



# Psychometric Evaluation of the PedsQL GCS and CHU9D in Australian Children and Adolescents with Common Chronic Health Conditions

Rakhee Raghunandan<sup>1,2</sup> · Kirsten Howard<sup>1,2</sup> · Sarah Smith<sup>3</sup> · Anagha Killedar<sup>2</sup> · Erin Cvejic<sup>1</sup> · Martin Howell<sup>1,2</sup> · Stavros Petrou<sup>4</sup> · Emily Lancsar<sup>5</sup> · Germaine Wong<sup>1</sup> · Jonathan Craig<sup>6</sup> · Alison Hayes<sup>1</sup>

Accepted: 13 September 2023 / Published online: 3 October 2023  
© The Author(s) 2023

## Abstract

**Background** Generic instruments such as the Pediatric Quality of Life Inventory™ v4.0 Generic Core Scales (PedsQL GCS) and Child Health Utility 9D (CHU9D) are widely used to assess health-related quality of life (HRQOL) of the general childhood population, but there is a paucity of information about their psychometric properties in children with specific health conditions. This study assessed psychometric properties, including acceptability, reliability, validity, and responsiveness, of the PedsQL GCS and the CHU9D in children and adolescents with a range of common chronic health problems.

**Methods** We used data from the Longitudinal Study of Australian Children (LSAC), for children aged 10–17 years with at least one of the following six parent-reported health conditions: asthma, anxiety/depression, attention deficit hyperactivity disorder (ADHD), autism/Asperger’s, epilepsy, and type 1 diabetes mellitus. The LSAC used parent proxy-reported PedsQL GCS and child self-reported CHU9D assessments. The performance of each instrument (PedsQL GCS and CHU9D) for each psychometric property (acceptability, reliability, validity, and responsiveness) was assessed against established criteria.

**Results** The study sample included 7201 children and adolescents (mean age = 14 years; range 10.1–17.9 years; 49% female) with 15,568 longitudinal observations available for analyses. Across the six health conditions, acceptability of the PedsQL GCS was high, while acceptability for the CHU9D was mixed. Both the PedsQL GCS and CHU9D showed strong internal consistency (Cronbach’s alpha range: PedsQL GCS = 0.70–0.95, CHU9D = 0.76–0.84; item-total correlations range: PedsQL GCS = 0.35–0.84, CHU9D = 0.32–0.70). However, convergent validity for both the PedsQL GCS and CHU9D was generally weak (Spearman’s correlations  $\leq 0.3$ ). Known group validity was strong for the PedsQL GCS (HRQOL differences were detected for children with and without asthma, anxiety/depression, ADHD, autism/Asperger’s, and epilepsy). CHU9D was only able to discriminate between children with and without anxiety/depression, ADHD, and autism/Asperger’s. The responsiveness of both the PedsQL GCS and CHU9D was variable across the six conditions, and most of the estimated effect sizes were relatively small ( $< 0.5$ ).

**Conclusion** This study expands the evidence base of psychometric performance of the PedsQL GCS and CHU9D and can aid in appropriate HRQOL instrument selection for the required context by researchers and clinicians.

## 1 Introduction

Patient-reported outcome measures (PROMs) have become increasingly important in health outcomes research in children and young people, including in clinical trials, evaluations of health systems, and informing decision making [1, 2]. Generic PROMs include measures that aim to assess a broad concept of health-related quality of life (HRQOL) that is relevant to all populations and therefore enable comparisons across clinical conditions and between clinical and general populations [3, 4]. Some generic measures of

HRQOL measure preference-based HRQOL on a utility scale (e.g. multi-attribute instruments) and are used to calculate quality-adjusted life years (QALYs) in health economic evaluations to inform decision making within the context of healthcare funding constraints [5]. Other non-preference-based PROMs are valuable as broad measures of health in clinical trials and population health monitoring. All PROMs must meet minimum standards for scientific quality. These standards are well established and are developed and used across research [6–8], professional bodies [9], and clinical trials and industry regulators [10].

A recent systematic review identified 89 generic PROMs for use in children and young people (aged  $\leq 18$  years old),

Extended author information available on the last page of the article

### Key Points for Decision Makers

Overall the Pediatric Quality of Life Inventory™ v4.0 Generic Core Scales (PedsQL GCS) and Child Health Utility 9D (CHU9D) instruments both demonstrated consistent psychometric robustness against a set of absolute criteria for the psychometric properties assessed in this study; however, the CHU9D was more variable in its performance with some of the psychometric properties such as acceptability and known group validity.

This study expands the evidence base of psychometric performance of the PedsQL GCS and CHU9D in a large representative cohort of Australian children and adolescents with common chronic health conditions.

Results from this study can aid in appropriate health-related quality-of-life instrument selection by researchers and clinicians, to ensure that clinical and policy decision-making outcomes are based on consistent and robust research and evaluations.

including preference-based HRQOL instruments [1]. The Pediatric Quality of Life Inventory™ v4.0 Generic Core Scales (PedsQL GCS), a non-preference-based HRQOL instrument, is one of the most widely used generic PROMs for HRQOL assessment [11]. Using an integrated generic core and disease/symptom-specific modular approach, the conceptual framework for the Pediatric Quality of Life Inventory™ (PedsQL) was derived from an earlier HRQOL measure that was developed using a cancer patient cohort aged 8–18 years old [12, 13]. The PedsQL model includes several condition-specific modules (e.g. asthma, rheumatology, diabetes) and the PedsQL 4.0 Generic Core Scales (PedsQL GCS). The PedsQL GCS was developed as a generic measure and validated in a cohort that included healthy children and children with chronic conditions, including asthma, attention deficit hyperactivity disorder (ADHD), and diabetes. [14, 15]. The Child Health Utility 9D (CHU9D) is a relatively recent preference-based HRQOL instrument that was developed specifically with and for children and includes value sets for different country contexts derived from adolescent populations to provide health utilities [16–19]. The Longitudinal Study of Australian Children (LSAC) administered both the PedsQL GCS and the CHU9D instruments to their participants and provided this study with the opportunity to evaluate the psychometric properties of these two instruments within the same sample of Australian children and adolescents.

Assessment of psychometric properties of HRQOL instruments is important to determine whether the instrument provides valid, reliable, and responsive measurement of the concept being assessed, i.e. HRQOL. However, many lack evidence of performance over a comprehensive range of psychometric properties [20, 21]. This lack of evidence may hinder appropriate instrument selection for the required context by researchers and clinicians, and raises potential issues such as the validity and reliability of outcome measurements and consequent clinical and policy decision making. Two systematic reviews identified evidence gaps in the psychometric performance of HRQOL instruments for children and young people, including the PedsQL and CHU9D [4, 22]. Moreover, Rowen and colleagues highlighted that good performance of preference-based instruments in a general population does not necessarily signal good performance in patient populations with specific clinical conditions [22]. The psychometric performance evidence base is more extensive for the PedsQL but relatively limited for the CHU9D [4, 22]. There is very little research that assesses the psychometric properties of the PedsQL and CHU9D in the same cohort, especially patient cohorts, to evaluate the performance of these instruments against a set of absolute criteria using the same dataset. Some studies have assessed the psychometric properties of the PedsQL GCS and validated translations of the CHU9D using the same study sample, namely in Denmark, China, and Sweden [23–25]. The paucity of evidence around the psychometric performance of the PedsQL GCS and CHU9D in Australian cohorts is also evident, with only two studies that have examined both instruments in Australian adolescent populations [26, 27]. However, both studies used cross-sectional data and assessed psychometric properties in general populations of adolescents aged 11–12 years and 15–17 years, respectively. The first study [26] only indirectly assessed psychometric properties of the PedsQL GCS and CHU9D, while the second study [27] used the Short Form 15 version of the PedsQL GCS, rather than the full instrument.

Given the gap in the evidence base for the psychometric performance of the generic PedsQL GCS and CHU9D instruments in patient populations of Australian children, the aim of this study was to assess the acceptability, reliability, validity, and responsiveness of the PedsQL GCS and the CHU9D for the measurement of HRQOL among children with common chronic conditions, using a longitudinal dataset.

## 2 Methods

### 2.1 Participants

This study was a secondary analysis of existing data from the LSAC [28]. The LSAC is a continuing population representative survey of children and their families that collects data on child wellbeing and development over the paediatric life course of the child, using separate parent-reported and child-reported questionnaires. The LSAC data collection phases (referred to as waves) started with the initial recruitment of two cohorts of children during 2004 (wave 1). Children and their parents have been interviewed every 2 years using a mixture of data collection methods and modes, i.e. self-complete, interviewer-administered, mail-out, in-person, telephone, and computer-assisted methods [28]. The most recent wave of data collection was in 2020 (wave 9). Wave 1 included two cohorts of children with 5107 in the ‘baby’ (B) cohort aged 0–1 years old and 4983 children in the ‘kindergarten’ (K) cohort aged 4–5 years old at baseline [29]. In this study, we used a subset of the LSAC that included data collected using both the PedsQL GCS (parent proxy-reported questionnaire) and CHU9D (child self-reported questionnaire) instruments from the B and K cohorts. The resulting longitudinal dataset included data from 2013 to 2018, and included the B cohort in which children were aged 10–11 (wave 6; 2013–2014), 12–13 (wave 7; 2015–2016), and 14–15 years old (wave 8; 2017–2018), respectively, and the K cohort in which children were aged 14–15 (wave 6; 2013–2014) and 16–17 years old (wave 7; 2015–2016).

### 2.2 HRQOL Instruments

#### 2.2.1 PedsQL GCS

The PedsQL GCS is a non-preference-based generic PROM that was developed to measure childhood HRQOL [30]. Parent proxy-reported age-appropriate PedsQL GCS versions were used in the LSAC. For children aged 10–13 years old, the LSAC used the PedsQL GCS parent report for children (ages 8–12) version for all the relevant waves in the B and K cohorts, except for children aged 10–11 years (wave 4) in the K cohort, which used the PedsQL GCS parent report for young children (ages 5–7) version. For children aged 14–17 years, the PedsQL GCS parent report for teenagers (ages 13–18) version was used for all the relevant waves in both the B and K cohorts. The age-specific PedsQL GCS versions used in this study have the same scoring structure; however, some questions are worded differently to be age appropriate. The PedsQL GCS versions used in the study consist of 23 items within four summary functional subscales/domains, namely, physical (8 items), emotional (5 items), social (5

items), and school (5 items) [30]. Each item is scored on a 5-point scale (0 = never a problem; 1 = almost never a problem; 2 = sometimes a problem; 3 = often a problem; 4 = almost always a problem). To score the PedsQL GCS, the items scores are reversed scored (0 = 100, 1 = 75, 2 = 50, 3 = 25, 4 = 0) and transformed onto a linear scale where the total score scale is the sum of items scores divided by the number of items answered, and ranges from 0 to 100, with higher scores indicating better HRQOL [30].

#### 2.2.2 CHU9D

The CHU9D is a generic childhood preference-based HRQOL instrument designed to measure health utility. It was developed with children from the United Kingdom and initially validated for a target age of 7–11 years old in a general population sample and a clinical sample [17, 18]. The clinical sample included a wide range and severity of health conditions, with children recruited from a medical ward that covered acute and chronic medical conditions, surgical wards that covered a range of renal, gastrointestinal, neurological, orthopaedic, limb reconstruction, and spinal surgeries, and a day care unit that covered children undergoing procedures involving urology, gastroenterology, endocrinology, neurology, oncology, orthopaedics, general surgery, dental, ear, nose and throat, as well as allergy patients [17, 18]. It has also been validated in Australia in a sample of adolescents aged 11–17 years old from the general population [19, 31]. The CHU9D is self-reported by the child in the LSAC dataset and comprises nine dimensions (worried, sad, pain, tired, annoyed, schoolwork, sleep, daily routine, and activities), each of which are scored across five levels (0 = "It is never a problem"; 1 = "It is almost never a problem"; 2 = "It is sometimes a problem"; 3 = "It is often a problem"; 4 = "It is almost always a problem") [18]. A preference-based value set is applied to the participant responses to provide an overall health utility score indexed from 0 (usually indicating dead) to 1 (perfect health) and can include negative values that indicate health states considered worse than being dead. Value sets for the CHU9D have been developed for the United Kingdom, Australia, China, and the Netherlands [1]. In the LSAC dataset, the overall CHU9D utility scores were calculated using the Australian valuation algorithm developed in adolescents [32], which results in a utility scale that ranges from -0.1059 (poorest health) to 1 (perfect health).

### 2.3 Health Conditions

At each wave of data collection in the LSAC, the parent proxy reported the presence or absence of a range of health conditions for their child. While the LSAC includes children with various health conditions, this study focussed only on the children with parent proxy-reported presence of any of

the six common chronic health conditions (asthma, anxiety/depression, ADHD, autism/Asperger's, epilepsy, type 1 diabetes) and children without any of these six conditions. Children may have had multiple health conditions in LSAC; however, this study identified the presence of any of these six specified health conditions as an individual observation when analysing by each condition and did not account for children with multiple conditions. An end user group (EUG) for this study, which consisted of paediatric and non-paediatric clinicians, health economists, and decision makers (14 members in total), stated and ranked paediatric clinical health conditions in order of priority. The choice of these final six common chronic health conditions was based on a combination of the prioritisation feedback from the study EUG and the availability of data for the health conditions within the LSAC.

## 2.4 Sociodemographic Characteristics

Where appropriate, the statistical analyses controlled for sociodemographic characteristics of the participating children that are known predictors of childhood HRQOL, such as age [33, 34], gender [34], culturally and linguistically diverse (CALD) status [35], being Aboriginal or Torres Strait Islander [36], and socioeconomic position (SEP) [37, 38]. The sociodemographic characteristics were parent proxy-reported and coded in the analyses as follows: age (in years), sex (male/female), SEP (high/low), CALD status (CALD/not CALD), and Aboriginal or Torres Strait Islander (yes/no). The LSAC study measures SEP using a composite variable developed for the study, which combines education level, occupation type, and income of the child's parents into a  $z$ -score [39]. The present analysis further categorised the SEP variable into low SEP [yes = SEP  $z$ -score < 0 (low SEP); no = SEP  $z$ -score  $\geq$  0 (high SEP)]. The variable that documented whether a language other than English was regularly spoken to the child at home was collected at age 2–3 years for the B cohort and age 4–5 years for the K cohort and was used as a proxy measure for CALD status in this study.

## 2.5 Evaluation of Psychometric Properties

For each instrument, we evaluated acceptability, reliability, validity, and responsiveness and compared them against criteria from psychometric best-practice guidelines and established standards [6–10, 40–42]. Except when assessing acceptability through the level of missing data, the psychometric evaluation for each health condition used complete observations for the specified condition, PedsQL GCS, and CHU9D. Acceptability, reliability, and convergent validity analyses were conducted separately for each of the six health conditions and also for children without any of the six conditions. Known group validity and responsiveness

were assessed separately for children with each of the six health conditions.

*Acceptability* typically measures the quality of the data and is assessed by data completeness and distribution of scores, including floor and ceiling effects [10, 41]. Acceptability can also refer to the feasibility and practicality of using a particular instrument and may include measures of comprehension or burden of completion [10, 40]; however, these data were not available within the LSAC. Hence, acceptability was limited to assessment of missing data (missingness) and the proportion of ceiling and floor values for the PedsQL GCS total scores and CHU9D utility scores. Missing data levels of < 5% and proportions of floor and ceiling values < 10% are considered relevant thresholds for acceptability [41].

*Reliability* relates to the degree to which an instrument measurement is free from random error [7, 9]. The cohort dataset only enabled assessment of reliability in terms of *internal consistency*, which measures the inter-relatedness among items from the same scale [7, 42]. We assessed internal consistency for the four subscales of the PedsQL GCS, the total PedsQL GCS, and the total CHU9D scales using Cronbach's alpha and item-total correlations. Cronbach's alpha values  $\geq$  0.7 and item-total correlations  $\geq$  0.2 are considered minimum standards for internal reliability consistency [41–44].

*Validity* assesses whether the instrument measures the concept that it is intended to measure [7, 42]. This study assessed convergent validity and known group validity, both of which provide evidence of construct validity [7]. *Convergent validity* assesses the extent to which there is an association between the scale under investigation and other scales that measure the same or similar constructs [10, 40]. The PedsQL GCS and CHU9D are both established and accepted measures of the HRQOL construct, and as there was no other independent validated HRQOL instrument within the LSAC dataset, convergent validity was assessed using these instruments. This is not ideal as both instruments were under scrutiny in our study. However, given that this was a secondary analysis, and we were limited to the data collection undertaken by the LSAC, this test was better than having no test at all for convergent validity. Convergent validity was assessed using Spearman's correlations between the PedsQL GCS total score and the CHU9D utility score for the children and adolescents in the study sample. Convergent validity was determined by the expected correlation and was categorised based on strength of the correlation coefficients, i.e. weak (< 0.40), moderate (0.41–0.60), good (0.61–0.80), and strong (> 0.80) [22]. Although both the PedsQL GCS and CHU9D measure HRQOL, the LSAC used only parent proxy-reported versions of the PedsQL GCS, while the CHU9D was child self-reported; hence, we hypothesised

that there may be only moderate correlations between the two instruments.

*Known group validity* assesses if the instrument can differentiate between clinically distinct groups [10]. We hypothesised that children with each of the six selected clinical conditions individually would have lower HRQOL compared to those without that condition. Known group validity was evaluated using general estimating equations (GEE) to account for the repeated measures of HRQOL and reported clinical condition among the same children, with adjustment for sociodemographic characteristics known to impact on HRQOL [45]. The PedsQL GCS total score, which is scored on a 0–100 scale, was transformed to a 0–1 scale so that the PedsQL GCS total score and CHU9D utility score variables within the GEE models used the same scale. The GEE models specified a binomial family with a log-link function to account for the distribution of the HRQOL scores being on a 0–1 scale in this study, and robust variance estimation. Separate GEE models were estimated for each of the six health conditions (asthma, anxiety /depression, ADHD, autism/Asperger’s, epilepsy, type 1 diabetes) for both the PedsQL GCS and CHU9D, where the effect of the presence of each selected condition was relative to the absence of that selected condition only (e.g. asthma vs no asthma only, ADHD vs no ADHD only); hence, the reference level sample differs for each of the individual conditions because health conditions other than the condition of interest may have been present. This resulted in six separate GEE models using the PedsQL GCS total scale score (transformed to a 0–1 scale) as the response variable and the presence of each of the six selected conditions individually as the univariate explanatory variable, adjusted for the sociodemographic characteristics (i.e. age, sex, whether participants identified as Aboriginal or Torres Strait Islander, CALD status, and SEP), and a further six GEE models were specified in the same way but with the CHU9D utility score as the response variable. Significance levels were set at  $p < 0.05$  for main effects and  $p < 0.01$  for interaction terms. Interaction terms of clinical condition and significant sociodemographic variables were explored; however, none were significant and only the main effects GEE models are reported. The main effects GEE models were used to predict the marginal effects of the clinical condition on HRQOL, i.e. the adjusted predicted mean PedsQL GCS total score transformed back to a 0–100 scale and mean CHU9D utility scores in the presence and absence of the individual clinical condition.

*Responsiveness* assesses the ability of the instrument to detect change over time in the construct being measured [7] in relation to an intervention of known efficacy [22, 41]. We hypothesised a change in HRQOL (i.e. a change in the

PedsQL GCS total score or CHU9D utility score) if there was a change in the status of the selected health conditions between consecutive waves of the LSAC data. Children were classified as to whether each of the selected clinical conditions persisted, resolved, or manifested between consecutive waves of the LSAC dataset. As autism/Asperger’s, epilepsy, and type 1 diabetes are lifelong conditions [46–48], they were not categorised as resolved as this was considered not to be clinically possible. The B and K cohorts provided data on the change in HRQOL total scores for individual children over 2-year intervals for the following three age group progressions: 10–11 to 12–13 years old, 12–13 to 14–15 years old, and 14–15 to 16–17 years old. We hypothesised that a child moving from absence to presence of the condition (e.g. no asthma to asthma) would result in a negative HRQOL score change, whilst moving from presence to absence of the selected conditions (e.g. asthma to no asthma) would result in a positive HRQOL score change, and no change in presence or absence of the condition would result in an HRQOL score change close to zero. Responsiveness was evaluated using effect sizes (ES), which account for the change in HRQOL score in relation to the standard deviation (SD) of the baseline score, and were calculated according to the method outlined by Pink and colleagues [49]. We used the commonly accepted, although arbitrary, categorisations of magnitude of ES being small ( $ES = 0.2$ ), moderate ( $ES = 0.5$ ), and large ( $ES \geq 0.8$ ) [44, 50].

## 3 Results

### 3.1 Participants

Descriptive statistics for the study analysis sample are shown in Table 1. Across both cohorts and all age groups, data consisted of 15,568 observations from 7201 children with 9529 observations from 3760 children in the B cohort (‘baby’ cohort) and 6039 observations from 3441 children in the K cohort (‘kindergarten’ cohort).

There were low numbers of children reported with epilepsy and diabetes (type 1) over all waves of data collection for the B and K cohorts. The distribution of sociodemographic characteristics varied across the six conditions, with a higher proportion of boys having ADHD or autism/Asperger’s compared with the other health conditions and the sample without any of these conditions.

### 3.2 Psychometric Properties

#### 3.2.1 Acceptability

Acceptability of the PedsQL GCS was high, based on the overall low level of missingness ( $< 2\%$ ) for the total scores and no floor or ceiling effects across all samples of children

with each of the six conditions and without any of the conditions, over all the age groups (see Supplementary Table S1 in the Electronic Supplementary Material [ESM]). Acceptability for the CHU9D was mixed. While missingness was < 5% for CHU9D utility scores for most age groups for asthma, type 1 diabetes, and children without any of the six conditions, missingness was > 5% for most age groups in children with epilepsy, ADHD, autism/Asperger's, and anxiety/depression. There were no floor effects for the CHU9D utility total score. However, ceiling effects > 10% were observed for most of the age groups for children with each of the six conditions and without any of those conditions.

### 3.2.2 Internal Consistency Reliability

Among children and adolescents with each of the six conditions individually and without any of the six conditions, internal consistency reliability for the PedsQL GCS total score scale and the individual summary score subscales for physical health, emotional, social, and school functioning was strong (Cronbach's alpha range 0.70–0.95; item-total correlations range 0.35–0.84) (Table 2). CHU9D utility scores also showed strong internal consistency reliability for children and adolescents without any of the six conditions individually and over all the conditions (Cronbach's alpha range 0.76–0.84; item-total correlations range 0.32–0.70), except for type 1 diabetes (Cronbach's alpha = 0.65; item-total correlations range 0.09–0.56).

### 3.2.3 Convergent Validity

Convergent validity between the PedsQL GCS and CHU9D among children with each of the six conditions individually and without any of the six conditions was weak overall (i.e. < 0.4 threshold), and the statistically significant Spearman's correlation coefficients ranged from 0.13 to 0.30 (Table 3). Children aged 14–15 years old (K cohort) with type 1 diabetes were the exception (Spearman's correlation coefficient = 0.62); however, this was a small sample ( $n = 9$ ).

### 3.2.4 Known Group Validity

The outputs from the main effects models using GEE for quality of life associated with the six clinical conditions are presented in Supplementary Table S4 in the ESM. These univariate analyses of the health conditions, adjusted for the selected sociodemographic variables, indicated that the presence of asthma, anxiety/depression, ADHD, and autism/Asperger's coefficients were statistically significant, but epilepsy and type 1 diabetes were not.

Marginal predictions for the mean PedsQL GCS total score from the GEE main effects models adjusted for sex, age, identifying as Aboriginal or Torres Strait Islander,

CALD status, and SEP (Fig. 1 and Supplementary Table S2 in the ESM), indicated that the PedsQL GCS discriminated ( $p < 0.05$ ) between children with and without anxiety/depression, ADHD, autism/Asperger's, and epilepsy, but not for asthma and type 1 diabetes.

Marginal predictions for the mean CHU9D utility score from the GEE main effects models adjusted for sex, age, identifying as Aboriginal or Torres Strait Islander, CALD status, and SEP (Fig. 1 and Supplementary Table S3 in the ESM) indicated mixed known group validity for the CHU9D. The CHU9D only discriminated between children with and without anxiety/depression, ADHD, and autism/Asperger's, but not for those with asthma, epilepsy, and type 1 diabetes. Note that the marginal predictions for the mean PedsQL GCS total score and mean CHU9D utility score varied for the reference level (i.e. without the condition) for each of the individual health conditions (Fig. 1 and Supplementary Table S3). This is because the effect of the presence of the selected condition in each GEE model was relative to the absence of that selected condition only; hence, the reference level sample differed for each of the individual conditions. Figure 1 does include the 95% confidence intervals (CIs) for the marginal prediction estimates; however, due to the large sample sizes, the 95% CIs are very precise and therefore not apparent in the figure.

### 3.2.5 Responsiveness

In Figs. 2 and 3 respectively, the ES estimates would be expected to trend from left to right for a responsive instrument, with a negative ES expected for the "worse" category, an ES close to zero for "same", and a positive ES for "better". The responsiveness of both PedsQL GCS and CHU9D was variable for the six conditions across the available age groups in the B and K cohorts (Figures 2 and 3).

The responsiveness of the PedsQL GCS and CHU9D is consistent with the expected trend in ES for children with anxiety/depression in all the three age group progressions, asthma in the 14–15- to 16–17-years-old progression, ADHD in the 10–11- to 12–13-years-old and 12–13- to 14–15-years-old progressions, autism/Asperger's in the 10–11- to 12–13-years-old and 14–15- to 16–17-years-old progressions, and type 1 diabetes in the 10–11- to 12–13-years-old progression. The responsiveness of PedsQL GCS is also consistent with the expected trend in ES for children with epilepsy for the 12–13- to 14–15-years-old progression, while it is inconsistent for CHU9D for these children. The responsiveness of CHU9D is consistent with the expected trend in ES observed in children with type 1 diabetes for the 12–13- to 14–15-years-old progression, and is inconsistent for PedsQL GCS for these children. The responsiveness of PedsQL GCS and CHU9D are both inconsistent with the expected trend in ES observed in children

**Table 1** Descriptive statistics for study analysis sample of children and adolescents by health condition and age group

Cohort	Variables	Sample without any of the 6 conditions	Asthma	Anxiety/depression	ADHD	Autism/Asperger's	Epilepsy	Diabetes (type 1)
<b>B cohort</b>								
10–11 years ( <i>N</i> = 3481)	<i>n</i> (%)	2736 (78.6)	452 (13.0)	161 (4.6)	119 (3.4)	117 (3.4)	19 (0.6)	11 (0.3)
	PedsQL GCS mean total score (SD)	82.0 (12.0)	79.5 (14.1)	64.5 (16.0)	68.9 (15.0)	62.0 (15.9)	67.3 (19.3)	81.2 (12.0)
	CHU9D mean utility score (SD)	0.81 (0.17)	0.78 (0.18)	0.71 (0.20)	0.75 (0.20)	0.78 (0.20)	0.80 (0.20)	0.75 (0.20)
	Female, <i>n</i> (%)	1396 (51)	199 (44)	80 (50)	33 (28)	23 (20)	9 (47.4)	6 (54.6)
	Low socioeconomic position, <i>n</i> (%)	1361 (50)	251 (56)	97 (60)	68 (57)	64 (55)	12 (63.2)	6 (54.6)
	Aboriginal or Torres Strait Islander, <i>n</i> (%)	66 (2.4)	17 (3.8)	3 (1.9)	6 (5.0)	3 (2.6)	1 (5.3)	0 (0)
	Culturally and linguistically diverse, <i>n</i> (%)	463 (17)	87 (19)	25 (16)	17 (14)	19 (16)	4 (21.1)	1 (9.1)
12–13 years ( <i>N</i> = 3124)	<i>n</i> (%)	2419 (77.4)	403 (12.9)	217 (7.0)	114 (3.7)	123 (3.9)	16 (0.5)	14 (0.5)
	PedsQL GCS mean total score (SD)	82.8 (12.2)	78.4 (14.8)	65.2 (15.6)	67.1 (16.0)	63.6 (15.7)	61.4 (25.1)	77.9 (13.4)
	CHU9D mean utility score (SD)	0.82 (0.17)	0.79 (0.20)	0.69 (0.24)	0.77 (0.24)	0.74 (0.24)	0.79 (0.16)	0.81 (0.12)
	Female, <i>n</i> (%)	1203 (50)	190 (47)	113 (52)	36 (32)	34 (28)	5 (31.3)	9 (64.3)
	Low socioeconomic position, <i>n</i> (%)	1194 (49)	222 (55)	114 (53)	66 (58)	69 (56)	9 (56.3)	8 (57.1)
	Aboriginal or Torres Strait Islander, <i>n</i> (%)	49 (2.0)	19 (4.7)	7 (3.2)	8 (7.0)	4 (3.3)	1 (6.3)	0 (0)
	Culturally and linguistically diverse, <i>n</i> (%)	427 (18)	71 (18)	32 (15)	15 (13)	18 (15)	1 (6.3)	2 (14.3)
14–15 years ( <i>N</i> = 2924)	<i>n</i> (%)	2288 (78.2)	372 (12.7)	225 (7.7)	114 (3.9)	126 (4.3)	16 (0.6)	16 (0.6)
	PedsQL GCS mean total score (SD)	80.6 (14.8)	77.6 (16.9)	63.8 (16.7)	68.9 (16.6)	65.0 (16.5)	64.1 (19.4)	74.0 (22.4)
	CHU9D mean utility score (SD)	0.81 (0.19)	0.79 (0.20)	0.69 (0.27)	0.79 (0.24)	0.78 (0.23)	0.73 (0.29)	0.83 (0.18)
	Female, <i>n</i> (%)	1120 (49)	187 (50)	105 (47)	38 (33)	43 (34)	7 (43.8)	9 (56.3)
	Low socioeconomic position, <i>n</i> (%)	1111 (49)	204 (55)	126 (56)	64 (56)	63 (50)	10 (62.5)	7 (43.8)
	Aboriginal or Torres Strait Islander, <i>n</i> (%)	44 (1.9)	12 (3.2)	6 (2.7)	8 (7.0)	3 (2.4)	1 (6.25)	0 (0)
	Culturally and linguistically diverse, <i>n</i> (%)	401 (18)	70 (19)	37 (16)	16 (14)	22 (17)	0 (0)	2 (12.5)
<b>K cohort</b>								

**Table 1** (continued)

Cohort	Variables	Sample without any of the 6 conditions	Asthma	Anxiety/depression	ADHD	Autism/Asperger's	Epilepsy	Diabetes (type 1)
14–15 years ( <i>N</i> = 3233)	<i>n</i> (%)	2464 (76.2)	409 (12.7)	305 (9.4)	95 (2.9)	77 (2.4)	14 (0.4)	14 (0.4)
	PedsQL GCS mean total score (SD)	81.9 (14.0)	76.3 (16.9)	64.4 (16.9)	66.3 (16.8)	63.7 (17.8)	61.7 (17.7)	73.0 (23.4)
	CHU9D mean utility score (SD)	0.81 (0.19)	0.80 (0.19)	0.64 (0.26)	0.77 (0.22)	0.79 (0.21)	0.82 (0.18)	0.78 (0.20)
	Female, <i>n</i> (%)	1218 (49)	179 (44)	193 (63)	22 (23.2)	17 (22.1)	7 (50.0)	6 (42.9)
	Low socioeconomic position, <i>n</i> (%)	1231 (50)	208 (51)	182 (60)	49 (51.6)	44 (57.1)	10 (71.4)	8 (57.1)
	Aboriginal or Torres Strait Islander, <i>n</i> (%)	53 (2.2)	11 (2.7)	5 (1.6)	1 (1.1)	2 (2.6)	0 (0)	0 (0)
	Culturally and linguistically diverse, <i>n</i> (%)	438 (18)	70 (17)	42 (14)	10 (10.5)	11 (14.3)	5 (35.7)	2 (14.3)
16–17 years ( <i>N</i> = 2806)	<i>n</i> (%)	2168 (77.3)	321 (11.4)	319 (11.4)	77 (2.7)	81 (2.9)	8 (0.3)	9 (0.3)
	PedsQL GCS mean total score (SD)	82.8 (12.9)	77.7 (15.0)	67.0 (15.2)	68.8 (16.3)	62.9 (17.6)	69.6 (23.0)	79.5 (12.0)
	CHU9D mean utility score (SD)	0.80 (0.20)	0.75 (0.24)	0.66 (0.25)	0.81 (0.20)	0.77 (0.22)	0.85 (0.19)	0.85 (0.19)
	Female, <i>n</i> (%)	1072 (49)	152 (47)	167 (52)	15 (19.5)	15 (18.5)	3 (37.5)	2 (22.2)
	Low socioeconomic position, <i>n</i> (%)	1095 (51)	157 (49)	162 (51)	37 (48.1)	43 (53.1)	4 (50.0)	4 (44.4)
	Aboriginal or Torres Strait Islander, <i>n</i> (%)	44 (2.0)	9 (2.8)	2 (0.6)	0 (0)	1 (1.2)	0 (0)	0 (0)
	Culturally and linguistically diverse, <i>n</i> (%)	398 (18)	52 (16)	38 (12)	6 (7.8)	10 (12.4)	4 (50.0)	1 (11.1)

ADHD attention deficit hyperactivity disorder, *B* baby, CHU9D Child Health Utility 9D, *K* kindergarten, PedsQL GCS Pediatric Quality of Life Inventory™ v4.0 Generic Core Scales, SD standard deviation

PedsQL GCS total score scale = 0–100

CHU9D utility score scale = 0–1

with asthma for the 10–11- to 12–13-years-old and 12–13- to 14–15-years-old progressions, ADHD in the 14–15- to 16–17-years-old progression, autism/Asperger's in the 12–13- to 14–15-years-old progression, and epilepsy in the 14–15- to 16–17-years-old progression.

For the responsiveness results that were consistent with the hypothesised direction of change for ES, most of the estimated ES were relatively small for both the PedsQL GCS and CHU9D. The responsiveness results in children with epilepsy and type 1 diabetes are likely underpowered due to the low numbers, and there were no observations in the

“worse” category. The point estimates and 95% CIs for the PedsQL GCS and CHU9D ES by condition and age group progressions are presented in Supplementary Tables S5 and S6, respectively, in the ESM.

## 4 Discussion

This study investigated the psychometric properties of two generic childhood HRQOL instruments, the PedsQL GCS and CHU9D. This was conducted in the context of a large



**Table 2** Internal consistency results for PedsQL GCS and CHU9D among children aged 10–17 years old by clinical condition

Condition	Variable	Cronbach's alpha value <sup>a</sup>	Item-total correlations range <sup>b</sup>
1. Sample without any of the 6 conditions ( <i>n</i> = 12,075)	<b>CHU9D</b> Total utility score scale	0.7855	0.3622–0.5861
	<b>PedsQL GCS</b> Total score scale	0.9048	0.4122–0.6002
	PedsQL GCS Physical health summary score subscale	0.8586	0.3691–0.7784
	PedsQL GCS Emotional functioning summary score subscale	0.8038	0.5158–0.6646
	PedsQL GCS Social functioning summary score subscale	0.7934	0.5114–0.6302
	PedsQL GCS School functioning summary score subscale	0.7506	0.3956–0.6354
2. Asthma ( <i>n</i> = 1957)	<b>CHU9D</b> Total utility score scale	0.8069	0.3833–0.5975
	<b>PedsQL GCS</b> Total score scale	0.9221	0.4588–0.6585
	PedsQL GCS Physical health summary score subscale	0.8658	0.4158–0.7972
	PedsQL GCS Emotional functioning summary score subscale	0.8477	0.5874–0.7186
	PedsQL GCS Social functioning summary score subscale	0.8331	0.5950–0.6879
	PedsQL GCS School functioning summary score subscale	0.7749	0.4616–0.6486
3. Anxiety/depression ( <i>n</i> = 1227)	<b>CHU9D</b> Total utility score scale	0.8390	0.4405–0.6996
	<b>PedsQL GCS</b> Total score scale	0.9037	0.4003–0.6343
	PedsQL GCS Physical health summary score subscale	0.8452	0.4395–0.7303
	PedsQL GCS Emotional functioning summary score subscale	0.7820	0.4642–0.6488
	PedsQL GCS Social functioning summary score subscale	0.8427	0.6174–0.7082
	PedsQL GCS School functioning summary score subscale	0.7718	0.4390–0.6396
4. ADHD ( <i>n</i> = 519)	<b>CHU9D</b> Total utility score scale	0.8312	0.3660–0.6026
	<b>PedsQL GCS</b> Total score scale	0.8977	0.3654–0.5849
	PedsQL GCS Physical health summary score subscale	0.8383	0.3807–0.7402
	PedsQL GCS Emotional functioning summary score subscale	0.8071	0.5005–0.6945
	PedsQL GCS Social functioning summary score subscale	0.8197	0.5040–0.7218
	PedsQL GCS School functioning summary score subscale	0.7034	0.3765–0.5328
5. Autism/Asperger's ( <i>n</i> = 524)	<b>CHU9D</b> Total utility score scale	0.8018	0.3859–0.5752
	<b>PedsQL GCS</b> Total score scale	0.8963	0.3540–0.6032
	PedsQL GCS Physical health summary score subscale	0.8458	0.4049–0.7359
	PedsQL GCS Emotional functioning summary score subscale	0.8200	0.4977–0.6989
	PedsQL GCS Social functioning summary score subscale	0.8327	0.5931–0.7145
	PedsQL GCS School functioning summary score subscale	0.7114	0.3735–0.5971
6. Epilepsy ( <i>n</i> = 73)	<b>CHU9D</b> Total utility score scale	0.7623	0.3170–0.5742
	<b>PedsQL GCS</b> Total score scale	0.9468	0.4727–0.8415
	PedsQL GCS Physical health summary score subscale	0.8826	0.5177–0.8170
	PedsQL GCS Emotional functioning summary score subscale	0.8726	0.6307–0.7770
	PedsQL GCS Social functioning summary score subscale	0.8879	0.6887–0.7779
	PedsQL GCS School functioning summary score subscale	0.8351	0.4956–0.7519
7. Diabetes – type 1 ( <i>n</i> = 64)	<b>CHU9D</b> Total utility score scale	0.6504	0.0946–0.5599
	<b>PedsQL GCS</b> Total score scale	0.9343	0.3600–0.7065
	PedsQL GCS Physical health summary score subscale	0.9121	0.4866–0.8228
	PedsQL GCS Emotional functioning summary score subscale	0.8962	0.7135–0.7942
	PedsQL GCS Social functioning summary score subscale	0.7885	0.4446–0.6469
	PedsQL GCS School functioning summary score subscale	0.7661	0.4165–0.6744

ADHD attention deficit hyperactivity disorder, CHU9D Child Health Utility 9D, PedsQL GCS Pediatric Quality of Life Inventory™ v4.0 Generic Core Scales

<sup>a</sup>Acceptable Cronbach's alpha threshold  $\geq 0.7$

<sup>b</sup>Acceptable item-total correlations threshold  $\geq 0.2$

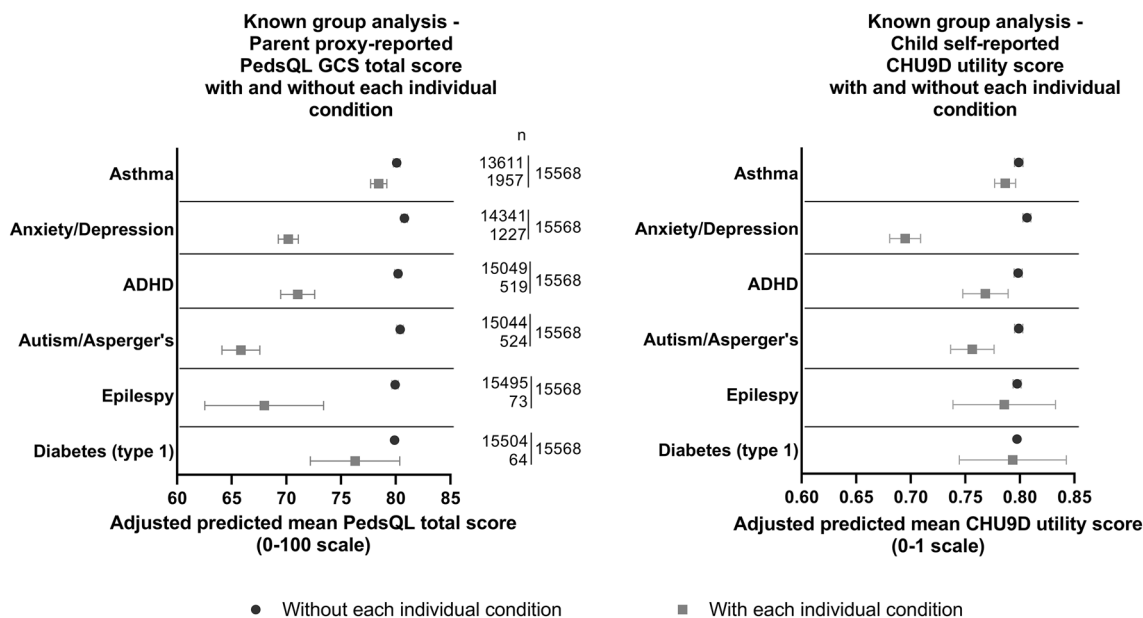
**Table 3** Convergent validity (i.e. Spearman's correlation coefficients) between PedsQL GCS total score (parent-reported) and CHU9D utility score (child-reported) for children aged 10–17 years old by clinical condition status and age group

Conditions	B cohort 10–11 years old		B cohort 12–13 years old		B cohort 14–15 years old		K cohort 14–15 years old		K cohort 16–17 years old		Total (B and K cohorts) 10–17 years old	
	<i>N</i>	Spearman's correlation coefficient	<i>N</i>	Spearman's correlation coefficient	<i>N</i>	Spearman's correlation coefficient	<i>N</i>	Spearman's correlation coefficient	<i>N</i>	Spearman's correlation coefficient	<i>N</i>	Spearman's correlation coefficient
Overall sample without any of the 6 conditions	2736	0.17	2419	0.19	2288	0.19	2464	0.24	2168	0.26	12,075	
Asthma	452	0.24	403	0.20	372	0.21	409	0.24	321	0.30	1957	
Anxiety/depression	161	0.22	217	0.23	225	0.24	305	0.30	319	0.13	1227	
ADHD	119	0.17 (NS)	114	0.21	114	0.15 (NS)	95	0.13 (NS)	77	-0.18 (NS)	519	
Autism/Asperger's	117	0.25	123	0.17 (NS)	126	0.21	77	0.09 (NS)	81	0.08 (NS)	524	
Epilepsy	19	-0.34 (NS)	16	-0.06 (NS)	16	0.05 (NS)	14	0.27 (NS)	8	0.39 (NS)	73	
Diabetes (type 1)	11	<b>0.59</b> (NS)	14	<b>0.42</b> (NS)	16	<b>0.46</b> (NS)	14	<b>0.62</b>	9	-0.33 (NS)	64	

ADHD attention deficit hyperactivity disorder, B baby, CHU9D Child Health Utility 9D, K kindergarten, NS non-significant, PedsQL GCS Pediatric Quality of Life Inventory™ v4.0 Generic Core Scales

Values are Spearman's correlation coefficients with those  $\geq 0.4$  threshold in bold

NS correlation *p* value  $> 0.05$



**Figure 1** Known group marginal predictions for PedsQL GCS total score and CHU9D utility score from the 12 separate adjusted GEE models with 95% CIs for children aged 10–17 years old with and without the individual clinical conditions. Due to the large sample sizes, the 95% CIs are very precise and therefore not apparent in the

figure. *ADHD* attention deficit hyperactivity disorder, *CHU9D* Child Health Utility 9D, *CI* confidence interval, *GEE* general estimating equation, *PedsQL GCS* Pediatric Quality of Life Inventory™ v4.0 Generic Core Scales

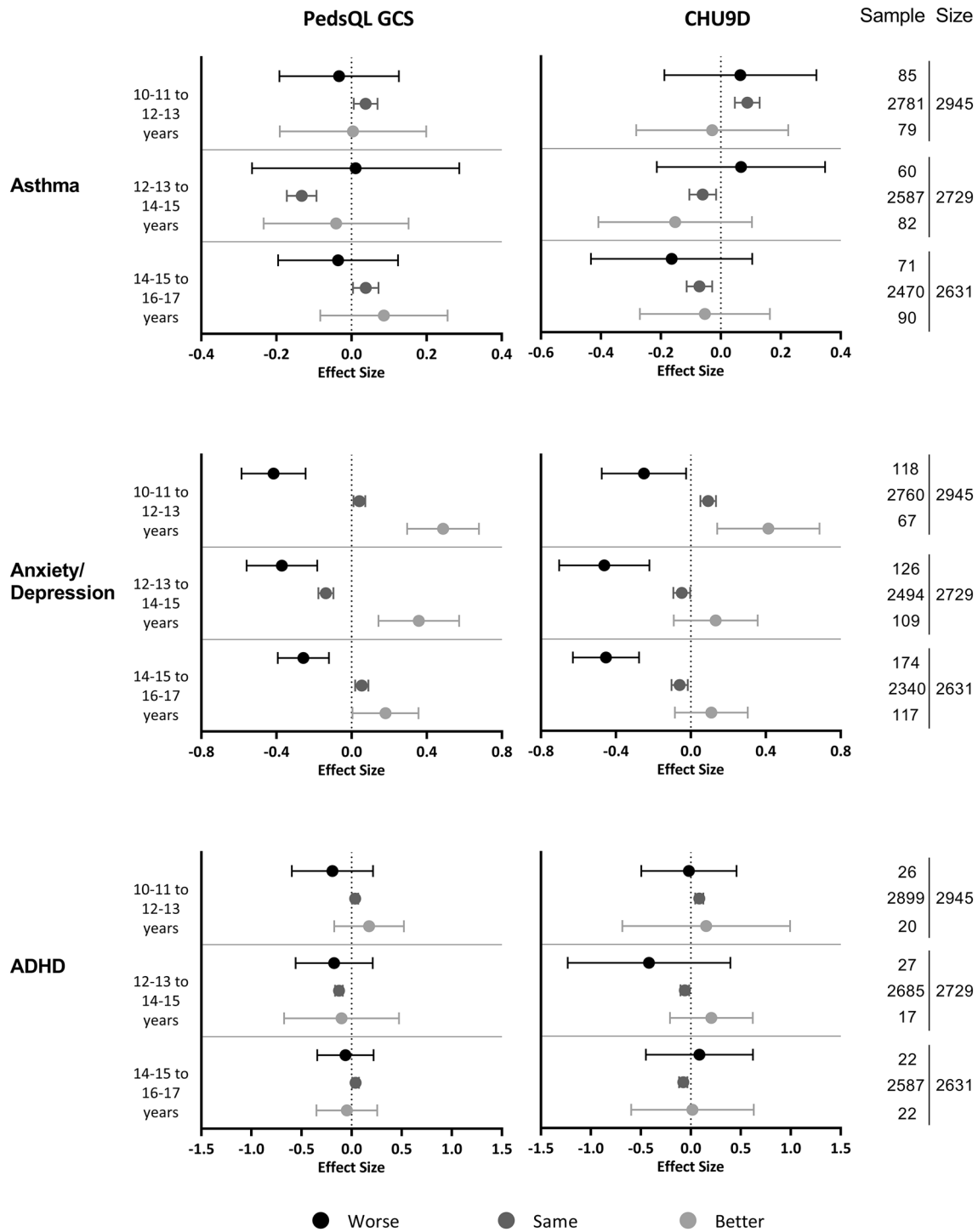
longitudinal cohort of Australian children and adolescents (aged 10–17 years old) with and without parent proxy-reported clinical conditions. The psychometric properties evaluated included acceptability, reliability, validity, and responsiveness, and the clinical conditions were asthma, anxiety/depression, ADHD, autism/Asperger’s, epilepsy, and type 1 diabetes. To our knowledge, this is the first time that a rigorous psychometric assessment of PedsQL GCS and CHU9D has been possible for a longitudinal cohort that includes reporting of a range of common chronic health conditions.

The findings of this study indicate that PedsQL GCS demonstrated good acceptability, while missingness and ceiling effects for the CHU9D exceeded acceptable levels. Overall, the PedsQL GCS and CHU9D both showed good internal consistency reliability in this cohort of children and provide users with confidence that the items within the summary subscales of PedsQL GCS and total score scales for the PedsQL GCS and CHU9D are internally consistent with the overall construct of HRQOL as intended. The low correlation between the PedsQL GCS total score and the CHU9D utility score in this study suggested poor convergent validity, even though both instruments are measuring the same construct, i.e. HRQOL. The evidence for known group validity suggested that PedsQL GCS demonstrated greater sensitivity than CHU9D at discriminating between children with and without the six common chronic conditions investigated. However, both instruments were able to detect significant

differences for children with and without anxiety/depression, ADHD, and autism/Asperger’s, which provides confidence that both instruments are able to discriminate between groups of children based on the status of these conditions. This study also assessed responsiveness, which is not often assessed in children [22]. The responsiveness performance of both PedsQL GCS and CHU9D was variable for the six conditions across the selected age groups in the B and K cohorts, and the majority of the estimated ES were relatively small for both the PedsQL GCS and CHU9D.

A potential reason for the data missingness for the CHU9D could be that it is child self-reported in the LSAC and children with these clinical conditions may find it challenging when self-completing the CHU9D. The observed ceiling effects for CHU9D in the children and adolescents in this study suggest that the utility scores are skewed and clustered towards the top of the scale and may not measure participants on the higher end of the scale very well. A multi-comparison cross-sectional study of Australian children aged 11–12 years old also reported a skewed distribution of HRQOL measured by CHU9D, i.e. more children reported high HRQOL scores [26]. However, another Australian study found no evidence of ceiling effects for the CHU9D in an Australian community-based sample of adolescents aged 11–17 years old [51].

The poor convergent validity between the PedsQL GCS and CHU9D likely reflects the method variance as a consequence of the PedsQL GCS being parent proxy-reported

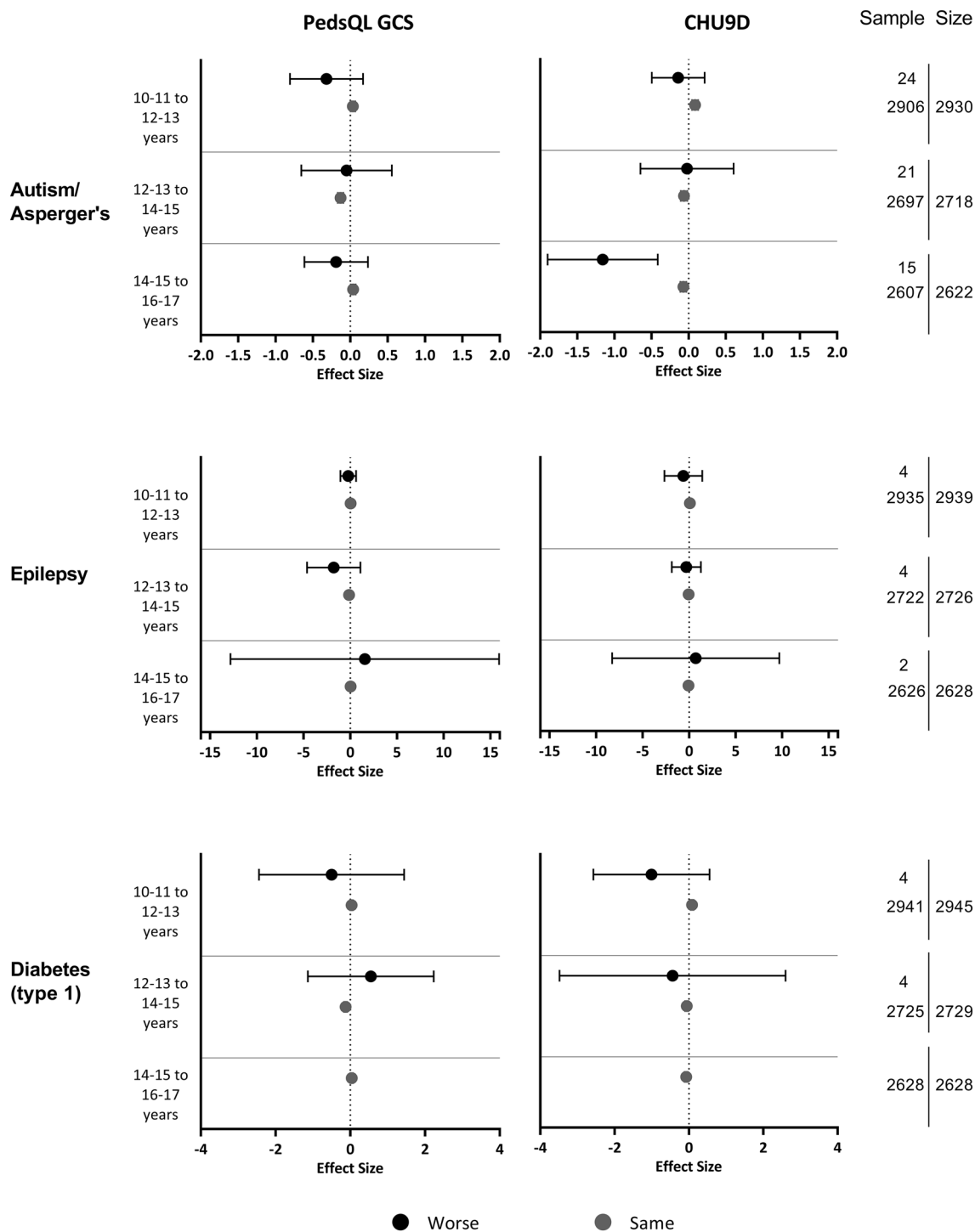


**Fig. 2** Responsiveness results for PedsQL GCS and CHU9D, i.e. effect sizes and 95% CIs for changes in condition status for asthma, anxiety/depression, and ADHD over the 3 age group progressions. The effect size point estimates would be expected to trend from *left to right* for a responsive instrument, with a negative effect size expected

for the “worse” category, an effect size close to zero for “same”, and a positive effect size for “better”. *ADHD* attention deficit hyperactivity disorder, *CHU9D* Child Health Utility 9D, *CI* confidence interval, *PedsQL GCS* Pediatric Quality of Life Inventory™ v4.0 Generic Core Scales

and the CHU9D being child self-reported in the LSAC dataset. A systematic review concluded that there were differences in the level of agreement between proxy-reported

and self-reported HRQOL in children and adolescents [52]. Moreover, in their systematic review, which included preference- and non-preference-based measures, Jardine and



**Fig. 3** Responsiveness results for PedsQL GCS and CHU9D, i.e. effect sizes and 95% CIs for changes in condition status for autism/Asperger's, epilepsy, and type 1 diabetes over the 3 age group progressions. The effect size point estimates would be expected to trend from left to right for a responsive instrument, with a negative effect

size expected for the “worse” category and an effect size close to zero for “same”. CHU9D Child Health Utility 9D, CI confidence interval, PedsQL GCS Pediatric Quality of Life Inventory™ v4.0 Generic Core Scales

colleagues [53] noted that children and parents differ in their perception of HRQOL, particularly for subjective domains (e.g. emotional, psychosocial). This further supports the

potential contribution of a proxy-reported instrument to the poor correlation noted between the PedsQL GCS and CHU9D in this study. Although both instruments are

purported to measure the same construct, i.e. HRQOL, the PedsQL GCS and CHU9D have different numbers of items (23 and 9 items, respectively), and different scoring systems, and may therefore capture different aspects of HRQOL slightly differently from each other. Each instrument may also include different types of items, which means that different aspects of the construct are measured, thereby potentially reducing association within the convergent validity assessment. The poor convergent validity between the PedsQL GCS and CHU9D also highlights the differences in the purpose of these instruments and their content validity, which are important considerations for instrument selection. The PedsQL GCS is designed to measure HRQOL on an unweighted scale, while the CHU9D is a preference-based HRQOL instrument designed to measure health utility that is weighted by preferences. They also differ in the instrument development process. The PedsQL GCS builds on previous iterations of instrument development, and only the initial PedsQL measurement model was developed in a cancer patient cohort aged 8–18 years old [12, 30]. While the CHU9D has been validated for use in general population adolescents aged 11–17 years old, it was initially developed and validated in two paediatric samples (aged 7–11 years old), i.e. a general sample and a clinical sample that included a wide range of health conditions and severity [17–19, 31]. Although not formally evaluated in this study, the choice between the two instruments should also take account of the content validity and original purpose of the instrument.

Known group validity and responsiveness are important properties for the measurement of HRQOL in clinical trials as intervention effectiveness is contingent on being able to detect meaningful differences between groups and changes over time because of the intervention. By using a single cohort, this study demonstrated that for known group validity the PedsQL GCS was more sensitive than the CHU9D at discriminating between the same children and adolescents with and without the health conditions, with larger changes in HRQOL scores estimated for the PedsQL GCS than for the CHU9D for most of the health conditions except anxiety/depression and asthma. The incorporation of a larger number of items in the total score for the PedsQL GCS and/or the questions themselves may be more relevant to the impacts of these health conditions and may account for the PedsQL GCS demonstrating stronger performance with known group validity than the CHU9D. The variable responsiveness found for the CHU9D may be due to the ceiling effects as evidenced by the clustering of the utility scores at the higher end of the scale, which may impair its ability to detect change over time. This evaluation raises the question of whether the CHU9D is sensitive enough to detect changes in condition status over time in the common health conditions in this study. Potential implications for this issue

relate to the use of the CHU9D to calculate QALYs, which is relevant to cost-utility analyses in economic evaluations.

#### 4.1 Strengths and Limitations

Strengths of this study include its large, diverse, and population-representative sample ( $n = 15,568$  observations from 7201 children) drawn from longitudinal data that enabled the assessment of multiple psychometric properties for the PedsQL GCS and CHU9D in a single cohort of Australian children and adolescents. The longitudinal nature allowed us to investigate responsiveness, which is not often possible. The psychometric methods used were rigorous and based on established practice guidelines and criteria.

This study had limitations, which included the potential for reporting bias as the health conditions were proxy-reported by the child's parent, did not report severity of conditions, and were not supported by clinical or other data. Additionally, the PedsQL GCS data were limited to parent proxy-reporting of PedsQL GCS as the LSAC did not use the child self-report versions, while CHU9D was child self-reported. This may have impacted the broad comparison between the two instruments against the absolute set of psychometric assessment criteria. As this was a secondary analysis using an existing dataset from the LSAC, some of the analysis and methods were constrained by the availability of variables and measures in this existing dataset.

Patient burden and comprehension and test-retest reliability could not be assessed using the LSAC data. Both these psychometric properties are important features to assess in future studies for PedsQL GCS and CHU9D. Future work with the modern psychometric paradigm (e.g. based on Rasch measurement theory) could undertake analyses that considered the extent to which the data for each health condition showed differential item functioning by age, as this was beyond the scope of this study, which adhered to the classical psychometric paradigm. With PedsQL GCS being one of the most commonly used paediatric PROMs, there is also the potential for future research to build on the psychometric evidence base for the PedsQL GCS and develop it into an HRQOL preference-based instrument for use in economic evaluations.

## 5 Conclusion

This study provides a valuable contribution to the evidence base of psychometric performance of the PedsQL GCS and CHU9D in a large representative cohort of Australian children and adolescents with common chronic health conditions. The PedsQL GCS (parent proxy-report) instrument demonstrated consistent psychometric robustness against a set of absolute criteria for the psychometric properties

assessed in this study. The CHU9D (child self-report) also demonstrated good psychometric performance; however, it was more variable in its performance with some of the psychometric properties such as acceptability and known group validity. Evidence and discussion from this study can aid in appropriate HRQOL instrument selection for the required context by researchers and clinicians, to ensure that clinical and policy decision-making outcomes are based on consistent and robust research and evaluations.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40258-023-00836-2>.

#### Declarations

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions. This research has been funded by the Australian Government's Medical Research Future Fund (MRFF), under grant APP1199902. Author R. Raghunandan was funded through a Medical Research Futures Fund Preventive and Public Health Research Grant, APP1199902. Author S. Petrou receives support as a UK National Institute for Health Research Senior Investigator (NF-SI-0616-10103) and from the UK NIHR Applied Research Collaboration Oxford and Thames Valley. The funders 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; and decision to submit the manuscript for publication.

**Conflict of interest** The authors declare no conflicts of interest or competing interests.

**Ethics Approval** This study was approved by the University of Sydney Human Ethics Committee (project number 2021/139).

**Consent to Participant** Informed consent was obtained from all participants of the LSAC to use their data to further the understanding of child and adolescent development and wellbeing.

**Consent for Publication** Informed consent was obtained from all participants of the LSAC to use their data to further the understanding of child and adolescent development and wellbeing.

**Availability of Data and Material** The data from the Longitudinal Study of Australian Children (LSAC) used in this study are available by application to the data custodians: Longitudinal Studies, Data Strategy Branch, Australian Government Department of Social Services.

**Code Availability** The STATA code used in the analysis can be made available upon reasonable request from the authors.

**Authors' Contributions** RR, AH, KH, SP, EL, and SS conceived the study. RR and AH carried out the analyses. EC and AK provided statistical support. RR wrote the first draft of the manuscript. All authors contributed to the manuscript.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the

material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## References

1. Kwon J, Freijser L, Huynh E, Howell M, Chen G, Khan K, et al. Systematic review of conceptual, age, measurement and valuation considerations for generic multidimensional childhood patient-reported outcome measures. *Pharmacoeconomics*. 2022;40(4):379–431. <https://doi.org/10.1007/s40273-021-01128-0>.
2. Janssens A, Thompson Coon J, Rogers M, Allen K, Green C, Jenkinson C, et al. A Systematic Review of Generic Multidimensional Patient-Reported Outcome Measures for Children, Part I: Descriptive Characteristics. *Value in Health*. 2015;18(2):315–33. <https://doi.org/10.1016/j.jval.2014.12.006>.
3. Verstraete J, Scott D. Comparison of the EQ-5D-Y-5L, EQ-5D-Y-3L and PedsQL in children and adolescents. *Journal of Patient-Reported Outcomes*. 2022;6(1):67. <https://doi.org/10.1186/s41687-022-00480-9>.
4. Janssens A, Rogers M, Thompson Coon J, Allen K, Green C, Jenkinson C, et al. 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. *Value Health*. 2015;18(2):334–45. <https://doi.org/10.1016/j.jval.2015.01.004>.
5. Williams KSJ, Morris D, Grootemaat P, Thompson C. Patient-reported outcome measures: literature review. Sydney: ACSQHC; 2016.
6. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick 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(1):22. <https://doi.org/10.1186/1471-2288-10-22>.
7. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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(7):737–45. <https://doi.org/10.1016/j.jclinepi.2010.02.006>.
8. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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(4):539–49. <https://doi.org/10.1007/s11136-010-9606-8>.
9. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013;22(8):1889–905. <https://doi.org/10.1007/s11136-012-0344-y>.
10. Food and Drug Administration. Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims>. Accessed 12 Dec 2022.
11. Arsiwala T, Afroz N, Kordy K, Naujoks C, Patalano F. Measuring what matters for children: a systematic review of

- frequently used pediatric generic pro instruments. *Therap Innov Regul Sci.* 2021;55(5):1082–95. <https://doi.org/10.1007/s43441-021-00311-x>.
12. Varni JW, Seid M, Rode CA. The PedsQL™: Measurement Model for the Pediatric Quality of Life Inventory. *Med Care.* 1999;37(2):126–39.
  13. Sprangers MA, Cull A, Bjordal K, Groenvold M, Aaronson NK. The European Organization for Research and Treatment of Cancer. Approach to quality of life assessment: guidelines for developing questionnaire modules. *Qual Life Res.* 1993;2(4):287–95. <https://doi.org/10.1007/bf00434800>.
  14. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. *Ambul Pediatr.* 2003;3(6):329–41. [https://doi.org/10.1367/1539-4409\(2003\)003%3c0329:tpaapp%3e2.0.co;2](https://doi.org/10.1367/1539-4409(2003)003%3c0329:tpaapp%3e2.0.co;2).
  15. Varni JW. The PedsQL Measurement Model for the Paediatric Quality of Life Inventory—About the Model [Internet]. 2022. Available from: [https://www.pedsqol.org/about\\_pedsqol.html](https://www.pedsqol.org/about_pedsqol.html). Accessed 1 Nov 2022.
  16. Stevens K. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Qual Life Res.* 2009;18(8):1105–13. <https://doi.org/10.1007/s11136-009-9524-9>.
  17. Stevens KJ. Working with children to develop dimensions for a preference-based, generic, pediatric, health-related quality-of-life measure. *Qualit Health Res.* 2010;20(3):340–51. <https://doi.org/10.1177/1049732309358328>.
  18. Stevens K. Assessing the performance of a new generic measure of health-related quality of life for children and refining it for use in health state valuation. *Appl Health Econ Health Policy.* 2011;9(3):157–69. <https://doi.org/10.2165/11587350-00000000-00000>.
  19. Ratcliffe J, Stevens K, Flynn T, Brazier J, Sawyer M. An assessment of the construct validity of the CHU9D in the Australian adolescent general population. *Qual Life Res.* 2012;21(4):717–25. <https://doi.org/10.1007/s11136-011-9971-y>.
  20. Andrich DSI. Report on the psychometric analysis of the early development instrument (EDI) using the Rasch model. Perth, WA: Murdoch University; 2004.
  21. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess.* 2009;13(12):iii. <https://doi.org/10.3310/hta13120>. (ix-x, 1-177).
  22. Rowen D, Keetharuth AD, Poku E, Wong R, Pennington B, Wailoo A. 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.* 2021;24(3):443–60. <https://doi.org/10.1016/j.jval.2020.09.012>.
  23. Yang P, Chen G, Wang P, Zhang K, Deng F, Yang H, et al. Psychometric evaluation of the Chinese version of the Child Health Utility 9D (CHU9D-CHN): a school-based study in China. *Qual Life Res.* 2018;27(7):1921–31. <https://doi.org/10.1007/s11136-018-1864-x>.
  24. Petersen KD, Ratcliffe J, Chen G, Serles D, Frøsig CS, Olesen AV. The construct validity of the Child Health Utility 9D-DK instrument. *Health Qual Life Outcomes.* 2019;17(1):187. <https://doi.org/10.1186/s12955-019-1256-0>.
  25. Lindvall K, Vaezghasemi M, Feldman I, Ivarsson A, Stevens KJ, Petersen S. Feasibility, reliability and validity of the health-related quality of life instrument Child Health Utility 9D (CHU9D) among school-aged children and adolescents in Sweden. *Health Qual Life Outcomes.* 2021;19(1):193. <https://doi.org/10.1186/s12955-021-01830-9>.
  26. Catchpool M, Gold L, Grobler AC, Clifford SA, Wake M. Health-related quality of life: population epidemiology and concordance in Australian children aged 11–12 years and their parents. *BMJ Open.* 2019;9(Suppl 3):157. <https://doi.org/10.1136/bmjopen-2018-022398>.
  27. Petersen KD, Chen G, Mpundu-Kaambwa C, Stevens K, Brazier J, Ratcliffe J. Measuring health-related quality of life in adolescent populations: an empirical comparison of the CHU9D and the PedsQL™ 40 Short Form 15. *The Patient Patient-Cent Outcomes Res.* 2018;11(1):29–37. <https://doi.org/10.1007/s40271-017-0265-5>.
  28. Mohal J, Lansangan C, Gasser C, Howell L, Duffy J, Renda J, Scovelle A, Jessup K, Daraganova G, Mundy L. Growing up in Australia: The Longitudinal Study of Australian Children—Data User Guide, Release 9.0C2, June 2022. Melbourne: Australian Institute of Family Studies. 2022. Accessed 5 Oct 2022.
  29. Australian Institute of Family Studies. Growing Up In Australia – The Longitudinal Study of Australian Children, Annual Statistical Report 2018. Melbourne: Australian Institute of Family Studies. 2018. Accessed 5 Oct 2022.
  30. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Med Care.* 2001;39(8):800–12. <https://doi.org/10.1097/00005650-200108000-00006>.
  31. 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(8):1092–9. <https://doi.org/10.1016/j.jval.2012.07.011>.
  32. Ratcliffe J, Huynh E, Chen G, Stevens K, Swait J, Brazier J, 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. <https://doi.org/10.1016/j.socscimed.2016.03.042>.
  33. Varni JW, Limbers CA, Burwinkle TM. Impaired health-related quality of life in children and adolescents with chronic conditions: a comparative analysis of 10 disease clusters and 33 disease categories/severities utilizing the PedsQL™ 4.0 Generic Core Scales. *Health Qual Life Outcomes.* 2007;5(1):43. <https://doi.org/10.1186/1477-7525-5-43>.
  34. Killedar A, Lung T, Petrou S, Teixeira-Pinto A, Tan EJ, Hayes A. Weight status and health-related quality of life during childhood and adolescence: effects of age and socioeconomic position. *Int J Obes.* 2020;44(3):637–45. <https://doi.org/10.1038/s41366-020-0529-3>.
  35. Vella SA, Magee CA, Cliff DP. Trajectories and Predictors of Health-Related Quality of Life during Childhood. *J Pediatr.* 2015;167(2):422–7. <https://doi.org/10.1016/j.jpeds.2015.04.079>.
  36. Jansen PW, Mensah FK, Clifford S, Nicholson JM, Wake M. Bidirectional associations between overweight and health-related quality of life from 4–11 years: Longitudinal Study of Australian Children. *Int J Obes.* 2013;37(10):1307–13. <https://doi.org/10.1038/ijo.2013.71>.
  37. Rajmil L, Herdman M, Ravens-Sieberer U, Erhart M, Alonso J. Socioeconomic inequalities in mental health and health-related quality of life (HRQOL) in children and adolescents from 11 European countries. *Int J Public Health.* 2014;59(1):95–105. <https://doi.org/10.1007/s00038-013-0479-9>.
  38. Otto C, Haller A-C, Klasen F, Hölling H, Bullinger M, Ravens-Sieberer U, et al. Risk and protective factors of health-related quality of life in children and adolescents: results of the longitudinal BELLA study. *PLoS ONE.* 2017;12(12): e0190363. <https://doi.org/10.1371/journal.pone.0190363>.
  39. Baker K, Siphthorp M, Edwards B. A Longitudinal Measure of Socioeconomic Position in LSAC. Australian Institute of Family Studies. 2017. Accessed 1 Sep 2021.



40. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Econ.* 1999;8(1):41–51.
41. Smith SC, Lamping DL, Banerjee S, Harwood R, Foley B, Smith P, et al. Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technol Assess.* 2005;9(10):1–93. <https://doi.org/10.3310/hta9100>. (iii-iv).
42. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11(3):193–205.
43. Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw-Hill; 1994.
44. Streiner DL, Norman GR. *Health measurement scales : a practical guide to their development and use*. 2nd ed. Oxford: Oxford University Press, Oxford Medical Publications; 1995.
45. Von Rueden U, Gosch A, Rajmil L, Bisegger C, Ravens-Sieberer U. Socioeconomic determinants of health related quality of life in childhood and adolescence: results from a European study. *J Epidemiol Community Health.* 2006;60(2):130–5. <https://doi.org/10.1136/jech.2005.039792>.
46. Tonge B, Brereton A. Autism spectrum disorders. *Austral J Gener Pract.* 2012;40:672–7.
47. Allers K, Essue BM, Hackett ML, Muhunthan J, Anderson CS, Pickles K, et al. The economic impact of epilepsy: a systematic review. *BMC Neurol.* 2015;15(1):245. <https://doi.org/10.1186/s12883-015-0494-y>.
48. Australian Institute of Health Welfare. *Diabetes: Australian facts*. Canberra: AIHW. 2022. <https://www.aihw.gov.au/reports/diabetes/diabetes>. Accessed 12 Dec 2022.
49. Pink J, Petrou S, Williamson E, Williams M, Lamb SE. Properties of patient-reported outcome measures in individuals following acute whiplash injury. *Health Qual Life Outcomes.* 2014;12:38. <https://doi.org/10.1186/1477-7525-12-38>.
50. Cohen J. *Statistical power analysis for the behavioural sciences*. 2nd ed. New Jersey: Routledge; 1988.
51. Chen G, Flynn T, Stevens K, Brazier J, Huynh E, Sawyer M, 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(4):432–8. <https://doi.org/10.1016/j.jval.2015.02.014>.
52. Khadka J, Kwon J, Petrou S, Lancsar E, Ratcliffe J. Mind the (inter-rater) gap. An investigation of self-reported versus proxy-reported assessments in the derivation of childhood utility values for economic evaluation: a systematic review. *Soc Sci Med.* 2019;240:112543. <https://doi.org/10.1016/j.socscimed.2019.112543>.
53. Jardine J, Glinianaia SV, McConachie H, Embleton ND, Rankin J. Self-reported quality of life of young children with conditions from early infancy: a systematic review. *Pediatrics.* 2014;134(4):e1129–48. <https://doi.org/10.1542/peds.2014-0352>.

## Authors and Affiliations

Rakhee Raghunandan<sup>1,2</sup>  · Kirsten Howard<sup>1,2</sup>  · Sarah Smith<sup>3</sup>  · Anagha Killedar<sup>2</sup>  · Erin Cvejic<sup>1</sup>  · Martin Howell<sup>1,2</sup>  · Stavros Petrou<sup>4</sup>  · Emily Lancsar<sup>5</sup>  · Germaine Wong<sup>1</sup>  · Jonathan Craig<sup>6</sup>  · Alison Hayes<sup>1</sup> 

✉ Rakhee Raghunandan  
rakhee.raghunandan@sydney.edu.au

<sup>1</sup> School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, NSW 2006, Australia

<sup>2</sup> Menzies Centre for Health Policy and Economics, Faculty of Medicine and Health, University of Sydney, Charles Perkins Centre D17, Sydney, NSW 2006, Australia

<sup>3</sup> Department of Health Services Research and Policy, London School of Hygiene & Tropical Medicine, London, England, UK

<sup>4</sup> Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

<sup>5</sup> Department of Health Services Research and Policy, Australian National University, Canberra, Australia

<sup>6</sup> College of Medicine and Public Health, Flinders University, Adelaide, Australia