Depression	Yes	No	13-17	392
Feeny, 2004[42]	Canada	Canada	Yes	Extremely low weight at birth	Yes	No	12-16	275
Francis, 2019[93]	Australia New Zealand	Canada	No	Chronic kidney disease	Yes (≥13)	Yes (<13)	6-18	375
Furlong, 2012[43]	Canada	Canada	Yes	Acute lymphoblastic leukemia	Yes	Yes	5-18	196
Janse, 2008[94]	The Netherlands	Canada	No	Chronic illness—cystic fibrosis admitted for pneumonia, newly diagnosed acute lymphatic leukemia, juvenile idiopathic arthritis, or asthma	Yes	Yes	10-17	60

continued on next page

Table 1. Continued

Study reference	Country	Value set	General population*	Condition, where relevant	Self-report	Proxy report	Age range of children (yr)	N
Kennes, 2002[95]	Canada	N/A	No	Cerebral palsy	Yes	Yes	5-13	408
Klaassen, 2010a[87]	Canada	Canada	No	Hodgkin disease	Yes	No	8-17	51
Klaassen, 2010b[44]	Canada	Canada	No	Hodgkin disease	Yes	Yes	8.9-18	49
Kulkarni, 2010[96]	Canada	Canada	No	Obstructive hydrocephalus	No	Yes	5-18	47
Kulpeng, 2013[45]	Thailand	Canada	No	Meningitis, bacteremia, pneumonia, acute otitis media, hearing loss, chronic lung disease, epilepsy, mild mental retardation	Yes, (≥7 and able)	Yes	5-14	173
Le Gales, 1999[46]	France	N/A	No	Medulloblastoma and ependymoma	Yes	Yes	5-19	43
Lee, 2011[97]	USA	Canada	No	Type 1 diabetes	Yes	Yes	8-18	238
Livingston, 2008[59]	Canada	Canada	No	Cerebral palsy	No	Yes	13-20	185
Lovett, 2010[47,98]	UK	Canada	No	Deafness	No	Yes	18 mo–16 yr	50
Lynch, 2016[47]	USA	Canada	Yes	Depression	Yes	No	13-17	392
Mattera, 2018[62]	UK and USA	N/A	No	Hunter syndrome	Yes (≥12 and able)	Yes (<12 or unable)[§]	12-17	7
Mok, 2014[48]	Hong Kong	Canada	No	Down syndrome	No	Yes	5-18	30
Morrow, 2012[49]	Australia	N/A	No	Chronic illness	Yes (≥12 and able)	Yes	5-18	131
Nixon Speechley, 1999[50]	Canada	Canada	No	Childhood cancer survivors	No	Yes	7-16	250
Penn, 2011[99]	UK	Canada	Yes	Childhood brain tumors	Yes (≥8)	Yes	3-16	61
Petrou, 2013[51]	UK and Republic of Ireland	Canada	Yes	Neurological disability and preterm births	No	Yes	Clinical population 10 yr 1 mo–11 yr 1 mo General population 9 yr 9 mo–12 yr 3 mo	331
Rhodes, 2012[100]	USA	Canada	No	Adolescents with BMI ≥ 85th percentile with type 2 diabetes, prediabetes or insulin resistance	Yes	Yes	12-18	107
Roposch, 2011[101]	UK	Canada	No	Osteonecrosis secondary to treatment of developmental dysplasia of the hip	Yes	No	4-18	72
Rosenbaum, 2007[60]	Canada	Canada	No	Cerebral palsy	No	Yes	13-20	203
Smith-Olinde, 2008[102]	USA	Canada	No	Permanent hearing loss	No	Yes	5-10	103
Stade, 2006[103]	Canada	Canada	No	Children and youth prenatally exposed to alcohol, fetal alcohol spectrum disorder	Yes	Yes	8-21	126

continued on next page

Table 1. Continued

Study reference	Country	Value set	General population*	Condition, where relevant	Self-report	Proxy report	Age range of children (yr)	N
Stevens, 2012b[34]	UK	Canada	No	Intensive care	Yes (>11)	Yes	5 and older	685
Sung, 2003[52]	Canada	Canada	No	Cancer	No	Yes	1-18	36
Sung, 2004[53]	Canada	Canada	No	Chronic Illness	Yes	Yes	12-17	19
Tan, 2018[104]	Australia	Canada	No	Part of an obesity prevention intervention	No	Yes	2-5 (unclear)	368
Tilford, 2012[61]	USA	Canada	Yes	Autism spectrum disorders	No	Yes	4-17	150
Trevino, 2013[54]	USA	Canada	Yes	Obesity	Yes	No	10-12	4979
Ungar, 2012[65]	Canada	Canada	No	Asthma	Yes—solo then dyad	Yes—solo then dyad	8-17	91
Verrips, 2001[105]	Netherlands	Canada	No	Very low birth weight children	Yes	Yes	14	684
Wolke, 2013[106]	Germany	Canada	Yes	Very low birth weight and very preterm children	Yes where able	Yes	13	554

Note. Wong et al (2019) compare the EQ-5D-Y 3-level and 5-level versions.
BMI indicates body mass index; CHU9D, Child Health Utility; HUI2, Health Utilities Index Mark 2; HUI3, Health Utilities Index Mark 3; N/A, not applicable; N/R, not reported.
*General population is recorded where the sample was recruited from the general population, and studies that are included both in the general population column and in the condition column, where relevant, compare the performance of the measure in a clinical population sample to a general population sample.
†Excluded from the analysis because of low numbers.
‡These data cannot be extracted because they were merged with caregiver report up to aged 26.

Self-report and Proxy Report

Thirty studies administered the measures to the children/adolescents using only self-report, and 14 studies administered the measures using only proxy report. Twenty-seven studies used both self-report and proxy report for the same children, though for 11 of these studies restrictions were given about when self-complete was administered (eg, a minimum age or only where the child was able to self-complete), and one of the studies administered the measures separately and then as a dyad. Three studies used either self- or proxy report depending on the age of the child, and 2 studies did not report who completed the measure.

Age and Gender

Mean age varied from 6.4 years[55] to 16.[59,60] The age range of children and adolescents included in each study varied. Eleven studies included children aged younger than 5 years, which is younger than the recommended age for the measures used in these studies (note the minimum recommended age for CHU9D and EQ-5D-Y-3L is 4 and for HUI2 and HUI3 is 5).[3] The percentage of female participants in the samples ranged from 14.7%[61] to 80.6%.[24]

Sample Size

Sample size varied considerably, from 7 participants[62] to 9949.[63] Thirteen studies had sample sizes less than 50, 15 studies had sample sizes between 50 and 99, and 7 studies had sample sizes of 1000 or more.

Psychometric Performance

No study assesses all properties extracted in this review. Overall 48 studies assessed known-group validity, 33 studies assessed convergent validity, 14 studies assessed responsiveness, 24 studies assessed reliability, and 19 studies assessed acceptability and feasibility. Table 2 summarizes whether the study found evidence, mixed evidence, or no evidence for each of the properties assessed per measure, ordered within each measure by population assessed (general population, clinical population, or whether both general and clinical populations are assessed). For CHU9D the review found evidence of known-group validity, convergent validity, acceptability and feasibility, and mixed evidence of responsiveness, but the only study assessing test-retest reliability did not find evidence of reliability. For EQ-5D-Y-3L dimensions, the review found evidence of known-group validity, convergent validity, responsiveness, acceptability, and feasibility and mixed evidence of test-retest reliability, but the only study assessing interrater reliability did not find evidence of reliability. The evidence was mainly about the performance of the dimensions because there is no recommended value set. For HUI2 the review found evidence of convergent validity and test-retest reliability and mixed evidence of known-group validity, responsiveness, interrater reliability, acceptability, and feasibility because good performance was not found unanimously across these aspects of psychometric performance. For HUI3 the review found evidence of convergent validity and mixed evidence of known-group validity, responsiveness, interrater reliability, test-retest reliability, and acceptability and feasibility, with a proportion of studies not reporting evidence of known-group validity, responsiveness, or reliability.

Performance by Study Population

For CHU9D the most evidence was about performance in a general population sample (n = 8), and performance appeared similar across clinical and general population samples. For EQ-5D-Y-3L performance appeared similar across clinical and general

Table 2. Evidence of psychometric performance in included studies.

Study reference	Index, dimensions, or both assessed	General population, clinical population, or both	Sample size*	Known-group validity	Convergent validity	Responsiveness	Interrater reliability	Test-retest reliability	Acceptability and feasibility
CHU9D									
Canaway, 2013[55]	Both	General	160	✓	✓			✗	✓
Chen, 2015[56]	Both	General	2020	✓	✓				
Oluboyede, 2019[70]	Index	General	975	✓	✓				
Petersen, 2018[71]	Index	General	775	✓	✓				
Ratcliffe, 2012a[58]	Both	General	500	±	✓				
Ratcliffe, 2012b[23]	Index	General	500	✓					
Stevens, 2012a[72]	Both	General	961	✓	±				
Xu, 2014[73]	Both	General	815	✓					
Foster Page, 2015[67]	Index	Clinical	87		✓	✗			
Furber, 2015[69]	Both	Clinical	200	✓	✓				
Sach, 2017[36]	Index	Clinical	137	✓	✓	✓			
Frew, 2015[68]	Both	Both	1344	✗	±				
EQ-5D-Y-3L									
Åström, 2018[74]	Dimensions	General	6574	✓					
Canaway, 2013[55]	Both	General	160	✓	✓			✗	✓
Chen, 2015[56]	Both	General	2020	✓	✓				
Jelsma, 2010[78]	Dimensions	General	522						✓
Oluboyede, 2013[57]	Dimensions	General	49						✗
Ravens-Sieberer, 2010[13]	Dimensions	General	2809	±	✓			✓	✓
Robles, 2015[81]	Dimensions	General	923	✓					✓
Wille, 2010[14]	Dimensions	General	1987						✓
Bergfors, 2015[64]	Dimensions	Clinical	94		✓				
Eidt-Koch, 2009[76]	Dimensions	Clinical	96		✓				
Hernandez, 2015[35]	Index	Clinical	69	✓					
Hsu, 2018[77]	Dimensions	Clinical	68					±	
Kim, 2018[63]	Dimensions	Clinical	9949	✓					✓
Mayoral, 2017[37]	Both	Clinical	136	±	✓	✓			
Perez-Sousa, 2018[80]	Dimensions	Clinical	151			✓	✗		
Wong, 2019[24],	Dimensions	Clinical	129					±	

continued on next page

Table 2. Continued

Study reference	Index, dimensions, or both assessed	General population, clinical population, or both	Sample size*	Known-group validity	Convergent validity	Responsiveness	Interrater reliability	Test-retest reliability	Acceptability and feasibility
Burstrom, 2014[75]	Dimensions	Both	478	✓	✓				
Loof, 2019[79]	Dimensions	Both	215	±					
Scalone, 2011[82]	Dimensions	Both	440	±	✓			✓	✓
Scott, 2017[83]	Both	Both	329	✓	±	✓		±	✓
HUI2									
Oluboyede, 2013[57]	Dimensions	General	49						✗
Ratcliffe, 2012a[58]	Both	General	500	±	✓				
Trevino, 2013[54]	Index	General	4979	±					
Banks, 2008[38]	Index	Clinical	29		±	✗			
Barr, 1997[39]	Both	Clinical	18			✓			
Belfort, 2011[40]	Both	Clinical	76	✗					
Boran, 2011[41]	Both	Clinical	50	✓		✓			
Dickerson, 2018[84]	Index	Clinical	392		✓	✓			
Glaser, 1999[85]	Index	Clinical	30				✓		✓
Kennedy, 1999[86]	Index	Clinical	32	±					
Klaassen, 2010a[87]	Index	Clinical	51		✓	±			
Klaassen, 2010b[44]	Index	Clinical	49				±		
Kulpeng, 2013[45]	Index	Clinical	173		✓		✓		
Le Gales, 1999[46]	Dimensions	Clinical	43	✗					✓
Lynch, 2016[47]	Index	Clinical	392	✓					
Mok, 2014[48]	Index	Clinical	30	✓					[1]
Morrow, 2012[49]	Dimensions	Clinical	131				✗		
Nixon Speechley, 1999[50]	Both	Clinical	250		✓				
Stevens, 2012b[34]	Index	Clinical	685	[†]		[†]	✓		✓
Sung, 2003[52]	Both	Clinical	36		±				✓
Sung, 2004[53]	Index	Clinical	19				✗		
Trudel, 1998[88]	Both	Clinical	61	✓	✓			✓	
Ungar, 2012[65]	Both	Clinical	91	✗	✓	✗	✗	✓	
Feeny, 2004[42]	Index	Both	275	✓	✓				

continued on next page

Table 2. Continued

Study reference	Index, dimensions, or both assessed	General population, clinical population, or both	Sample size*	Known-group validity	Convergent validity	Responsiveness	Interrater reliability	Test-retest reliability	Acceptability and feasibility
Furlong, 2012[43]	Index	Both	196	✓					✗
Petrou, 2013[51]	Both	Both	331	✓					
HUI3									
Trevino, 2013[54]	Index	General	4979	±					
Banks, 2008[38]	Index	Clinical	29		±	✗			
Barr, 1997[39]	Both	Clinical	18			✓			✓
Belfort, 2011[40]	Both	Clinical	76	✗			±		
Boran, 2011[41]	Both	Clinical	50	±		±			
Boulton, 2006[89]	Both	Clinical	79	✓					
Cheng, 2000[90]	Index	Clinical	22			✓			
de Sonneville-Koedoot, 2015[92]	Index	Clinical	198			✗			
Dickerson, 2018[84]	Index	Clinical	392		✓	✓			
Francis, 2019[93]	Both	Clinical	375	✓					
Janse, 2008[94]	Both	Clinical	60				✗		
Kennes, 2002[95]	Dimensions	Clinical	408		✓				
Klaassen, 2010a[87]	Index	Clinical	51		✓	±			
Klaassen, 2010b[44]	Index	Clinical	49				±		
Kulkarni, 2010[96]	Index	Clinical	47		✓				
Kulpeng, 2013[45]	Index	Clinical	173		✓		✓		
Le Gales, 1999[46]	Dimensions	Clinical	43	✗			✗		✓
Lee, 2011[97]	Index	Clinical	238				✓	✓	✓
Livingston, 2008[59]	Index	Clinical	185		✗				
Lovett, 2010[47,98]	Index	Clinical	50	✗					
Mattera, 2018[62]	Dimensions	Clinical	7		✓				
Mok, 2014[48]	Index	Clinical	30	✓					‡
Morrow, 2012[49]	Dimensions	Clinical	131				✗		
Nixon Speechley, 1999[50]	Both	Clinical	250		✓				

continued on next page

Table 2. Continued

Study reference	Index, dimensions, or both assessed	General population, clinical population, or both	Sample size*	Known-group validity	Convergent validity	Responsiveness	Interrater reliability	Test-retest reliability	Acceptability and feasibility
Rhodes, 2012[100]	Index	Clinical	107	✗	✓		±		
Roposch, 2011[101]	Both	Clinical	72	±					✗
Rosenbaum, 2007[60]	Both	Clinical	203	±	✗				
Smith-Olinde, 2008[102]	Both	Clinical	103	✗					✓
Stade, 2006[103]	Both	Clinical	126	±			✓		
Stevens, 2012b[34]	Index	Clinical	685	[1]		[†]	✓		✓
Sung, 2003[52]	Both	Clinical	36		±				✓
Sung, 2004[53]	Index	Clinical	19				✗		
Tan, 2018[104]	Both	Clinical	368	✗					
Ungar, 2012[65]	Both	Clinical	91	✗	✓	✗	✗	✗	
Verrips, 2001[105]	Both	Clinical	684				✗		
de Sonneville-Koedoot, 2014[91]	Index	Both	197	✓					
Feeny, 2004[42]	Index	Both	275	✓	✓				
Furlong, 2012[43]	Index	Both	196	✓					✗
Lynch, 2016[47]	Index	Both	392	✓					
Penn, 2011[99]	Both	Both	61	✓			±		
Petrou, 2013[51]	Both	Both	331	✓					
Tilford, 2012[61]	Both	Both	150	✓	✓				
Wolke, 2013[106]	Dimensions	Both	554	✓					

Note. Wong et al (2019) compare the EQ-5D-Y 3-level and 5-level versions.
CHU9D indicates Child Health Utility; HUI2, Health Utilities Index Mark 2; HUI3, Health Utilities Index Mark 3.
✓ Evidence indicated significant performance. ✗ Property was examined but no significant evidence was found. ± Evidence was mixed or inconclusive evidence was found.
*Sample size was reported for the reported sample assessed in the paper and may differ within the analyses reported.
[†]Property assessed but results not reported.

population samples, with the same number of studies in each population (n = 8 each, n = 4 for both clinical and general population), though note that acceptability and feasibility was mainly examined using a general population sample (6 of 9 samples assessing feasibility). For HUI2 the majority of evidence was from studies with clinical samples (n = 20), and the 3 studies using a general population sample did not find evidence of good psychometric performance. For HUI3 the majority of evidence was from studies with clinical samples (n = 34), with only 1 study with a general population sample, though studies with both general and clinical population samples (n = 8) found evidence of known-group validity and convergence validity in all studies where this was assessed.

Summary of Psychometric Performance

Table 3 summarizes the results of all analyses. The number of entries reflects the number of studies where each psychometric property is assessed. EQ-5D-Y-3L has the largest amount of evidence of good psychometric performance in proportion to the number of studies that have examined its psychometric

Table 3. Summary of psychometric performance by measure and utility index (ie, country value set).

	Dimensions or utility index (ie, country value set)	Known group validity	Convergent validity	Responsiveness	Interrater reliability	Test-retest reliability	Inter-modality reliability	Acceptability and feasibility
CHU9D	Dimensions	✓✓✗	✓✓✓✓✓±✗			✗		✓
	Australian adolescent value set	✓✓✓±	✓✓✓					
	Australian adult value set	✓±	✓✓					
	UK value set	✓✓✓✓✓✓✓±✗	✓✓✓✓	✓✗				
EQ-5D-Y-3L	Dimensions	✓✓✓✓✓±±±	✓✓✓✓✓✓✓✓✓±	✓	✗	✓✓±±±✗	✓	✓✓✓✓✓✓✓✗
	EQ-5D UK value set	✓	✓					
	Australian EQ-5D value set	✓						
	French EQ-5D value set	✓						
	Spanish EQ-5D value set	±	✓	✓				
	US EQ-5D-Y value set							
HUI2	Dimensions	✓✓✗✗	✓✓✓	✓✓✗	✓✗✗✗	✓		✓✓✓✓✗✗
	Canadian value set	✓✓✓✓✓✓✓±✗✗	✓✓✓✓✓✓✓✓±±	✓✓✓±✗✗	✓✓±✗✗	✓✓		
	UK value set	✓±	✓		✓			
HUI3	Dimensions	✓✓✓✓✓✓±±±±±✗✗	✓✓✓✓±✗	✓✓±✗	✓±±±✗✗✗✗✗		✗	
	Canadian value set	✓✓✓✓✓✓✓✓✓✓✓✓±±±✗✗✗✗✗✗	✓✓✓✓✓✓✓✓✓✓±±✗	✓✓✓✓±✗✗✗	✓✓✓✓✓✓±✗✗	✓✗		✓✓✓✓✓✓✗✗

Note. ✓ Evidence indicating significant performance. ✗ Property was examined but no significant evidence was found. ± Evidence was mixed or inconclusive evidence was found. Each symbol represents the findings of one study assessing that psychometric property. Where studies assessed multiple psychometric properties, a symbol is recorded for each psychometric property assessed.
CHU9D indicates Child Health Utility; HUI2, Health Utilities Index Mark 2; HUI3, Health Utilities Index Mark 3.

performance (note this is for dimensions). The CHU9D was assessed in fewer studies, but the majority of studies found evidence of good psychometric performance. The evidence for HUI2 and HUI3 was more mixed. More detailed results are available in the online Supplemental Table (found at https://doi.org/10.1016/j.jval.2020.09.012), where studies can also be separated by condition.

Discussion

The review outlined the evidence about the psychometric performance of the child and adolescent-specific measures of CHU9D, EQ-5D-Y-3L, HUI2, and HUI3. The heterogeneity of published studies means that the evidence was based on studies across a range of countries, populations, and conditions, using different study designs, different languages, different value sets, and different statistical techniques. The wide variation in studies makes it difficult to synthesize the evidence to generate a consistent picture of the overall performance of each measure. The evidence included several studies with small sample sizes that may not have been powered to detect statistical significance, and there were only a relatively small number of studies within the same condition. Evidence for CHU9D was based on a limited number of studies (n = 12), as well as evidence assessing responsiveness (n = 14). Only HUI2 performed strongly for test-retest reliability. None of the measures performed strongly for interrater reliability between child self-report and parent proxy report (though CHU9D was not assessed), suggesting there is reason for concern about the comparability of self-report and proxy report responses to measure HRQL of children and adolescents.

More studies assessed the psychometric performance of HUI3 than the other measures, but the evidence of HUI3 was more mixed. This means that for HUI3 there was a large number of studies finding evidence of good psychometric performance, but the proportion of studies that did not find evidence of good psychometric performance was larger than for the other measures. HUI2 was also assessed in a large number of studies, though the performance was mixed. In contrast, EQ-5D-Y-3L and CHU9D were assessed in fewer studies, but the proportion of studies that found evidence of good psychometric performance was larger.

For EQ-5D-Y-3L there were the same number of studies assessing performance in general population and clinical population samples, and performance was similar across these different samples. Although the performance of EQ-5D-Y-3L was assessed by a similar number of studies with clinical samples and general population samples, CHU9D was mainly assessed in general

population samples and HUI2 and HUI3 were mainly assessed in clinical population samples. To inform resource allocation decisions using health technology assessment, it is important that these generic preference-based measures have good performance across different conditions; however, it is also important that these generic preference-based measures have good performance in general population samples to enable accurate evaluations of the health of general population samples. Therefore evidence across both clinical samples and general population samples is important, and it does not follow that good psychometric performance in one condition or even in clinical population samples necessarily means that the preference-based measure also has good psychometric performance in general population samples.

For EQ-5D-Y-3L there is no official value set, and the good psychometric performance that was identified is based mainly on the performance on the dimensions. Although it could be anticipated that a utility index would have the same psychometric performance, this can only be confirmed through data analyses. A value set may not have sufficiently large differences in utility decrements for different severity levels of each dimension.

Few studies assessed measures within the same clinical area; however, even where there were multiple studies within a clinical area the evidence was limited. For example, 3 studies assessed the performance of measures in patients with asthma: 2 assessed EQ-5D-Y-3L[35,64] and 1 assessed HUI2 and HUI3.[65] EQ-5D-Y-3L was found to have known-group validity and convergent validity, with no assessment of responsiveness, reliability, acceptability, or feasibility. HUI2 and HUI3 were found to have convergent validity, but the study assessed and found no evidence for known-group validity, responsiveness, or interrater reliability and found evidence of test-retest reliability for HUI2 but not HUI3. On the basis of these findings it is difficult to recommend use of either measure over the other because for EQ-5D-Y-3L there is limited evidence available but the evidence that is available suggests good performance, whereas for HUI2 and HUI3 there is wider evidence available but the evidence is mixed. Equally, although the evidence is mixed, it is difficult to determine whether this is affected by the sample size of 91. Differences in samples may also potentially affect results.

Some studies had small sample sizes, with 28 out of the 76 studies having a sample size less than 100. Sample size was not been used to assess the studies, but it should be taken into consideration that some studies may not have found significant evidence of the psychometric performance because of the sample size, meaning that the result may not be indicative of the performance of the measure. In particular for HUI2 and HUI3 this may have affected the results because for HUI2 15 of 26 studies assessing performance had sample sizes less than 100 and for HUI3 18 of 43 studies had sample sizes less than 100. In the literature there are no clear guidelines or accepted practice about how to generate sample sizes for studies assessing psychometric performance of patient-reported outcome measures[66] nor, to our knowledge, preference-based measures.

The review assessed the performance of child- and adolescent-specific preference-based measures deemed appropriate for informing UK public policy and hence did not assess the performance of all child- and adolescent-specific preference-based measures. The literature search included search terms to identify studies assessing AQoL-6D and EQ-5D-Y-5L but identified only 1 relevant study for each measure. The one study assessing AQoL-6D examined known-group validity, for which evidence was found.[23] The one study assessing EQ-5D-Y-5L examined test-retest reliability of the dimensions, for which evidence was found.[24] The review did not include Adolescent Health Utility Measure, AQoL-6D, EQ-5D-Y-5L, 15D, 16D, 17D, or QWB, and it is recommended that the psychometric performance of these measures are reviewed to inform researchers and policy makers worldwide in their selection of appropriate child- and adolescent-specific preference-based measures.

Methodologic limitations of the review include potential missing studies of child and adolescent preference-based measures in mixed adolescent and adult populations because of the pediatrics filter applied in the database search. Statistical mapping analyses were not included in the review because mapping assessments are undertaken to generate predictions rather than assess association per se, though it is recognized that mapping analyses can provide some evidence of associations between measures. Content validity of the measures was not assessed as part of the review, though it is recognized that this is an important consideration when deciding which measure to use. Studies that administered 1 or more measures and summarized their results were not included in the review unless they assessed psychometric properties. It is possible that some relevant studies may not have been captured in our search if they did not use the vocabulary of responsiveness or validity, for example, even though they reported change over time or difference across treatments. Appropriateness of the statistical analyses undertaken was also not assessed. In addition, psychometric performance data were single extracted rather than double extracted (with the exception of 3 studies). Quality assessment of the studies was not undertaken, though we have reported where there was a small sample size.

Conclusion

The review of published evidence on the psychometric performance of a selection of child- and adolescent-specific generic preference-based measures found that EQ-5D-Y-3L has the largest proportion of evidence of good psychometric performance out of the studies that have examined its psychometric performance, followed by CHU9D; however, the majority of the evidence related to EQ-5D-Y-3L is based on dimensions, and the same psychometric performance of any utility index is not guaranteed. This review enables researchers and policy makers to identify critical evidence gaps in the performance of these measures, and it is hoped that knowing this will encourage collection of this psychometric evidence. Further research is recommended to provide greater evidence both on psychometric performance of the measure (dimensions and utilities) administered by self-report and proxy report and assessed by the age of the child. The review tables are informative in indicating patient populations where the psychometric performance of 1 or more measures was assessed and providing an overview of the evidence found. Concerns were raised about the comparability of self-report and proxy responses to measure HRQL of children and adolescents.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2020.09.012.

Article and Author Information

Accepted for Publication: September 25, 2020

Published Online: Month xx, 2020

doi: https://doi.org/10.1016/j.jval.2020.09.012

Author Affiliation: School of Health and Related Research, University of Sheffield, Sheffield, England, UK.

Author Contributions: DR lead the study and wrote the first version of the manuscript. RW undertook the searches, AK, DR and EP undertook data sifting and extraction. All authors contributed to study design, interpretation of results and the manuscript. All authors approved the final manuscript.

Conflict of Interest Disclosures: The authors have reported no conflicts of interest.

Funding/Support: This research was funded by National Institute of Health and Care Excellence (NICE) Decision Support Unit (NICE DSU) and the National Institute for Health Research (NIHR) Policy Research Programme, conducted through the Policy Research Unit in Economic Methods of Evaluation in Health and Social Care Interventions, PR-PRU-1217-20401. The views expressed are those of the author(s) and not necessarily those of NICE, the NIHR, or the Department of Health and Social Care.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The authors would like to thank Sophie Cooper, Sarah Davis, Alan Lamb, Rosie Lovett, and Tracey Young for commenting on previous drafts. They would also like to thank Donna Davis and Liz Metham for project management.

REFERENCES

1. Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *Eur J Health Econ*. 2018;19:557–570.
2. Chen G, Ratcliffe J. A review of the development and application of generic multi-attribute utility instruments for paediatric populations. *Pharmacoeconomics*. 2015;33:1013–1028.
3. Rowen D, Rivero-Arias O, Devlin N, et al. Review of valuation methods of preference-based measures of health for economic evaluation in child and adolescent populations: where are we now and where are we going. *Pharmacoeconomics*. 2020;38:325–340.
4. Kwon J, Kim SW, Ungar WJ, et al. A systematic review and meta-analysis of childhood health utilities. *Med Decis Making*. 2018;38:277–305.
5. Stevens K. Valuation of the Child Health Utility 9D Index. *Pharmacoeconomics*. 2012;30:729–747.
6. Ratcliffe J, Couzner L, Flynn T, et al. Valuing Child Health Utility 9D health states with a young adolescent sample: a feasibility study to compare best-worst scaling discrete-choice experiment, standard gamble and time trade-off methods. *Appl Health Econ Health Policy*. 2011;9:15–27.
7. Ratcliffe J, Flynn T, Terlich F, et al. Developing adolescent-specific health state values for economic evaluation: an application of profile case best-worst scaling to the Child Health Utility 9D. *Pharmacoeconomics*. 2012;30:713–727.
8. Ratcliffe J, Chen G, Stevens K, et al. Valuing Child Health Utility 9D health states with young adults: insights from a time trade off study. *Appl Health Econ Health Policy*. 2015;13:485–492.
9. Ratcliffe J, Huynh E, Chen G, et al. Valuing the Child Health Utility 9D: using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. *Soc Sci Med*. 2016;157:48–59.
10. Rowen D, Mulhern B, Stevens K, et al. Estimating a Dutch value set for the pediatric preference-based CHU9D using a discrete choice experiment with duration. *Value Health*. 2018;21:1234–1242.
11. Chen G, Xu F, Huynh E, et al. Scoring the Child Health Utility 9D instrument: estimation of a Chinese child and adolescent-specific tariff. *Qual Life Res*. 2019;28:163–176.
12. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: past, present and future. *Appl Health Econ Health Policy*. 2017;15:127–137.
13. Ravens-Sieberer U, Wille N, Badia X, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res*. 2010;19:887–897.
14. 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;19:875–886.
15. Hu RL. [Connected pathways of the thalamus and clinical manifestations of lesions of its various parts]. *Zhonghua Shen Jing Jing Shen Ke Za Zhi*. 1988;21:362–365.
16. Tatsuoka H, Reese TS. New structural features of synapses in the anteroventral cochlear nucleus prepared by direct freezing and freeze-substitution. *J Comp Neurol*. 1989;290:343–357.
17. Kind P, Klose K, Gusi N, et al. Can adult weights be used to value child health states? Testing the influence of perspective in valuing EQ-5D-Y. *Qual Life Res*. 2015;24:2519–2539.
18. Kreimeier S, Oppe M, Ramos-Goñi JM, et al. Valuation of EuroQol Five-Dimensional Questionnaire, Youth Version (EQ-5D-Y) and EuroQol Five-Dimensional Questionnaire, Three-Level Version (EQ-5D-3L) Health States: the impact of wording and perspective. *Value Health*. 2018;21:1291–1298.
19. Ramos-Goñi JM, Oppe M, Stolk E, et al. International valuation protocol for the EQ-5D-Y-3L. *Pharmacoeconomics*. 2020;38:653–663.
20. Torrance GW, Feeny DH, Furlong WJ, et al. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care*. 1996;34:702–722.
21. McCabe C, Stevens K, Roberts J, et al. Health state values for the HUI 2 descriptive system: results from a UK survey. *Health Econ*. 2005;14:231–244.
22. 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;40:113–128.
23. Ratcliffe J, Stevens K, Flynn T, et al. Whose values in health? An empirical comparison of the application of adolescent and adult values for the CHU-9D and AQOL-6D in the Australian adolescent general population. *Value Health*. 2012;15:730–736.
24. Wong CKH, Cheung PWH, Luo N, et al. A head-to-head comparison of five-level (EQ-5D-5L-Y) and three-level EQ-5D-Y questionnaires in paediatric patients. *Eur J Health Econ*. 2019;20:647–656.
25. 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;63:737–745.
26. Longworth L, Yang Y, Young T, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technol Assess*. 2014;18:1–224.
27. National Institute of Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal 2013*. London, UK: NICE; 2013.
28. Davidson M. Known-groups validity. In: Michalos A, ed. *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht, The Netherlands: Springer; 2014.
29. Chin C, Yao G. Convergent validity. In: Michalos A, ed. *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht, The Netherlands: Springer; 2014.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
31. Terwee C. Responsiveness to change. In: Michalos A, ed. *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht, The Netherlands: Springer; 2014.
32. Vilagut G. Test-retest reliability. In: Michalos A, ed. *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht, The Netherlands: Springer; 2014.
33. Kimel M, Revicki D. Inter-rater reliability. In: Michalos A, ed. *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht, The Netherlands: Springer; 2014.
34. Stevens KJ, Freeman JV. An assessment of the psychometric performance of the Health Utilities Index 2 and 3 in children following discharge from a U.K. pediatric intensive care unit. *Pediatr Crit Care Med*. 2012;13(4):387–392.
35. Hernandez G, Avila M, Garin O, et al. Validity of the EQ-5D-Y in a cohort of asthmatic children in Europe. *Qual Life Res*. 2015;24, 174-174.
36. Sach T, McManus E, Thomas K, Montgomery A, Harrison E, Williams H. on behalf of the CLOTHES trial team. The comparative performance of the CHU-9D and the ADQoL amongst children aged 5 years or more with eczema: evidence from the CLOTHES randomised controlled trial. *Value Health*. 2017;20(9). A406-A406.
37. Mayoral K, Rajmil L, Murillo M, et al. Validity of the EQ-5D-Y in children and adolescents with diabetes. *Qual Life Res*. 2017;26:71.
38. Banks BA, Barrowman NJ, Klaassen R. Health-related quality of life: changes in children undergoing chemotherapy. *J Pediatr Hematol Oncol*. 2008;30:292–297.
39. Barr R, Petrie C, Furlong W, et al. Health-related quality of life during post-induction chemotherapy in children with acute lymphoblastic leukemia in remission. *Int J Oncol*. 1997;11:333–339.
40. Belfort MB, Zupancic JA, Riera KM, et al. Health state preferences associated with weight status in children and adolescents. *BMC Pediatr*. 2011;11:12.
41. Boran P, Horsman J, Tokuc G, et al. Translation and cultural adaptation of health utilities index with application to pediatric oncology patients during neutropenia and recovery in Turkey. *Pediatr Blood Cancer*. 2011;56:812–817.
42. Feeny D, Furlong W, Saigal S, et al. Comparing directly measured standard gamble scores to HUI2 and HUI3 utility scores: group- and individual-level comparisons. *Soc Sci Med*. 2004;58:799–809.
43. Furlong W, Rae C, Feeny D, et al. Health-related quality of life among children with acute lymphoblastic leukemia. *Pediatr Blood Cancer*. 2012;59:717–724.
44. Klaassen RJ, Barr RD, Hughes J, et al. Nurses provide valuable proxy assessment of the health-related quality of life of children with Hodgkin disease. *Cancer*. 2010;116:1602–1607.
45. Kulpeng W, Sornsrivichai V, Chongsuvivatwong V, et al. Variation of health-related quality of life assessed by caregivers and patients affected by severe childhood infections. *BMC Pediatr*. 2013;13:122.
46. Le Galès C, Costet N, Gentet JC, et al. Cross-cultural adaptation of a health status classification system in children with cancer. First results of the French adaptation of the Health Utilities Index Marks 2 and 3. *Int J Cancer Suppl*. 1999;12:112–118.

47. Lynch FL, Dickerson JF, Feeny DH, et al. Measuring health-related quality of life in teens with and without depression. *Med Care*. 2016;54:1089–1097.
48. Mok WK, Wong WH, Mok GT, et al. Validation and application of health utilities index in Chinese subjects with down syndrome. *Health Qual Life Outcomes*. 2014;12:144.
49. Morrow AM, Hayen A, Quine S, et al. A comparison of doctors', parents' and children's reports of health states and health-related quality of life in children with chronic conditions. *Child Care Health Dev*. 2012;38:186–195.
50. Nixon Speechley K, Maunsell E, Desmeules M, 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.
51. Petrou S, Johnson S, Wolke D, et al. The association between neurodevelopmental disability and economic outcomes during mid-childhood. *Child Care Health Dev*. 2013;39:345–357.
52. Sung L, Greenberg ML, Doyle JJ, et al. Construct validation of the Health Utilities Index and the Child Health Questionnaire in children undergoing cancer chemotherapy. *Br J Cancer*. 2003;88:1185–1190.
53. Sung L, Young NL, Greenberg ML, et al. Health-related quality of life (HRQL) scores reported from parents and their children with chronic illness differed depending on utility elicitation method. *J Clin Epidemiol*. 2004;57:1161–1166.
54. Treviño RP, Pham TH, Edelstein SL. Obesity and preference-weighted quality of life of ethnically diverse middle school children: the HEALTHY study. *J Obes*. 2013;2013:206074.
55. Canaway AG, Frew EJ. 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;22:173–183.
56. Chen G, Flynn T, Stevens K, et al. Assessing the health-related quality of life of Australian adolescents: an empirical comparison of the Child Health Utility 9D and EQ-5D-Y Instruments. *Value Health*. 2015;18:432–438.
57. Oluboyede Y, Tubeuf S, McCabe C. Measuring health outcomes of adolescents: report from a pilot study. *Eur J Health Econ*. 2013;14:11–19.
58. Ratcliffe J, Stevens K, Flynn T, et al. An assessment of the construct validity of the CHU9D in the Australian adolescent general population. *Qual Life Res*. 2012;21:717–725.
59. Livingston MH, Rosenbaum PL. Adolescents with cerebral palsy: stability in measurement of quality of life and health-related quality of life over 1 year. *Dev Med Child Neurol*. 2008;50:696–701.
60. Rosenbaum PL, Livingston MH, Palisano RJ, et al. Quality of life and health-related quality of life of adolescents with cerebral palsy. *Dev Med Child Neurol*. 2007;49:516–521.
61. Tilford JM, Payakachat N, Kovacs E, et al. Preference-based health-related quality-of-life outcomes in children with autism spectrum disorders: a comparison of generic instruments. *Pharmacoeconomics*. 2012;30:661–679.
62. Mattera M, Vernon MK, Raluy-Callado M, et al. Validation of the shortened Hunter Syndrome-Functional Outcomes for Clinical Understanding Scale (HS-FOCUS). *Health Qual Life Outcomes*. 2018;16:209.
63. Kim SK, Jo MW, Kim SH. Health-related quality of life by allergy symptoms in elementary school students. *Health Qual Life Outcomes*. 2018;16:93.
64. Bergfors S, Åström M, Burström K, et al. Measuring health-related quality of life with the EQ-5D-Y instrument in children and adolescents with asthma. *Acta Paediatr*. 2015;104:167–173.
65. Ungar WJ, Boydell K, Dell S, et al. A parent-child dyad approach to the assessment of health status and health-related quality of life in children with asthma. *Pharmacoeconomics*. 2012;30:697–712.
66. Anthoine E, Moret L, Regnault A, et al. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes*. 2014;12:176.
67. Foster Page LA, Beckett DM, Cameron CM, et al. Can the Child Health Utility 9D measure be useful in oral health research. *Int J Paediatr Dent*. 2015;25:349–357.
68. Frew EJ, Pallan M, Lancashire E, et al. Is utility-based quality of life associated with overweight in children? Evidence from the UK WAVES randomised controlled study. *BMC Pediatr*. 2015;15:211.
69. Furber G, Segal L. The validity of the Child Health Utility instrument (CHU9D) as a routine outcome measure for use in child and adolescent mental health services. *Health Qual Life Outcomes*. 2015;13:22.
70. Oluboyede Y, Tomos R. Measuring weight-specific quality of life in adolescents: an examination of the concurrent validity and test-retest reliability of the WAItE. *Value Health*. 2019;22(3):348–354.
71. Petersen K, Chen G, Mpundu-Kaambwa C, et al. Measuring health-related quality of life in adolescent populations: an empirical comparison of the CHU9D and the PedsQLTM 4.0 Short Form 15. *Patient*. 2018;11(1):29–37.
72. Stevens K, Ratcliffe J. Measuring and valuing health benefits for economic evaluation in adolescence: an assessment of the practicality and validity of the child health utility 9D in the Australian adolescent population. *Value Health*. 2012;15:1092–1099.
73. Xu F, Chen G, Stevens K, et al. Measuring and valuing health-related quality of life among children and adolescents in mainland China—a pilot study. *PLoS One*. 2014;9:e89222.
74. Åström M, Persson C, Lindén-Boström M, et al. Population health status based on the EQ-5D-Y-3L among adolescents in Sweden: results by sociodemographic factors and self-reported comorbidity. *Qual Life Res*. 2018;27:2859–2871.
75. Burström K, Bartonek Å, Broström EW, et al. EQ-5D-Y as a health-related quality of life measure in children and adolescents with functional disability in Sweden: testing feasibility and validity. *Acta Paediatr*. 2014;103:426–435.
76. Eidt-Koch D, Mittendorf T, Greiner W. Cross-sectional validity of the EQ-5D-Y as a generic health outcome instrument in children and adolescents with cystic fibrosis in Germany. *BMC Pediatr*. 2009;9:55.
77. Hsu C, Lin HW, Pickard AS, Tain YL. EQ-5D-Y for the assessment of health-related quality of life among Taiwanese youth with mild-to-moderate chronic kidney disease. *Int J Qual Health Care*. 2018;30:298–305.
78. Jelsma J. A comparison of the performance of the EQ-5D and the EQ-5D-Y health-related quality of life instruments in South African children. *Int J Rehabil Res*. 2010;33:172–177.
79. Lööf E, Andriesse H, Broström EW, et al. Neurodevelopmental difficulties negatively affect health-related quality of life in children with idiopathic clubfoot. *Acta Paediatr*. 2019;108:1492–1498.
80. Perez-Sousa MA, Olivares PR, Garcia-Hermoso A, et al. Does anthropometric and fitness parameters mediate the effect of exercise on the HRQoL of overweight and obese children/adolescents. *Qual Life Res*. 2018;27:2305–2312.
81. Robles N, Rajmil L, Rodriguez-Arjona D, et al. Development of the web-based Spanish and Catalan versions of the Euroqol 5D-Y (EQ-5D-Y) and comparison of results with the paper version. *Health Qual Life Outcomes*. 2015;13:72.
82. Scalone L, Tomasetto C, Matteucci MC, et al. Assessing quality of life in children and adolescents: development and validation of the Italian version of the EQ-5D-Y. *Ital J Public Health*. 2011;8:331–341.
83. Scott D, Ferguson GD, Jelsma J. The use of the EQ-5D-Y health related quality of life outcome measure in children in the Western Cape, South Africa: psychometric properties, feasibility and usefulness—a longitudinal, analytical study. *Health Qual Life Outcomes*. 2017;15:12.
84. Dickerson JF, Feeny DH, Clarke GN, et al. Evidence on the longitudinal construct validity of major generic and utility measures of health-related quality of life in teens with depression. *Qual Life Res*. 2018;27:447–454.
85. Glaser AW, Furlong W, Walker DA, et al. Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the U.K. *Eur J Cancer*. 1999;35:256–261.
86. Kennedy CR, Leyland K. Comparison of screening instruments for disability and emotional/behavioral disorders with a generic measure of health-related quality of life in survivors of childhood brain tumors. *Int J Cancer Suppl*. 1999;12:106–111.
87. Klaassen RJ, Krahn M, Gaboury I, et al. Evaluating the ability to detect change of health-related quality of life in children with Hodgkin disease. *Cancer*. 2010;116:1608–1614.
88. Trudel JG, Rivard M, Dobkin PL, et al. Psychometric properties of the Health Utilities Index Mark 2 system in paediatric oncology patients. *Qual Life Res*. 1998;7:421–432.
89. Boulton M, Haines L, Smyth D, et al. Health-related quality of life of children with vision impairment or blindness. *Dev Med Child Neurol*. 2006;48:656–661.
90. Cheng AK, Rubin HR, Powe NR, et al. Cost-utility analysis of the cochlear implant in children. *JAMA*. 2000;284:850–856.
91. de Sonneville-Koedoot C, Stolk EA, Raat H, et al. Health-related quality of life of preschool children who stutter. *J Fluency Disord*. 2014;42:1–12.
92. de Sonneville-Koedoot C, Bouwmans C, Franken MC, et al. Economic evaluation of stuttering treatment in preschool children: the RESTART-study. *J Commun Disord*. 2015;58:106–118.
93. Francis A, Didsbury MS, van Zwieten A, et al. Quality of life of children and adolescents with chronic kidney disease: a cross-sectional study. *Arch Dis Child*. 2019;104:134–140.
94. Janse AJ, Sinnema G, Uiterwaal CS, et al. Quality of life in chronic illness: children, parents and paediatricians have different, but stable perceptions. *Acta Paediatr*. 2008;97:1118–1124.
95. Kennes J, Rosenbaum P, Hanna SE, et al. Health status of school-aged children with cerebral palsy: information from a population-based sample. *Dev Med Child Neurol*. 2002;44:240–247.
96. Kulkarni AV, Shams I, Cochrane DD, et al. Quality of life after endoscopic third ventriculostomy and cerebrospinal fluid shunting: an adjusted multivariable analysis in a large cohort. *J Neurosurg Pediatr*. 2010;6:11–16.
97. Lee JM, Rhee K, O'Grady MJ, et al. Health utilities for children and adults with type 1 diabetes. *Med Care*. 2011;49:924–931.
98. Lovett RE, Vickers DA, Summerfield AQ. Bilateral cochlear implantation for hearing-impaired children: criterion of candidacy derived from an observational study. *Ear Hear*. 2015;36:14–23.
99. Penn A, Lowis SP, Stevens MC, et al. A detailed prospective longitudinal assessment of health status in children with brain tumors in the first year after diagnosis. *J Pediatr Hematol Oncol*. 2011;33:592–599.
100. Rhodes ET, Goran MI, Lieu TA, et al. Health-related quality of life in adolescents with or at risk for type 2 diabetes mellitus. *J Pediatr*. 2012;160:911–917.
101. Roposch A, Liu LQ, Offiah AC, et al. Functional outcomes in children with osteonecrosis secondary to treatment of developmental dysplasia of the hip. *J Bone Joint Surg Am*. 2011;93:e145.
102. Smith-Olinde L, Grosse SD, Olinde F, et al. Health state preference scores for children with permanent childhood hearing loss: a comparative analysis of the QWB and HUI3. *Qual Life Res*. 2008;17:943–953.

103. Stade BC, Stevens B, Ungar WJ, et al. Health-related quality of life of Canadian children and youth prenatally exposed to alcohol. *Health Qual Life Outcomes*. 2006;4:81.
104. Tan EJ, Brown V, Petrou S, et al. Is there an association between early weight status and utility-based health-related quality of life in young children. *Qual Life Res*. 2018;27:2851–2858.
105. Verrips GH, Stuifbergen MC, den Ouden AL, et al. Measuring health status using the Health Utilities Index: agreement between raters and between modalities of administration. *J Clin Epidemiol*. 2001;54:475–481.
106. Wolke D, Chernova J, Eryigit-Madzwamuse S, et al. Self and parent perspectives on health-related quality of life of adolescents born very preterm. *J Pediatr*. 2013;163:1020–1026.e2.