



An Investigation of Inter-Rater and Intra-Proxy Agreement in Measuring Quality of Life of Children in the Community Using the EQ-5D-Y-3L

Diana Khanna¹ · Jyoti Khadka^{1,2} · Christine Mpundu-Kaambwa¹ · Gang Chen⁵ · Kim Dalziel^{3,4} · Nancy Devlin³ · Julie Ratcliffe¹ · in collaboration with the Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team

Accepted: 9 November 2023
© The Author(s) 2024

Abstract

Background Self-reporting of health-related quality of life (HRQoL) in children is not always feasible. To date, proxy perspectives (Proxy versions 1 and 2) using the EQ-5D-Y-3L have not been explored for its impact on agreement with child self-report. Proxy version 1 requires the proxy to consider their own view of the child's HRQoL (proxy-proxy), while with Proxy version 2, the proxy is asked to respond as they believe their child would self-report their HRQoL (proxy-child). This study compared the inter-rater and intra-proxy agreement (overall and dimension level) using the EQ-5D-Y-3L self, proxy-proxy, and proxy-child reports.

Methods A community-based sample of child (aged 6–12 years) and parent dyads were invited to participate in a semi-structured interview. The child self-completed the EQ-5D-Y-3L independently of the parent who completed the EQ-5D-Y-3L from proxy-proxy and proxy-child perspectives. Agreement was determined using Concordance Correlation Coefficients (CCCs) for the overall (preference-weighted) HRQoL, while agreement at the dimension level was evaluated using Gwet's agreement coefficient (AC_1). To assess the differences between the self and the two proxy reports, the Wilcoxon matched-pair signed-rank test was used.

Results This study involved 85 child-parent dyads. The agreement between self and proxy overall HRQoL was low (fair) with both proxy-proxy ($CCC = 0.28$) and proxy-child ($CCC = 0.26$) reports. The largest discrepancy in the child-proxy agreement at dimension level with both the proxy versions was observed for 'feeling worried, sad or unhappy'. Within this dimension, the proxy-child perspective resulted in a stronger agreement ($AC_1 = 0.7$, good) with child self-report compared with the traditional proxy-proxy perspective ($AC_1 = 0.58$, moderate). Although the preference-weighted HRQoL was consistent across both the proxy perspectives, a significant difference was observed in the EQ VAS scores ($p = 0.02$).

Conclusions This study demonstrates that choice of proxy perspective may have an impact on the problems reported on HRQoL dimensions and EQ VAS scores. However, in this community-based sample of generally healthy children, no significant difference was observed in the inter-rater agreement for child-self and proxy preference-weighted EQ-5D-Y-3L values based on proxy perspectives. While this suggests that preference-weighted data are not sensitive to the choice of perspective, these findings may differ for different HRQoL instruments and for alternative value sets with different properties.

1 Introduction

Measurement and valuation of health-related quality of life (HRQoL) in children is crucial for population health studies and for the assessment of outcomes, service quality, effectiveness and cost effectiveness of health and social care services for children [1, 2]. A number of concise generic child-specific HRQoL measures are available for

use in inviting children to self-report their health [3]. The EuroQol group have developed and validated two child-specific measures, the EQ-5D-Y three-level (EQ-5D-Y-3L) and five-level (EQ-5D-Y-5L). The EQ-5D-Y-3L is a youth version of the original EQ-5D-3L, an adult-specific HRQoL measure commonly used in economic evaluations. As with the EQ-5D-3L for adults, the EQ-5D-Y-3L includes five dimensions with three response levels per dimension, with wording adapted for use in child populations. Using a preference-based algorithm, the dimensions are synthesised

Extended author information available on the last page of the article

Key Points for Decision Makers

There are two ways to complete a proxy report—from the proxy’s own perspective (proxy-proxy, Proxy 1) or answer as the child would (proxy-child, Proxy 2). The implications of utilising these two perspectives on agreement with child self-report is not clear.

Based on a community-based dyadic sample comprising generally healthy children and their parents, who completed the EQ-5D-Y-3L (self, proxy-proxy and proxy-child), we found that proxy perspective influenced the agreement between child and proxy ratings of health-related quality of life (HRQoL).

Proxy-child perspective showed a stronger agreement at the dimension level for the psychosocial dimension compared with the traditional proxy-proxy perspective. While no statistically significant difference was observed for the preference-weighted HRQoL across the two proxy perspectives, the child- and parent-reported EQ VAS scores differed significantly when the proxy-child perspective was adopted, indicating that perspective may influence this aspect of HRQoL measurement.

into a single value representing the preference-weighted HRQoL. In addition, the EQ-5D-Y-3L includes a visual analogue scale (VAS) for indicating the level of health on a scale of 0 to 100, with 0 indicating the worst and 100 indicating the best possible health [3].

When feasible, the EQ-5D-Y measures (3L and 5L) have been recommended for self-completion by children aged ≥ 8 years. However, the minimum age at which children can accurately self-report their own HRQoL remains in question. Children aged 8 years and older are generally considered reliable for self-reporting HRQoL measures [9] and a number of studies have successfully administered the EQ-5D-Y-3L in cohorts of children aged 8–18 years [4–7]. Nevertheless, some studies have reported successful administration of the EQ-5D-Y-3L in younger populations. For example, Canaway and Frew demonstrated the feasibility and acceptability of the interviewer-administered EQ-5D-Y-3L measure in children aged 6–7 years [8]. In a study involving children aged 6–17 years, Gusi et al. showed the validity and reliability of the Spanish version of the EQ-5D-Y-3L [9]. Bray and colleagues also utilised a subsample of children aged 6–7 years to assess HRQoL in children with impaired mobility using measures including the EQ-5D-Y [10].

While self-reported child HRQoL measures are important, several methodological challenges remain. These challenges in assessing child-self and proxy reported HRQoL are nuanced and distinct from those in adult populations. In comparison with the adult population, the self-reporter in these instances is a child, introducing potential disparities in the interpretation of HRQoL dimensions. Such disparities may stem from differences in cognitive development stages, where a child’s reasoning may not align with that of an adult’s, or contextual factors such as peer influence or social dynamics at school [11]. A recent systematic review reported that children with cognitive processing challenges, particularly attention-deficit hyperactivity disorder (ADHD), learning disability, speech impairments or special health care needs, are more likely to have limited self-report capacity [12]. Children may also be unable to self-report their own HRQoL if they are too young or due to illness or lack of capacity [13, 14]. In such situations, parents, caregivers, teachers and/or health professionals may act as proxies to provide an informed estimate of the child’s HRQoL on their behalf.

In light of these limitations in self-reporting, the EQ-5D-Y-3L measure offers two proxy versions (Proxy versions 1 and 2), which differ in terms of the perspective the proxy is asked to adopt [4]. The EQ-5D-Y-3L Proxy version 1 uses the ‘proxy-proxy’ perspective, where the proxy is asked to think about their own view of the child’s HRQoL, while Proxy version 2 uses the ‘proxy-child’ approach, whereby the proxy is asked to respond as they believe their child would if they were reporting their own HRQoL [4]. The chosen proxy perspective provides assessments of child HRQoL that either reflect a viewpoint that may differ from the child’s own (proxy-proxy) or substituted judgement (proxy-child) while maintaining the construct validity of the measure [15].

The perspective from which HRQoL is valued is an important consideration in the context of health technology assessments [16, 17]. Likewise, in the measurement of HRQoL, the additional information obtained from various perspectives can contribute to a comprehensive understanding of the HRQoL [15] and offer additional methodological rigor by offering opportunities for triangulation of perspectives [18]. In adult populations, two previous studies conducted using the EQ-5D measures to assess HRQoL have compared the two proxy perspectives. One study found no systematic difference between self and proxy assessments from the two perspectives by clinicians (physiotherapists) in older hospital patients with intact cognition [19]. Another study found a higher inter-rater agreement using the EQ-5D-5L proxy-person perspective in an orthopaedic population than when the proxy-proxy perspective was adopted [20]. However, to our knowledge, no study to date has investigated the inter-rater agreement

with respect to the two proxy perspectives in a paediatric population. There is a need to determine whether the proxy perspective is useful depending on the degree to which it supports or complements information regarding the child's HRQoL (proxy-proxy) or attempts to replicate and substitute for child's self-assessment (proxy-child) [15]. This is identified through the difference between the HRQoL ratings produced using the two proxy versions, namely the intra-proxy gap.

A recent systematic review by our team examined the inter-rater agreement between child self-report and proxy-reports for preference-weighted generic HRQoL measures in children (below 18 years of age) [21]. The review included 30 published studies that reported the overall and/or dimension-level HRQoL agreement between proxies and children with and without health conditions, such as cancer, type 1 and 2 diabetes, asthma and cerebral palsy. Overall HRQoL was reported to have poor inter-rater agreement (intraclass correlation coefficient [ICC] < 0.5) in most of the identified studies. Psychosocial-related dimensions, e.g., 'feeling worried, sad, or unhappy' and 'having pain or discomfort' dimensions of the EQ-5D showed lower inter-rater agreement compared with dimensions related to physical health. Compared with health professionals, parents had higher levels of agreement with their children. Importantly, none of the reported studies that examined inter-rater agreement between self- and proxy-reports for preference-weighted generic HRQoL measures in children have compared the two proxy versions.

This study in a community-based sample therefore aims to measure (1) the inter-rater agreement, i.e., the level of agreement between EQ-5D-Y-3L responses (overall and dimension-level HRQoL) produced by child self-report and their parent-proxy (proxy-proxy and proxy-child) reports; and (2) the intra-proxy agreement, i.e., the overall and dimension-level agreement between the two proxy versions of the EQ-5D-Y-3L.

2 Methods

2.1 Participants and Study Design

Participant recruitment was conducted through a partnership between the research team and an independent social research company (Stable Research Australia). An invitation letter outlining the details of this study was sent to an active online panel of parents who had previously indicated their own and their child's interest in participating in research studies. Children aged 6–12 years and one of their parents living in the same household (i.e., parent/child dyads) were eligible to participate in this cross-sectional study according to prespecified inclusion and exclusion criteria.

The inclusion criteria required that children were able to read and understand written English. Additionally, children with reading disorders such as dyslexia were excluded. This study administered the child self-report questionnaire using REDCap, an online platform on a laptop embedded with a screen-based eye-tracker.¹ Therefore, criteria for exclusion also comprised contraindications for eye tracking, including eye conditions such as lazy eye (amblyopia), misaligned eyes (strabismus), and dancing eyes (nystagmus).

Information about the child participant's age, sex and household income was collected from parents following informed consent to participate from both the parent and child at the pre-interview stage. Parents also reported on any long-term child health condition(s): "Has your child been diagnosed by a health or education professional with a long-standing illness, medical condition, or disability? (yes/no)". If yes, parents were asked to specify the condition.

A broad representation in relation to key sociodemographic characteristics and common health conditions affecting children in the general population, such as asthma, anxiety disorders, conduct disorders, depressive disorders, autism spectrum disorders (ASD) and dental caries, was achieved using a stratified random sampling method [22]. Previous studies suggest a high percentage agreement (approximately 60%) can be expected between proxies and children from the general population [7, 9]. While Gwet recommends a sample size of 25 to estimate the agreement coefficient with an error margin of 20% under high percentage agreement [23], this study aimed for a larger sample size ($N > 25$) to increase the statistical power of the agreement analysis.

The study was conducted in South Australia and complied with the ethical guidelines of the Flinders University Human Research Ethics Committee (Project ID 4178).

2.2 Procedure

Child-parent dyads were invited to attend a semi-structured, face-to-face interview with a researcher at Flinders University. In the interview, the child was invited to self-complete their own HRQoL assessment using the EQ-5D-Y-3L measure and a self-rated general health (SRH) question, "In general, would you say your health is poor, fair, good, very good, or excellent?" [24], administered online via the REDCap software.

Simultaneously, the parent respondent was asked to self-complete both proxy versions of the EQ-5D-Y-3L as hard copies while using noise-cancelling headphones such

¹ The eye-tracking data, focusing on the analysis of children's gaze patterns and other metrics, will be explored as the subject of a subsequent paper.

that their responses were not unduly influenced by any conversations taking place between the interviewer and the child, and to ensure they were not influencing their child's responses. Moreover, the interviewer had minimal contact with the parent at the time of survey completion to mitigate any social desirability bias on behalf of the parent that might otherwise occur in an interviewer-led mode of administration [25]. The parent first completed the (traditional) proxy-proxy version on behalf of the child. This was followed by an assessment of their own HRQoL using the EQ-5D-3L. Following this, the (alternative) EQ-5D-Y-3L proxy-child version was administered to allow the parent time to switch between the perspectives of the two proxy versions. The proxy version asked the parent to select one statement that (1) 'you think best describes your child's health today' (proxy-proxy); and (2) 'you think your child would choose to describe their health today' (proxy-child). In addition, the parent completed a general health SRH item about themselves and a series of sociodemographic questions including their age, sex, and postcode.

2.3 Measures

The EQ-5D-Y-3L and its proxy versions were used to examine inter-rater agreement between self- and proxy-reported HRQoL [26]. For proxies, in version 1, the proxy is asked to rate their child's HRQoL according to their opinion (proxy-proxy), while in version 2, they are asked how the child would rate their own HRQoL if they were able to do so (proxy-child). There are five dimensions within the EQ-5D-Y-3L and its proxy versions: 'walking about', 'looking after myself', 'doing usual activities', 'having pain or discomfort', and 'feeling worried, sad or unhappy'. For each dimension, the respondent can indicate severity on any of three levels of problems (no problems, some problems, a lot of problems). The EQ-5D-Y-3L self-report and both its proxy versions also include a visual analogue scale (EQ VAS), where the respondent can rate their or their child's overall health status (or that of their own when self-reporting) on a scale from 0 to 100, with 0 indicating the worst possible state and 100 indicating the best possible state.

An Australian value set for the EQ-5D-Y-3L is not currently available. Hence, the EQ-5D-3L Australian adult value set derived using a time trade-off (TTO) approach was applied to both proxy and self HRQoL ratings to generate the overall HRQoL or HRQoL values (preference-weighted HRQoL) [27]. It is important to note that value sets for adult EQ-5D-3L are known to have different properties than value sets for EQ-5D-Y-3L, e.g., in terms of dimension ordering and length of value scale [17]. However, arguably, such differences are of lesser importance for this particular study, since our purpose is to determine the extent of agreement between the dyads according to the perspective adopted and

to examine any differences in this regard. The robustness of the main findings to the choice of value set was tested in a sensitivity analysis using a recently published EQ-5D-Y-3L value set (for Germany) [28]. Studies from European countries have indicated that the distribution of values for the adult and youth EQ-5D measures are similar to each other. Assuming that Australian values for the EQ-5D-Y-3L are more likely to resemble 'European' rather than 'Asian' preference patterns, the German EQ-5D-Y-3L value set was chosen [29]. The preferences for the EQ-5D-Y-3L health states were elicited from a German adult population using a discrete choice experiment (DCE) and composite TTO (cTTO) methods. The value set was applied to both self and proxy responses.

2.4 Statistical Analysis

Sociodemographic data were analysed using descriptive statistics. The relative socioeconomic disadvantage of the postal area was determined from the Socio-Economic Indexes for Australia (SEIFA) area decile number. The first six decile numbers were classified as disadvantaged quintiles (quintiles 1–3) and the last four as advantaged quintiles (quintiles 4 and 5) [30]. Subgroups were based on age and the presence of long-term health condition/s (yes/no) as reported by the parent. Three age classifications were applied for age group analysis: 6–7 years, 8–10 years and 11–12 years. The age-group segmentation was determined by our aim to contrast the responses from younger children, under the age of 8 years, with those in the older age group for which the EQ-5D-Y-3L is typically recommended for self-completion. Additionally, our sample composition was disproportionately skewed towards older children, with a notable overrepresentation of 11-year-olds as opposed to those aged 8, 9, or 10 years, thus necessitating their separate grouping in the analysis. Inter-rater differences and agreement were analysed for the overall sample and by subgroups for the two proxy versions of the EQ-5D-Y-3L for HRQoL values (preference-weighted), the dimension-level HRQoL and the EQ VAS scores. Further intra-proxy agreement was estimated for the overall and dimension-level HRQoL.

Medians (and interquartile range [IQR]) were used to describe the summary statistics for the HRQoL values and the EQ VAS scores by raters (self-report, proxy-proxy and proxy-child) as most study participants were in relatively good health and the HRQoL values were negatively skewed. Furthermore, agreement was assessed using the Concordance Correlation Coefficient (CCC) for the HRQoL values and VAS scores [31, 32]. The CCC is frequently used to evaluate agreement between two raters and does not rely on the analysis of variance (ANOVA) model assumptions, unlike the ICC [33]. Gwet's agreement coefficient (AC_1)

was used to analyse the dimension-level HRQoL [23]. The unweighted AC_1 was chosen due to the predominance of healthy children in the study sample as well as the EQ-5D-Y-3L's three-level response scale, which together reduce the likelihood of marked disagreements, rendering the weighted Gwet's AC_2 , which could overestimate agreement for adjacent category discrepancies, less advantageous for this study, although its analysis using linear weights is included in the appendix (electronic supplementary material [ESM] Table S1) for completeness [23]. Both CCC and Gwet's AC_1 take values between -1 and 1 , and their magnitude was qualified using Altman's scale for consistency of interpretation. Altman's scale is defined as poor, fair, moderate, good and very good for values less than or equal to 0.2, 0.4, 0.6, 0.8 and 1, respectively [23, 34]. The Wilcoxon matched-pairs signed-rank test was used to compare group differences for continuous variables and Fisher's exact test was used for categorical variables. In this study, the statistical significance level was set at 0.05. Analysis was carried out using Stata 16.1 (StataCorp LLC, College Station, TX, USA) [35].

3 Results

3.1 Sample Characteristics

In total, 89 dyads met the inclusion criteria and were invited to participate in the study. Of these, 85 dyads agreed and participated in the interview (response rate = 96%). Table 1 describes the sociodemographic characteristics of the child-parent respondents in the sample. The mean age of children in this sample was 9.13 years (standard deviation [SD] 2), with a slight overrepresentation of girls (56%). Parents in the sample had a mean age of 41.7 years (SD 5.6) and one-fifth of the dyads were father-child pairs. Unsurprisingly, almost all parents and children reported excellent to good health on the SRH item. Of the 85 children in the sample, 26 (31%) were reported by their parents to have at least one of the following conditions: asthma (42%), ASD (8%), dental caries (15%), ADHD (4%), anxiety/depression (15%), sleep problems (12%) and congenital heart disease (4%). Based on the SEIFA area decile numbers, in comparison with the Australian population, the sample had a lower representation of respondents residing in postcodes associated with relatively disadvantaged quintiles (37%) [36].

3.2 Dyad EQ-5D-Y-3L Values, EQ VAS Scores and Dimension-Level Responses

Table 2 reports the EQ-5D-Y-3L values and EQ VAS scores of the overall dyad sample and by raters and subgroups. Of the 85 dyad participants, two children did not report EQ

VAS scores. When compared with children's self-report, the HRQoL values were underestimated in proxy-proxy reports (self-report: median 1, IQR 0.81–1; proxy-proxy report: median 0.84, IQR 0.8–1). The median (IQR) value for the proxy-child report was identical to the child-self report. Nevertheless, as indicated by the Wilcoxon matched-pair signed-rank test, these differences were not statistically significant. Although the median EQ VAS scores were consistent at 90 across the three reports, the self-report (IQR 75–98) exhibited a greater degree of variability as compared with the two parent-proxy versions. Moreover, despite the identical medians, the child-self- and proxy-child-reported EQ VAS scores were significantly different ($p = 0.02$).

Across the age groups, the only statistically significant difference based on the Wilcoxon matched-pair signed-rank test was observed between older children aged 11–12 years and proxies where parents reported significantly higher EQ VAS scores from the proxy-child perspective. In view of these findings, the 11- to 12-year-old dyad subgroup may potentially account for the significant self- and proxy-child-reported heterogeneity in EQ VAS scores. A difference in medians of 0.2 in the HRQoL values, the largest among the subgroups, was observed between children with a health condition and their parents, from both proxy perspectives, although this difference was not statistically significant. However, the only significant difference within the subgroups based on the presence of long-term health condition/s was noted between the EQ VAS child-self and proxy-child ratings in the subsample of children without any health condition.

Figure 1 presents the distribution of child-self and the two parent-proxy (proxy-proxy and proxy-child perspective) responses for the EQ-5D-Y-3L dimensions. As reported in Table 3, using Fisher's exact test as an omnibus test, statistically significant differences were identified across the report types (self, proxy-proxy and proxy-child) in the dimensions 'walking about' ($p = 0.02$), 'doing usual activities' ($p < 0.001$) and 'feeling worried, sad or unhappy' ($p < 0.001$). Notably, in the 'feeling worried, sad or unhappy' dimension, parents were more likely to report problems than children themselves. Subsequent post hoc pairwise comparison indicated, for the 'walking about' dimension only, significant differences for the self and proxy-proxy perspective ($p = 0.02$), but not between the self and proxy-child perspective ($p = 0.06$). No differences were found between the two proxy perspectives across the dimensions.

Further subgroup analysis yielded statistically significant differences across the three reports among the 6- to 7-year-olds for 'walking about' ($p = 0.03$) and 'doing usual activities' ($p < 0.01$), and among the 8- to 10-year-olds for 'doing usual activities' ($p < 0.01$). In contrast to children with reported health conditions, among the children categorised

Table 1 Sociodemographic characteristics for all study participants (children and parent proxies)

	Child [<i>N</i> = 85]	Parent [<i>N</i> = 85]
Age, years		
Mean (SD)	9.13 (2)	41.7 (5.6)
Median (IQR)	9 (7–11)	41 (37–46)
Sex		
Female	47 (56)	68 (81)
Male	37 (44)	16 (19)
SRGH		
Excellent	20 (24)	34 (41)
Very good	44 (53)	43 (51)
Good	16 (19)	6 (7)
Fair	3 (4)	1 (1)
Long-term health condition/s		
Yes	26 (31)	–
No	59 (69)	–
Health condition		
Mental or behavioural disorder	7 (27)	–
Asthma	11 (42)	–
Dental caries	4 (15)	–
Congenital heart disease	1 (4)	–
Sleep disorders	3 (12)	–
Socioeconomic condition according to postcode		
Relatively advantaged quintile (SEIFA decile 7, 8, 9, 10)	–	52 (63)
Relatively disadvantaged quintile (SEIFA decile 1, 2, 3, 4, 5, 6)	–	31 (37)

Data are expressed as *n* (%) unless otherwise specified

SEIFA Socio-Economic Indexes for Australia, SD standard deviation, IQR interquartile range, SRGH self-rated general health

as relatively healthy (no reported health condition), a significant difference was observed across all the three dimensions: ‘walking about’ ($p < 0.01$), ‘doing usual activities’ ($p < 0.001$) and ‘feeling worried, sad or unhappy’ ($p < 0.01$) [see Table 3].

3.3 Inter-Rater Agreement for EQ-5D-Y-3L Values and Dimensions

Table 4 presents the dyad agreement for overall HRQoL and across dimensions, by rater and subgroups, along with the 95% confidence intervals (CI). The dyadic agreement using CCC was slightly higher for proxy-proxy (0.28) than proxy-child (0.26). For both the younger age groups (6–7 years and 8–10 years), a fair level of parent-child agreement was observed, with a higher agreement for proxy-child than proxy-proxy. In contrast, dyads with older children aged 11–12 years reported a poor level of agreement (CCC < 0.2) regardless of the perspective, with almost no agreement when the proxy-child report was considered. Similarly, a poor self and proxy-child agreement was observed in the dyad comprising children without any health condition.

The dimension-level agreement ranged between good and very good, with AC₁ values exceeding 0.6 for all dimensions using the two proxy versions, except for ‘feeling worried, sad or unhappy’. The highest level of agreement was observed for the physical health-related dimension of ‘walking about’, followed by ‘looking after myself’ and ‘doing usual activities’. The agreement within the ‘having pain/discomfort’ dimension was good but relatively lower with both versions. When comparing the two proxy versions, proxy-proxy report showed only a moderate agreement (0.58) for the ‘feeling worried, sad or unhappy’ dimension, while proxy-child report provided a higher (good) agreement estimate (0.7).

Inter-rater agreement was mostly consistent across both subgroups (as categorised by age groups and presence of a long-term health condition) for both versions within all dimensions except ‘feeling sad or worried’. The child-proxy agreement within this dimension was consistently higher across the subgroups when parents were asked to consider the proxy-child perspective.

Across the age groups, a low (moderate) agreement was evident among children aged 6–7 years and their parents for the ‘doing usual activities’ dimension from both proxy perspectives (proxy-proxy = 0.42, proxy-child = 0.48) and

Table 2 Description of EQ-5D-Y-3L (self, proxy-proxy, and proxy-child) preference-weighted health states (overall and based on subgroups)

	EQ-5D-Y-3L (self)	EQ VAS (self)	EQ-5D-Y-3L (proxy-proxy)	EQ VAS (proxy-proxy)	EQ-5D-Y-3L (proxy-child)	EQ VAS (proxy-child)
<i>Overall</i>						
<i>N</i>	85	83	85	85	85	85
Median (IQR)	1 (0.81–1)	90 (75–98)	0.84 (0.8–1)	90 (85–95)	1 (0.8–1)	90 (90–100)
Self vs. proxy difference (<i>p</i> value)			0.32	0.19	1	0.01 ^a
<i>Age group</i>						
6–7 years						
<i>N</i> (%)	30 (0.35)	29 (0.34)	30 (0.35)	30 (0.35)	30 (0.35)	30 (0.35)
Median (IQR)	1 (0.76–1)	89 (74–97)	0.83 (0.8–1)	90 (82–95)	1 (0.8–1)	90 (80–95)
Self vs. proxy difference (<i>p</i> value)			0.78	0.49	0.79	0.48
8–10 years						
<i>N</i> (%)	30 (0.35)	29 (0.34)	30 (0.35)	30 (0.35)	30 (0.35)	30 (0.35)
Median (IQR)	1 (0.76–1)	89 (74–97)	0.83 (0.8–1)	90 (82–95)	1 (0.8–1)	90 (80–95)
Self vs. proxy difference (<i>p</i> value)			0.78	0.49	0.79	0.48
11–12 years						
<i>N</i> (%)	32 (0.38)	32 (0.38)	32 (0.38)	32 (0.38)	32 (0.38)	32 (0.38)
Median (IQR)	1 (0.83–1)	85 (72.5–93)	0.92 (0.8–1)	90 (77.5–95)	1 (0.8–1)	90 (89–96.5)
Self vs. proxy difference (<i>p</i> value)			0.09	0.15	0.43	0.01 ^a
<i>Long-term health condition/s</i>						
No						
<i>N</i> (%)	59 (0.69)	58 (0.68)	59 (0.69)	59 (0.69)	59 (0.69)	59 (0.69)
Median (IQR)	1 (0.83–1)	89.5 (75–98)	1 (0.8–1)	90 (85–95)	1 (0.8–1)	90 (88–100)
Self vs. proxy difference (<i>p</i> value)			0.84	0.29	0.48	0.03 ^a
Yes						
<i>N</i> (%)	26 (0.31)	25 (0.29)	26 (0.31)	26 (0.31)	26 (0.31)	26 (0.31)
Median (IQR)	1 (0.76–1)	90 (75–95)	0.8 (0.73–1)	90 (81–95)	0.8 (0.73–1)	90 (90–100)
Self vs. proxy difference (<i>p</i> value)			0.18	0.47	0.3	0.21

VAS visual analogue scale, IQR interquartile range

^a*p* value significant at $\alpha = 0.05$ for the Wilcoxon matched-pair signed-rank test. EQ-5D-Y-3L, EQ VAS proxy-proxy = Proxy version 1 report; EQ-5D-Y-3L, EQ VAS proxy-child = Proxy version 2 report

‘looking after myself’ (both = 0.54). In comparison, very good agreement was observed within the same dimension in the 11- to 12-year-old age group with both proxy-proxy (0.87) and proxy-child (0.9) reports. However, for this age group (11–12 years), a lower level of (moderate) agreement was noted in the ‘feeling worried, sad or unhappy’ dimension with proxy-proxy (0.52) and for the ‘having pain or discomfort’ dimension with proxy-child (0.59). For the 8- to 10-year-olds, the dimension-level agreement was categorised as either good or very good.

Among children with reported health conditions, a moderate agreement was observed in the dimensions of ‘doing usual activities’ (0.54) and ‘feeling worried, sad or unhappy’

(0.49) with the proxy-proxy report, and in the dimension ‘having pain or discomfort’ (0.56) with the proxy-child report. In contrast, for children without any reported health conditions, agreement levels ranged between good and very good across all dimensions.

3.4 Intra-Proxy Agreement for the EQ-5D-Y-3L Proxy Measures

The EQ-5D-Y-3L proxy-proxy and proxy-child reports yielded similar HRQoL values. The Wilcoxon matched-pair signed-rank test revealed no significant differences in the

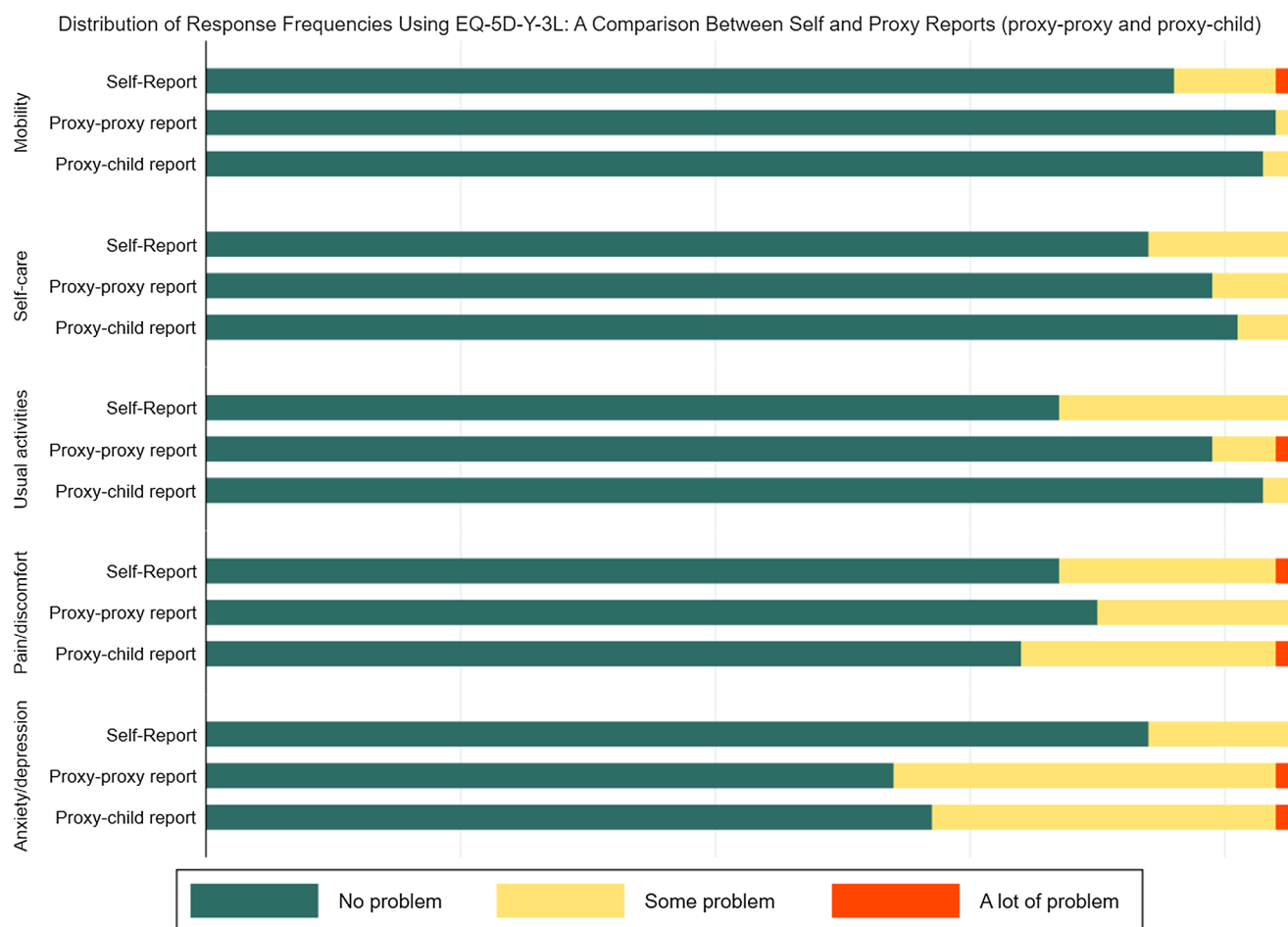


Fig. 1 An overview of the distribution of responses using the EQ-5D-Y-3L self, proxy-proxy and proxy-child reports. (Dimension labels: mobility = walking about, self-care = looking after myself,

usual activities = doing usual activities, pain/discomfort = having pain/discomfort, anxiety/depression = feeling worried, sad or unhappy)

HRQoL values ($p = 0.95$) and across subgroups. However, the EQ VAS scores for the proxy-proxy version were significantly lower than for the proxy-child version ($p = 0.02$).

Figure 2 shows the intra-proxy gap between the two proxy versions of the EQ-5D-Y-3L. A significant but moderate agreement was observed for the HRQoL values between the two proxy versions (CCC 0.53, 95% CI 0.35–0.66). The dimension-level agreement was found to be very good ($AC_1 > 0.9$) for all dimensions except ‘having pain or discomfort’ (0.64) and ‘feeling worried, sad or unhappy’ (0.59) [see ESM Appendix Table S2].

3.5 Sensitivity Analysis

The findings using the German EQ-5D-Y-3L value set indicate similar inter-rater agreements in terms of overall HRQoL. The agreement was 0.29 (0.08–0.47) for the proxy-proxy report and 0.25 (0.04–0.44) for the proxy-child report in the overall sample. The intra-proxy agreement was 0.52

(0.35–0.66) and the HRQoL values were similar to those produced by applying the Australian value set for the EQ-5D-3L (see ESM Appendix Table S3 for more details).

4 Discussion

This paper contributes to the literature by examining the impact of differing proxy perspectives on the inter-rater and intra-rater agreement in the assessment of children’s HRQoL. To our knowledge, this is the first study that compares the two proxy perspectives of the EQ-5D-Y-3L measure on child-proxy agreement in a sample of children from the community aged 6–12 years and their parents. The results of our study are consistent with the findings of our recent systematic review: the inter-rater agreement for HRQoL values was generally low, ranging from poor to fair, from both perspectives (proxy-proxy = 0.28, proxy-child = 0.26). We also found when the proxy-child

Table 3 Distribution of EQ-5D-Y-3L (self, proxy-proxy, and proxy-child) dimension-level responses (overall and based on subgroups)

Dimensions	Response level	EQ-5D-Y-3L (self) [n (%)]	EQ-5D-Y-3L (proxy-proxy) [n (%)]	EQ-5D-Y-3L (proxy-child) [n (%)]	Fisher exact test <i>p</i> value ^a
Overall (N = 85)					
Walking about	No problems	76 (0.89)	84 (0.99)	83 (0.98)	0.02 ^b
	Some problems	8 (0.09)	1 (0.01)	2 (0.02)	
	A lot of problems	1 (0.01)			
Looking after myself	No problems	74 (0.87)	79 (0.93)	81 (0.95)	0.18
	Some problems	11 (0.13)	6 (0.07)	4 (0.05)	
	A lot of problems				
Doing usual activities	No problems	67 (0.79)	79 (0.93)	83 (0.98)	< 0.001 ^b
	Some problems	18 (0.21)	5 (0.06)	2 (0.02)	
	A lot of problems		1 (0.01)		
Having pain/discomfort	No pain/discomfort	67 (0.79)	70 (0.82)	64 (0.75)	0.74
	Some pain/discomfort	17 (0.2)	15 (0.18)	20 (0.24)	
	A lot of pain/discomfort	1 (0.01)		1 (0.01)	
Feeling worried, sad or unhappy	Not worried, sad or unhappy	74 (0.87)	54 (0.64)	57 (0.67)	< 0.001 ^b
	A little worried, sad or unhappy	11 (0.13)	30 (0.35)	27 (0.32)	
	Very worried, sad or unhappy		1 (0.01)	1 (0.01)	
<i>Age group</i>					
6–7 years (n = 23)					
Walking about	No problems	19 (0.83)	23 (1)	23 (1)	0.03 ^b
	Some problems	3 (0.13)			
	A lot of problems	1 (0.04)			
Looking after myself	No problems	17 (0.74)	20 (0.87)	20 (0.87)	0.56
	Some problems	6 (0.26)	3 (0.13)	3 (0.13)	
	A lot of problems				
Doing usual activities	No problems	14 (0.61)	21 (0.91)	22 (0.96)	< 0.01 ^b
	Some problems	9 (0.39)	2 (0.09)	1 (0.04)	
	A lot of problems				
Having pain/discomfort	No pain/discomfort	20 (0.87)	19 (0.83)	17 (0.74)	0.64
	Some pain/discomfort	3 (0.13)	4 (0.17)	6 (0.26)	
	A lot of pain/discomfort				
Feeling worried, sad or unhappy	Not worried, sad or unhappy	21 (0.91)	16 (0.7)	16 (0.7)	0.15
	A little worried, sad or unhappy	2 (0.09)	6 (0.26)	7 (0.3)	
	Very worried, sad or unhappy		1 (0.04)		
8–10 years (n = 30)					
Walking about	No problems	30 (0.94)	31 (0.97)	31 (0.97)	0.32
	Some problems	2 (0.06)	1 (0.03)	1 (0.03)	
	A lot of problems				
Looking after myself	No problems	30 (0.94)	31 (0.97)	31 (0.97)	0.36
	Some problems	2 (0.06)	1 (0.03)	1 (0.03)	
	A lot of problems				
Doing usual activities	No problems	30 (0.94)	30 (0.94)	31 (0.97)	< 0.01 ^b
	Some problems	2 (0.06)	2 (0.06)	1 (0.03)	
	A lot of problems				
Having pain/discomfort	No pain/discomfort	26 (0.81)	28 (0.88)	25 (0.78)	0.97
	Some pain/discomfort	6 (0.19)	4 (0.13)	7 (0.22)	
	A lot of pain/discomfort				

Table 3 (continued)

Dimensions	Response level	EQ-5D-Y-3L (self) [n (%)]	EQ-5D-Y-3L (proxy-proxy) [n (%)]	EQ-5D-Y-3L (proxy-child) [n (%)]	Fisher exact test <i>p</i> value ^a
Feeling worried, sad or unhappy	Not worried, sad or unhappy	28 (0.88)	20 (0.63)	23 (0.72)	0.08
	A little worried, sad or unhappy	4 (0.13)	12 (0.38)	8 (0.25)	
	Very worried, sad or unhappy				
11–12 years (<i>n</i> = 32)					
Walking about	No problems	30 (0.94)	31 (0.97)	31 (0.97)	> 0.99
	Some problems	2 (0.06)	1 (0.03)	1 (0.03)	
	A lot of problems				
Looking after myself	No problems	30 (0.94)	31 (0.97)	31 (0.97)	> 0.99
	Some problems	2 (0.06)	1 (0.03)	1 (0.03)	
	A lot of problems				
Doing usual activities	No problems	30 (0.94)	30 (0.94)	31 (0.97)	> 0.99
	Some problems	2 (0.06)	2 (0.06)	1 (0.03)	
	A lot of problems				
Having pain/discomfort	No pain/discomfort	26 (0.81)	28 (0.88)	25 (0.78)	0.71
	Some pain/discomfort	6 (0.19)	4 (0.13)	7 (0.22)	
	A lot of pain/discomfort				
Feeling worried, sad or unhappy	Not worried, sad or unhappy	28 (0.88)	20 (0.63)	23 (0.72)	0.07
	A little worried, sad or unhappy	4 (0.13)	12 (0.38)	8 (0.25)	
	Very worried, sad or unhappy				
<i>Health condition</i>					
No (<i>n</i> = 59)					
Walking about	No problems	52 (0.88)	59 (1)	58 (0.98)	< 0.01 ^b
	Some problems	6 (0.1)		1 (0.02)	
	A lot of problems	1 (0.02)			
Looking after myself	No problems	52 (0.88)	54 (0.92)	56 (0.95)	0.35
	Some problems	7 (0.12)	5 (0.08)	3 (0.05)	
	A lot of problems				
Doing usual activities	No problems	47 (0.8)	58 (0.98)	58 (0.98)	< 0.001 ^b
	Some problems	12 (0.2)	1 (0.02)	1 (0.02)	
	A lot of problems				
Having pain/discomfort	No pain/discomfort	46 (0.78)	51 (0.86)	48 (0.81)	0.52
	Some pain/discomfort	13 (0.22)	8 (0.14)	11 (0.19)	
	A lot of pain/discomfort				
Feeling worried, sad or unhappy	Not worried, sad or unhappy	54 (0.92)	41 (0.69)	43 (0.73)	< 0.01 ^b
	A little worried, sad or unhappy	5 (0.08)	18 (0.31)	16 (0.27)	
	Very worried, sad or unhappy				
Yes (<i>n</i> = 26)					
Walking about	No problems	24 (0.92)	25 (0.96)	25 (0.96)	> 0.99
	Some problems	2 (0.08)	1 (0.04)	1 (0.04)	
	A lot of problems				
Looking after myself	No problems	22 (0.85)	25 (0.96)	25 (0.96)	0.47
	Some problems	4 (0.15)	1 (0.04)	1 (0.04)	
	A lot of problems				
Doing usual activities	No problems	20 (0.77)	21 (0.81)	25 (0.96)	0.14
	Some problems	6 (0.23)	4 (0.15)	1 (0.04)	
	A lot of problems		1 (0.04)		
Having pain/discomfort	No pain/discomfort	21 (0.81)	19 (0.73)	16 (0.62)	0.38

Table 3 (continued)

Dimensions	Response level	EQ-5D-Y-3L (self) [n (%)]	EQ-5D-Y-3L (proxy-proxy) [n (%)]	EQ-5D-Y-3L (proxy-child) [n (%)]	Fisher exact test <i>p</i> value ^a
Feeling worried, sad or unhappy	Some pain/discomfort	4 (0.15)	7 (0.27)	9 (0.35)	0.19
	A lot of pain/discomfort	1 (0.04)		1 (0.04)	
	Not worried, sad or unhappy	20 (0.77)	13 (0.5)	14 (0.54)	
	A little worried, sad or unhappy	6 (0.23)	12 (0.46)	11 (0.42)	
	Very worried, sad or unhappy		1 (0.04)	1 (0.04)	

HRQoL health-related quality of life

^a*p* value from the omnibus Fisher's exact test for comparison among self-, proxy-proxy-, and proxy-child-reported HRQoL

^b*p* value significant at $\alpha = 0.05$; EQ-5D-Y-3L proxy-proxy = proxy version 1 report; EQ-5D-Y-3L proxy-child = proxy version 2 report

perspective was adopted, the median HRQoL values for the child- and parent-respondent were almost identical, and this result persisted when an alternative value set was used. One plausible explanation for these findings could be that both Australian and German value sets assign a relatively small decrement in the utility in the transition between response levels 1 and 2 [27, 28]. In our study sample, the variations in self and proxy responses were mostly confined to levels 1 and 2, and the minimal disparity in the utility weights between the two levels may have had the effect of 'flattening' those differences. However, the median HRQoL values from the proxy-proxy perspective were lower relative to children's self-perspective. This discrepancy may reflect relevant information and insights from the proxies, which may differ from the child's own self-assessment. Proxies may have a more comprehensive understanding of the child's wellbeing and may often recognise subtle changes in the health state of the child, which the child may not be able to recognise or communicate depending on their age and development stage [14]. Furthermore, discrepancies between child and parent assessments might arise from differences in their inherent understanding of what constitutes various dimensions of HRQoL. These discrepancies may also be influenced by differing contextual considerations, such as their recent health experiences [37]. This may allow for a more nuanced assessment of the child's HRQoL, or alternatively, parents may not be fully aware of all HRQoL aspects that are important to children.

Consistent with earlier studies evaluating dimension-level HRQoL inter-rater agreement using other measures such as HUI 2/3 [38–42], this study found a lower overall agreement for the psychosocial health-related dimension. In this study, parents tended to report more problems in the 'feeling worried, sad, or unhappy' dimension than children themselves. However, a slightly higher level of agreement was observed for this for proxy-child report (proxy-proxy = 0.58 vs. proxy-child = 0.7). This was also evident in the intra-proxy gap,

which was the highest for this dimension. Therefore, when psychosocial wellbeing is a key consideration in the assessment of child HRQoL, the choice of proxy perspective may be crucial.

Another important finding was the differential effect of age group on the inter-rater gap. Previous research found that child-parent agreement decreases with age in a sample of children aged 8–18 years [43]. Our findings suggest a similar trend in agreement for the overall HRQoL values but not in the assessment of dimension-level HRQoL among children aged 11–12 years and their parents. This observed discrepancy between overall and dimension-level agreement was consistent across other child-specific generic measures, namely the CHU9D and the PedsQLTM, when assessed within the same sample in another study [44]. This could be due to the difference in the statistical properties of the method used to estimate the inter-rater agreement (CCC vs. AC_1). The CCC takes into account the high variation in ratings due to each child being rated by a different rater pair, which could potentially result in a lower estimated coefficient [23, 45]. On the other hand, AC_1 relies on the percentage agreement and chance-corrected agreement between the raters [23] and may be more informative due to its disaggregated approach.

The oldest age group also yielded a statistically significant difference in self and proxy reported EQ VAS scores when the proxy-child perspective was adopted. A statistically significant difference in the dyadic EQ VAS scores was reported in the study by Jelsma and Ramma involving school children using the EQ-5D-Y-3L proxy-child report [46]. These findings highlight that children apply a different set of internal standards to evaluate their health than parents who may be unable to replicate a child's self-assessment on the VAS. Research has also indicated that children under the age of 7 years may lack the conceptual ability to use a VAS [47]. However, no significant difference in the self and proxy EQ VAS scores was observed in this age group (6–7 years) in this study.

Table 4 Agreement overall and by subgroup (EQ-5D-Y-3L values and domain-level agreement [proxy-proxy and proxy-child])

	Dimension/HRQoL values	EQ-5D-Y-3L (self and proxy-proxy) AC ₁ (95% CI)	EQ-5D-Y-3L (self and proxy-child) AC ₁ (95% CI)
Overall (N = 85)	Walking about	0.87 (0.8, 0.95)	0.86 (0.78, 0.94)
	Looking after myself	0.78 (0.68, 0.88)	0.81 (0.71, 0.91)
	Doing usual activities	0.69 (0.57, 0.81)	0.74 (0.62, 0.85)
	Having pain/discomfort	0.68 (0.55, 0.8)	0.64 (0.51, 0.77)
	Feeling worried, sad or unhappy	0.58 (0.43, 0.72)	0.7 (0.58, 0.82)
CCC (95% CI)	EQ-5D-Y-3L values	0.28 (0.08, 0.46)	0.26 (0.05, 0.45)
<i>Age group</i>			
6–7 years (n = 23)	Walking about	0.81 (0.61, 1)	0.81 (0.61, 1)
	Looking after myself	0.54 (0.24, 0.83)	0.54 (0.24, 0.83)
	Doing usual activities	0.42 (0.1, 0.73)	0.48 (0.17, 0.78)
	Having pain/discomfort	0.75 (0.52, 0.98)	0.64 (0.37, 0.9)
	Feeling worried, sad or unhappy	0.59 (0.3, 0.87)	0.64 (0.37, 0.9)
CCC (95% CI)	EQ-5D-Y-3L values	0.28 (– 0.13, 0.61)	0.31 (– 0.08, 0.62)
8–10 years (n = 30)	Walking about	0.9 (0.77, 1)	0.86 (0.71, 1)
	Looking after myself	0.82 (0.65, 0.98)	0.9 (0.77, 1)
	Doing usual activities	0.69 (0.48, 0.91)	0.74 (0.54, 0.94)
	Having pain/discomfort	0.67 (0.44, 0.89)	0.7 (0.49, 0.92)
	Feeling worried, sad or unhappy	0.62 (0.39, 0.85)	0.71 (0.5, 0.92)
CCC (95% CI)	EQ-5D-Y-3L values	0.34 (– 0.01, 0.61)	0.36 (0.01, 0.63)
11–12 years (n = 32)	Walking about	0.9 (0.79, 1)	0.9 (0.79, 1)
	Looking after myself	0.9 (0.79, 1)	0.9 (0.79, 1)
	Doing usual activities	0.87 (0.73, 1)	0.9 (0.79, 1)
	Having pain/discomfort	0.64 (0.42, 0.86)	0.59 (0.36, 0.82)
	Feeling worried, sad or unhappy	0.52 (0.28, 0.77)	0.74 (0.54, 0.93)
CCC (95% CI)	EQ-5D-Y-3L values	0.16 (– 0.18, 0.47)	0.05 (– 0.29, 0.37)
<i>Health condition</i>			
No (n = 59)	Walking about	0.87 (0.78, 0.97)	0.86 (0.75, 0.96)
	Looking after myself	0.78 (0.65, 0.9)	0.82 (0.7, 0.93)
	Doing usual activities	0.76 (0.62, 0.89)	0.76 (0.62, 0.89)
	Having pain/discomfort	0.7 (0.56, 0.85)	0.68 (0.53, 0.83)
	Feeling worried, sad or unhappy	0.61 (0.45, 0.78)	0.7 (0.56, 0.85)
CCC (95% CI)	EQ-5D-Y-3L values	0.24 (– 0.01, 0.46)	0.19 (– 0.06, 0.42)
Yes (n = 26)	Walking about	0.88 (0.73, 1)	0.88 (0.73, 1)
	Looking after myself	0.79 (0.6, 0.98)	0.79 (0.6, 0.98)
	Doing usual activities	0.54 (0.26, 0.82)	0.7 (0.47, 0.93)
	Having pain/discomfort	0.62 (0.37, 0.88)	0.56 (0.29, 0.83)
	Feeling worried, sad or unhappy	0.49 (0.22, 0.77)	0.7 (0.47, 0.93)
CCC (95% CI)	EQ-5D-Y-3L values	0.31 (– 0.07, 0.61)	0.31 (– 0.07, 0.61)

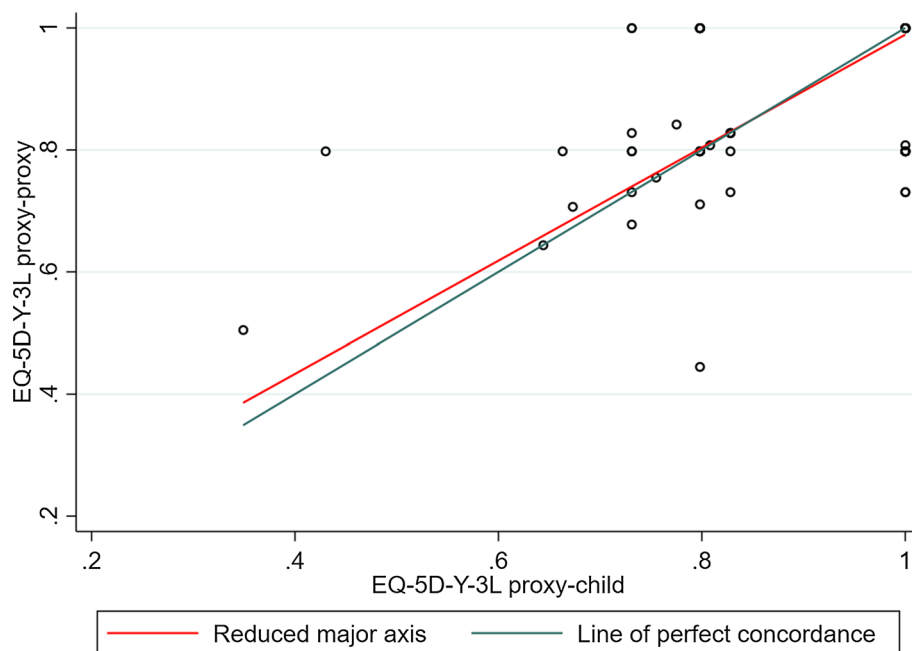
Altman's scale interpretation: ≤ 0.2 = poor, between 0.21 and 0.4 = fair, between 0.41 and 0.6 = moderate, between 0.61 and 0.8 = good, between 0.81 and 1 = very good; EQ-5D-Y-3L proxy-proxy = proxy version 1 report, EQ-5D-Y-3L proxy-child = proxy version 2 report

HRQoL health-related quality of life, AC₁ Gwet's agreement coefficient, CCC Concordance Correlation Coefficient, CI confidence interval

Contrary to the above findings, the youngest group of children (aged 6–7 years) had the lowest inter-rater agreement (moderate) for the observable dimensions of 'doing usual activities' and 'looking after myself'. Larger discrepancies in child-self and proxy reports have been commonly

seen within this age-group relative to cohorts of older children [48]. This has been attributed to either the inability of young children to accurately self-report or differences in the interpretation of the same construct [49]. For example, children may consider themselves too young to dress themselves

Fig. 2 Intra-proxy gap in agreement between the proxy-proxy and proxy-child versions of the EQ-5D-Y-3L. EQ-5D-Y-3L proxy-proxy = proxy version 1 report; EQ-5D-Y-3L proxy-child = proxy version 2 report



or look after themselves, leading to reporting problems in the associated dimensions. Alternatively, parents may simply interpret the construct differently to children. Therefore, additional research to explore how children understand and respond to the HRQoL measure is necessary. Furthermore, adapting the measure to accommodate the developmental stage of younger children (below 8 years of age) may also be needed.

Previous studies have indicated that children with severe health conditions tend to exhibit low levels of agreement with their proxy [50–52]. Interestingly, there is some evidence that children with no apparent health conditions have a lower agreement level than cohorts of children with existing health problems [21, 42]. In this study, significant differences were observed across the three ratings (self, proxy-proxy, and proxy-child) in the dimensions of ‘walking about’, ‘doing usual activities’ and ‘feeling worried, sad or unhappy’ for children with no reported health condition. Moreover, a higher level of inter-rater agreement, as estimated by the CCC, for HRQoL values was observed among children with health condition/s than those without. However, the dimension-level agreement did not exhibit this trend, except for ‘walking about’ wherein the agreement was slightly lower. Given that the children in this community-based sample were generally in good health and did not have any significant health issues, this may indicate that this discrepancy could result from a difference in interpretation of the HRQoL dimension.

Considering that self-reporting HRQoL is preferable, and the presence of a high inter-rater gap in agreement, it is important to assess whether children are meaningfully responding to the self-report measure and whether

differences with parents are based on genuine divergence in perspective. A qualitative investigation using a ‘think-aloud’ approach may provide further evidence to support the validity of the response processes in children of different age groups and to provide further evidence to inform guidance around the minimum age for child self-reporting their own HRQoL using the EQ-5D-Y-3L measure, and also to guide the choice on self-report versus proxy perspective where either are possible [8].

4.1 Limitations

It is important to note that prior studies have utilised weights for adult HRQoL to compute child values due to the absence of country-specific EQ-5D-Y-3L valuation sets [10, 53]. Given that the EQ-5D-Y-3L valuation set for Australia is not yet available and our aim is not to assess the HRQoL of children in this sample, we used Australian EQ-5D-3L weights to calculate self and proxy child values [27]. The same value set was applied to both child and proxy reports. In addition, a German value set specific to the EQ-5D-Y-3L measure was used to check the robustness of the analysis. Nevertheless, an Australian value set for the EQ-5D-Y-3L could weigh the dimensions differently than the adult value set. For instance, value sets for the EQ-5D-Y-3L have been observed to have different orders of dimension importance compared with corresponding adult value sets in the same country [17]. Different dimension-specific preference weights could, in principle, interact with dimension-specific differences in self and proxy ratings of HRQoL. This could either mask or amplify observed differences in inter- and intra-rater agreement by dimension. Further

investigation should be undertaken to determine the validity of the self-report in this sample. Additionally, despite the distraction task of completing the EQ-5D-3L for themselves in between each proxy task, given that the proxy-child report was completed subsequent to the proxy-proxy report, we are unable to rule out the possibility of an ordering effect, and proxy respondents may have potentially revisited their initial response to enforce consistency. Future research could mitigate this potential source of bias by introducing a longer time gap [54] or consider randomising the order in which the two proxy reports are administered.

The study did not capture whether the parent in the child-parent dyad was the primary caregiver. Additionally, the under-representation of fathers in the sample was insufficient for conducting a subgroup analysis. Children in this study were representative of the general community and hence tended to be in relatively good health overall. Furthermore, use of the EQ-5D-Y-3L version may have limited discriminative power in this population, potentially exaggerating agreement between proxy and self-reports. It is important to undertake further studies in clinical paediatric samples comprising children and parents with varying levels of overall health and regular engagement with health services to examine the impact of proxy perspective on the level of agreement across the range of levels of HRQoL dimensions comprising the EQ-5D-Y-3L measure.

5 Conclusions

Overall, for the preference-weighted HRQoL as measured by the EQ-5D-Y-3L, the child-proxy level of agreement was similar but low (poor to fair) regardless of the perspective adopted, especially with older children (11–12 years). This result did not appear to be sensitive to the choice of value set. Across the dimensions, the inter-rater agreement was similar from both perspectives, except for ‘feeling worried, sad, or unhappy’, where the proxy-child report showed higher concordance with child-self reports. The impact of the perspective adopted for measuring HRQoL, and the child-proxy agreement, is an important area for further research including qualitative investigation to better inform longitudinal assessments of child population health and for cost-effectiveness estimations and decision making regarding paediatric populations based on that evidence.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40273-024-01356-0>.

Acknowledgements The authors would like to thank Matthew Crocker from Deakin University, who assisted in conducting the interviews with the participants. The authors would also like to thank all the parents and children who generously gave up their time to participate in this study.

Declarations

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. DK is supported by a Ph.D. scholarship awarded from a project funded by the Department of Health, Australian Government, National Health and Medical Research Council (grant number MRF1200816). Project title: ‘Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA)’. The project aims to improve how quality of life is measured and valued within paediatric populations.

Conflict of interest Kim Dalziel and Nancy Devlin have received previous or current funding from the EuroQol Research Foundation, who is the developer of the measure included in this study. Nancy Devlin is a member of the EuroQol Group. The views expressed in this paper are those of the authors and are not necessarily those of the EuroQol Research Foundation. Diana Khanna, Jyoti Khadka, Christine Mpundu-Kaambwa, Gang Chen and Julie Ratcliffe declare no conflicts of interest.

Ethics approval This study was approved by the Flinders University’s Human Research Ethics Committee (project ID 4178).

Data availability statement The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Code availability Code is available upon request to the authors.

Author contributions JR conceptualised this study. All authors contributed to the study design and data analysis. DK wrote the first draft. All authors provided feedback on the first draft and agreed on the final draft. All authors reviewed and approved the final amendments. JR acts as guarantor.

Disclosure This article is published in a journal supplement wholly funded by the Australian Government MRFF funded QUOKKA research grant 1200816, the University of Melbourne, the EuroQol Research Foundation, the University of Technology Sydney and Flinders University.

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. Connolly MA, Johnson JA. Measuring quality of life in paediatric patients. *Pharmacoeconomics*. 1999;16(6):605–25. <https://doi.org/10.2165/00019053-199916060-00002>.

2. 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>.
3. Chen G, Ratcliffe J. A review of the development and application of generic multi-attribute utility instruments for paediatric populations. *Pharmacoeconomics*. 2015;33(10):1013–28. <https://doi.org/10.1007/s40273-015-0286-7>.
4. EuroQol Research Foundation. EQ-5D-Y User Guide. Rotterdam: QuroQol Research Foundation; 2020.
5. Amien R, Scott D, Verstraete J. Performance of the EQ-5D-Y interviewer administered version in young children. *Children*. 2022;9(1):93.
6. Zhou W, Shen A, Yang Z, Wang P, Wu B, Herdman M, et al. Patient-caregiver agreement and test-retest reliability of the EQ-5D-Y-3L and EQ-5D-Y-5L in paediatric patients with haematological malignancies. *Eur J Health Econ*. 2021;22(7):1103–13. <https://doi.org/10.1007/s10198-021-01309-w>.
7. Shirowai T, Fukuda T, Shimozuma K. Psychometric properties of the Japanese version of the EQ-5D-Y by self-report and proxy-report: reliability and construct validity. *Qual Life Res*. 2019;28(11):3093–105. <https://doi.org/10.1007/s11136-019-02238-1>.
8. 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(1):173–83. <https://doi.org/10.1007/s11136-012-0119-5>.
9. Gusi N, Perez-Sousa MA, Gozalo-Delgado M, Olivares PR. Validity and reliability of the spanish EQ-5D-Y proxy version [in Spanish]. *An Pediatr (Barc)*. 2014;81(4):212–9. <https://doi.org/10.1016/j.anpedi.2013.11.028>.
10. Bray N, Noyes J, Harris N, Edwards RT. Measuring the health-related quality of life of children with impaired mobility: examining correlation and agreement between children and parent proxies. *BMC Res Notes*. 2017;10(1):377. <https://doi.org/10.1186/s13104-017-2683-9>.
11. Matza LS, Patrick DL, Riley AW, Alexander JJ, Rajmil L, Pleil AM, et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. *Value Health*. 2013;16(4):461–79. <https://doi.org/10.1016/j.jval.2013.04.004>.
12. Bevans KB, Ahuvia IL, Hallock TM, Mendonca R, Roth S, Forrest CB, et al. Investigating child self-report capacity: a systematic review and utility analysis. *Qual Life Res*. 2020;29(5):1147–58. <https://doi.org/10.1007/s11136-019-02387-3>.
13. White-Koning M, Arnaud C, Dickinson HO, Thyen U, Beckung E, Fauconnier J, et al. Determinants of child-parent agreement in quality-of-life reports: a European study of children with cerebral palsy. *Pediatrics*. 2007;120(4):e804–14. <https://doi.org/10.1542/peds.2006-3272>.
14. Varni JW, Limbers CA, Burwinkle TM. Parent proxy-report of their children's health-related quality of life: an analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL 4.0 Generic Core Scales. *Health Qual Life Outcomes*. 2007;5:2. <https://doi.org/10.1186/1477-7525-5-2>.
15. Pickard AS, Knight SJ. Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. *Med Care*. 2005;43(5):493–9. <https://doi.org/10.1097/01.mlr.0000160419.27642.a8>.
16. Dolan P, Olsen JA, Menzel P, Richardson J. An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Econ*. 2003;12(7):545–51. <https://doi.org/10.1002/hec.760>.
17. Kreimeier S, Oppe M, Ramos-Goñi JM, Cole A, Devlin N, Herdman M, 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(11):1291–8. <https://doi.org/10.1016/j.jval.2018.05.002>.
18. Wyrwich KW, Metz SM, Kroenke K, Tierney WM, Babu AN, Wolinsky FD. Triangulating patient and clinician perspectives on clinically important differences in health-related quality of life among patients with heart disease. *Health Serv Res*. 2007;42(6 Pt 1):2257–74. <https://doi.org/10.1111/j.1475-6773.2007.00733.x>. (discussion 2294–323).
19. McPhail S, Beller E, Haines T. Two perspectives of proxy reporting of health-related quality of life using the Euroqol-5D, an Investigation of Agreement. *Med Care*. 2008;46(11):1140–8. <https://doi.org/10.1097/MLR.0b013e31817d69a6>.
20. Tol M, Kuipers JP, Willigenburg NW, Willems HC, Poolman RW. How are you doing in the eyes of your spouse? Level of agreement between the self-completed EQ-5D-5L and two proxy perspectives in an orthopaedic population: a randomized agreement study. *Health Qual Life Outcomes*. 2021;19(1):35. <https://doi.org/10.1186/s12955-021-01679-y>.
21. Khanna D, Khadka J, Mpundu-Kaambwa C, Lay K, Russo R, Ratcliffe J. Are we agreed? Self-versus proxy-reporting of paediatric health-related quality of life (HRQoL) using generic preference-based measures: a systematic review and meta-analysis. *Pharmacoeconomics*. 2022;40(11):1043–67. <https://doi.org/10.1007/s40273-022-01177-z>.
22. Australian Institute of Health and Welfare. Health of children. 2020. <https://www.aihw.gov.au/reports/australias-health/health-of-children>. Accessed 10 Aug 2023.
23. Gwet KL. Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters. Advanced Analytics, LLC; 2014.
24. Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav*. 1997;38(1):21–37.
25. Moom T. Mode of administration and interviewer effects in self-reported symptoms of anxiety and depression. *Soc Indic Res*. 1998;45(1):279–318. <https://doi.org/10.1023/A:1006958100504>.
26. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burström K, Cavrini G, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res*. 2010;19(6):887–97. <https://doi.org/10.1007/s11136-010-9649-x>.
27. Viney R, Norman R, King MT, Cronin P, Street DJ, Knox S, et al. Time trade-off derived EQ-5D weights for Australia. *Value Health*. 2011;14(6):928–36. <https://doi.org/10.1016/j.jval.2011.04.009>.
28. Kreimeier S, Mott D, Ludwig K, Greiner W, Prevolnik Rupel V, Ramos-Goñi JM, et al. EQ-5D-Y value set for Germany. *Pharmacoeconomics*. 2022;40(Suppl 2):217–29. <https://doi.org/10.1007/s40273-022-01143-9>.
29. Devlin N, Roudijk B, Viney R, Stolk E. EQ-5D-Y-3L value sets, valuation methods and conceptual questions. *Pharmacoeconomics*. 2022;40(2):123–7. <https://doi.org/10.1007/s40273-022-01226-7>.
30. Australian Bureau of Statistics. Socio-Economic Indexes for Areas (SEIFA). 2033.0.55.001. 2016. [https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/756EE3DBEFA869EFCA258259000BA746/\\$File/SEIFA%202016%20Technical%20Paper.pdf](https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/756EE3DBEFA869EFCA258259000BA746/$File/SEIFA%202016%20Technical%20Paper.pdf). Accessed 10 Aug 2023.
31. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45(1):255–68. <https://doi.org/10.2307/2532051>.
32. Lin L, Hedayat AS, Sinha B, Yang M. Statistical methods in assessing agreement: models, issues, and tools. *J Am Stat Assoc*. 2002;97(457):257–70.
33. Chen C-C, Barnhart HX. Comparison of ICC and CCC for assessing agreement for data without and with replications. *Comput*

- Stat Data Anal. 2008;53(2):554–64. <https://doi.org/10.1016/j.csda.2008.09.026>.
34. Klein D. Implementing a general framework for assessing inter-rater agreement in Stata. *Stata J.* 2018;18(4):871–901.
 35. StataCorp. *Stata Statistical Software: Release 16.* College Station: StataCorp LLC; 2019.
 36. Australian Bureau of Statistics. 2011.0—census of population and housing: reflecting Australia: stories from the Census, 2016. Canberra: Australian Bureau of Statistics; 2018.
 37. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and valuing health benefits for economic evaluation.* Oxford: Oxford University Press; 2016.
 38. Barr RD, Simpson T, Whitton A, Rush B, Furlong W, Feeny DH. Health-related quality of life in survivors of tumours of the central nervous system in childhood: a preference-based approach to measurement in a cross-sectional study. *Eur J Cancer.* 1999;35(2):248–55. [https://doi.org/10.1016/s0959-8049\(98\)00366-9](https://doi.org/10.1016/s0959-8049(98)00366-9).
 39. Glaser A, Kennedy C, Punt J, Walker D. Standardized quantitative assessment of brain tumor survivors treated within clinical trials in childhood. *Int J Cancer Suppl.* 1999;12:77–82. [https://doi.org/10.1002/\(sici\)1097-0215\(1999\)83:12+%3c77::aid-ijc14%3e3.0.co;2-x](https://doi.org/10.1002/(sici)1097-0215(1999)83:12+%3c77::aid-ijc14%3e3.0.co;2-x).
 40. Morrow AM, Hayen A, Quine S, Scheinberg A, Craig JC. 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(2):186–95. <https://doi.org/10.1111/j.1365-2214.2011.01240.x>.
 41. Verrips GH, Stuifbergen MC, den Ouden AL, Bonsel GJ, Gemke RJ, Paneth N, et al. Measuring health status using the Health Utilities Index: agreement between raters and between modalities of administration. *J Clin Epidemiol.* 2001;54(5):475–81. [https://doi.org/10.1016/s0895-4356\(00\)00317-6](https://doi.org/10.1016/s0895-4356(00)00317-6).
 42. Fluchel M, Horsman JR, Furlong W, Castillo L, Alfonz Y, Barr RD. Self and proxy-reported health status and health-related quality of life in survivors of childhood cancer in Uruguay. *Pediatr Blood Cancer.* 2008;50(4):838–43. <https://doi.org/10.1002/pbc.21299>.
 43. Rajmil L, López AR, López-Aguilá S, Alonso J. Parent-child agreement on health-related quality of life (HRQOL): a longitudinal study. *Health Qual Life Outcomes.* 2013;11:101. <https://doi.org/10.1186/1477-7525-11-101>.
 44. Khanna D, Khadka J, Mpundu-Kaambwa C, Ratcliffe J, in Collaboration with the Quality of Life in Kids: Key Evidence to Strengthen Decisions in 6 Australia Project T. Child-parent agreement in the assessment of health-related quality of life using the CHU9D and the PedsQLTM. *Appl Health Econ Health Policy.* 2023;21(6):937–47. <https://doi.org/10.1007/s40258-023-00831-7>.
 45. Lin J, Wong CKH, Cheung PWH, Luo N, Cheung JPY. Feasibility of proxy-reported EQ-5D-3L-Y and its agreement in self-reported EQ-5D-3L-Y for patients with adolescent idiopathic scoliosis. *Spine (Phila Pa 1976).* 2020;45(13):E799–807. <https://doi.org/10.1097/brs.00000000000003431>.
 46. Jelsma J, Ramma L. How do children at special schools and their parents perceive their HRQoL compared to children at open schools? *Health Qual Life Outcomes.* 2010;8:72–72. <https://doi.org/10.1186/1477-7525-8-72>.
 47. Shields BJ, Palermo TM, Powers JD, Grewe SD, Smith GA. Predictors of a child's ability to use a visual analogue scale. *Child Care Health Dev.* 2003;29(4):281–90. <https://doi.org/10.1046/j.1365-2214.2003.00343.x>.
 48. Conijn JM, Smits N, Hartman EE. Determining at what age children provide sound self-reports: an illustration of the validity-index approach. *Assessment.* 2019;27(7):1604–18. <https://doi.org/10.1177/1073191119832655>.
 49. Woolley ME, Bowen GL, Bowen NK. Cognitive pretesting and the developmental validity of child self-report instruments: theory and applications. *Res Soc Work Pract.* 2004;14(3):191–200. <https://doi.org/10.1177/1049731503257882>.
 50. Qadeer RA, Ferro MA. Child-parent agreement on health-related quality of life in children with newly diagnosed chronic health conditions: a longitudinal study. *Int J Adolesc Youth.* 2018;23(1):99–108. <https://doi.org/10.1080/02673843.2017.1297242>.
 51. Petsios K, Priftis KN, Tsoumakas C, Hatziaorou E, Tsanakas JN, Galanis P, et al. Level of parent-asthmatic child agreement on health-related quality of life. *J Asthma.* 2011;48(3):286–97. <https://doi.org/10.3109/02770903.2011.555031>.
 52. Fitriana TS, Purba FD, Stolk E, Busschbach JJV. EQ-5D-Y-3L and EQ-5D-Y-5L proxy report: psychometric performance and agreement with self-report. *Health Qual Life Outcomes.* 2022;20(1):88. <https://doi.org/10.1186/s12955-022-01996-w>.
 53. Kirkham FJ, Vigevano F, Raspall-Chaure M, Wilken B, Lee D, Le Reun C, et al. Health-related quality of life and the burden of prolonged seizures in noninstitutionalized children with epilepsy. *Epilepsy Behav.* 2020;102: 106340. <https://doi.org/10.1016/j.yebeh.2019.04.058>.
 54. Vilagut G. Test-retest reliability. In: Michalos AC, editor. *Encyclopedia of quality of life and well-being research.* Dordrecht: Springer; 2014. p. 6622–5.

Authors and Affiliations

Diana Khanna¹  · Jyoti Khadka^{1,2}  · Christine Mpundu-Kaambwa¹  · Gang Chen⁵  · Kim Dalziel^{3,4}  · Nancy Devlin³  · Julie Ratcliffe¹  · in collaboration with the Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team

✉ Diana Khanna
khan0420@flinders.edu.au

¹ Health and Social Care Economics Group, Caring Futures Institute, College of Nursing and Health Sciences, Flinders University, Adelaide, SA 5042, Australia

² Registry of Senior Australians, Healthy Ageing Research Consortium, South Australian Health and Medical Research Institute, Adelaide, SA, Australia

³ Health Economics Unit, University of Melbourne, Melbourne, VIC, Australia

⁴ Health Services and Economics, Murdoch Children's Research Institute, Parkville, VIC, Australia

⁵ Centre for Health Economics, Monash University, Caulfield East, VIC, Australia