Investigating agreement in Proxy and Self-Reported Health-Related Quality of Life in children for informing Economic Evaluation and Quality assessment

By

Diana Khanna (Masters in HealthEcon)
Supervisors:
Professor Julie Ratcliffe
Associate Professor Jyoti Khadka
Dr Christine Mpundu-Kaambwa

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy
College of Nursing and Health Sciences
31st January 2025

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Background: The application of child-specific preference-based measures to assess Health-Related Quality of Life (HRQoL) enables the derivation of utilities (preference weights) for incorporating into cost utility analysis of health and social care technologies or quality assessments targeted for paediatric populations. Challenges in paediatric self-assessment of HRQoL due to developmental and cognitive constraints necessitate proxy-reports alongside or in place of child self-reports. This thesis examines the inter-rater agreement between child-self and proxy assessments of child HRQoL using generic preference-based measures. The objective is to contribute to the existing evidence by focusing on reported measures of agreement in child and proxy assessments, particularly in highlighting dimension level discrepancies in addition to overall HRQoL.

Methods: The research methodology commenced with a systematic review, examining the existing evidence of agreement between child-self and proxy-reported HRQoL across both overall and at the dimension level using generic preference-based measures, the EQ-5D-Y-3L and the CHU9D. Subsequent empirical investigations involved two distinct studies. The first study was conducted through face-to-face interviews with 85 child-parent dyads, encompassing children aged 6-12 years. This study utilised a mixed-methods approach, combining quantitative and qualitative analyses to gain a comprehensive understanding of the HRQoL assessments. The second study was a larger quantitative analysis involving 845 online child-proxy dyads, where participants completed the EQ-5D-Y-3L for assessing child HRQoL.

Results: The systematic review revealed a generally low inter-rater agreement between child-self and proxy ratings of HRQoL, particularly in dimensions with less observable attributes such as emotional and social well-being. The first empirical study confirmed these findings, showing a consistent but low level of overall agreement across child-proxy assessments, with the lowest concordance observed in psychosocial health-related dimensions. However, the study also noted that adopting a proxy-child perspective resulted in higher agreement for the “feeling worried, sad, or unhappy” dimension in the EQ-5D-Y-3L. Additionally, with the CHU9D, children aged 6-7 years had comparable dimension level agreement to older age groups (8-10 and 11-12 years), whereas the PedsQL™ showed lower agreement for older age groups in similar psychosocial health dimensions. The mixed-methods approach further highlighted that the majority of younger children, especially those aged 6-10 years, tended to face significant challenges in reporting HRQoL dimensions as intended, leading to disparities with parental proxy reports. These discrepancies were not always reflective of actual health impairments but rather indicative of understanding and interpretation challenges. The second study’s findings, based on a larger sample, suggested that the Australian adult value set for the EQ-5D-Y-3L may provide agreement levels comparable to child-specific value sets in assessing HRQoL.

Conclusions: This thesis is one of the first to comprehensively examine the evidence that relates to the level of inter-rater agreement between child-self and proxy-reported child HRQoL, using generic preference-based measures. The findings significantly contribute to the methodological considerations in measuring child HRQoL, highlighting the complexities in using proxy reports as substitutes, particularly for subjective HRQoL, and in the intended interpretations of the dimensions by younger child-self reporters. The research underscores the need for enhancing current age-appropriate HRQoL measures and calls for ongoing research to improve the validity of both self and proxy reports. This research is vital for health economists, policymakers, and practitioners in paediatric health technology assessment, aiming to ensure that health technologies are evaluated accurately, reflecting the true needs and HRQoL experiences of the paediatric population.

LIST OF PUBLICATIONS AND ABSTRACTS ARISING FROM THIS THESIS

Peer-reviewed Publications (N=3)

Khanna, D., Khadka, J., Mpundu-Kaambwa, C., Lay, K., Russo, R., Ratcliffe, J., & Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team (2022). 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, 40(11), 1043–1067.

Khanna, D., Khadka, J., Mpundu-Kaambwa, C., Chen, G., Dalziel, K., Devlin, N., Ratcliffe, J., & Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team (2024). 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. PharmacoEconomics, 42(Suppl 1), 113–128.

Khanna, D., Khadka, J., Mpundu-Kaambwa, C., Ratcliffe, J., & Quality of Life in Kids: Key Evidence to Strengthen Decisions in 6 Australia (QUOKKA) Project Team (2023). Child-Parent Agreement in the Assessment of Health-Related Quality of Life Using the CHU9D and the PedsQLTM. Applied health economics and health policy, 21(6), 937–947.

Conference Presentations (N=5)

*Denotes presenting author

Khanna, D.*, Khadka, J., Mpundu-Kaambwa, C., Lay, K., Russo, R., Ratcliffe, J. 2022, ‘Are we agreed? Self-versus proxy-reporting of paediatric Health-related Quality of Life (HRQoL) using EuroQol measures: A systematic review’, 2nd Early Career Researcher EuroQol meeting, 20th April, Noordwijk, The Netherlands.

Khanna, D.*, Khadka, J., Mpundu-Kaambwa, C., Ratcliffe, J. 2022, ‘Do they agree? Self and proxy reports of health-related quality of life (HRQoL) among children in the general population using EQ-5D-Y-3L’, 39th Euroqol Plenary, 20-23 September, Chicago, United States.

Khanna, D.*, Khadka, J., Mpundu-Kaambwa, C., Ratcliffe, J. 2022, ‘Do they agree? Self and proxy reports of health-related quality of life among children in the general population via deliberation approach’, 29th ISOQOL Annual Conference, 19-22 October, Prague, Czech Republic.

Khanna, D.*, Khadka, J., Lay, K., Mpundu-Kaambwa, C., Ratcliffe, J. 2023, ‘Understanding Children's Perception of Health-Related Quality of Life (HRQoL): A Mixed-Methods Investigation of the EQ-5D-Y-3L in Community Settings’, 30th ISOQOL Annual Conference, 18-21 October, Alberta, Canada.

Khanna, D.*, Khadka, J., Lay, K., Mpundu-Kaambwa, C., Ratcliffe, J. 2023, ‘Understanding Children's Perception of Health-Related Quality of Life (HRQoL): A Mixed-Methods Approach’, 44th AHES Annual Conference, 21-22 September, Adelaide, Australia.

Conference Poster (Accepted)

Khanna, D.*, Khadka, J., Mpundu-Kaambwa, C., Ratcliffe, J. 2024, ‘Evaluating Health-Related Quality of Life in Children: Selecting the value-set for the EQ-5D-Y-3L’, 31st ISOQOL Annual Conference, 13-16 October, Cologne, Germany.

DECLARATION

I certify that this thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university

2. and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University; and

3. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

DIANA KHANNA

January 2025

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest gratitude to my supervisors. A heartfelt thank you to Professor Julie Ratclife for providing me with this opportunity to work on this meaningful and interesting project. I am immensely grateful for her guidance, generosity and support, both professionally and personally, which have been instrumental in achieving my milestones. I am also deeply thankful to Associate Professor Dr. Jyoti Khadka for his continuous guidance and insights that have significantly benefited my research and personal growth. My sincere appreciation goes to Dr. Christine Mpundu-Kaambwa for sharing her expertise and offering valuable advice. And importantly, I extend my heartfelt thanks to all the participants, the children, and their parents, for taking the time to engage with the research and making this thesis possible. Their keen involvement and cooperation have been crucial to this research.

A special thank you to my beloved Juan for his unwavering support through every challenge and triumph of my candidature. I am also profoundly grateful to the incredible Health Economics team, Associate Professor Rachel Milte, Kiri Lay, and Jenny Cleland. Their expertise, support and encouragement have been an invaluable source of motivation for me during this journey. I am fortunate to have a PhD family that made this journey memorable and enjoyable. Tracy, Lynne, Paul, Phoebe, Nurul, Jess, Yumi, Dhanushi and Azfira, without them, I might have finished my PhD earlier, but it would not have been as memorable. Their friendship and belief in me have been a source of strength and comfort. I also want to thank my dear friends Ruchita, Archana, Sunidhi, Jasmine and Vijaya for their love, patience and understanding over the years.

Finally, I extend my heartfelt gratitude to my family, Ignatius, Mary, Mangala, Anniruddha, Rudrakshi, Melvin, and Shiro. Despite the distance, their unwavering love has been a source of strength throughout my journey.

To each of them, I owe a debt of gratitude for your support, guidance, and encouragement throughout my PhD journey. This accomplishment would not have been possible without them.

ABBREVIATIONS

16D	Sixteen-dimensional Measure of HRQoL
17D	Seventeen-dimensional Measure of HRQoL
AC	Agreement Coefficient
AHUM	Adolescent Health Utility Measure
AQoL-6D	Adolescent Quality of Life-6D
AUD	Australian dollars
CBA	Cost Benefit Analysis
CCA	Cost Consequence Analysis
CCC	Concordance Correlation Coefficient
CEA	Cost Effectiveness Analysis
CHU9D	Child Health Utility 9D measure
CI	Confidence Intervals
CMA	Cost Minimisation Analysis
CUA	Cost Utility Analysis
DALY	Disability Adjusted Life Year
DCE	Discrete Choice Experiment
HRQoL	Health-Related Quality of Life
HTA	Health Technology Assessment

HUI	Health Utilities Index
ICC	Interclass Correlation Coefficient
ICER	Incremental Cost Effectiveness Ratio
IQR	Inter-Quartile Range
NICE	National Institute for Health and Care Excellence
PBAC	Pharmaceutical Benefits Advisory Committee
PBM	Preference-Based Measure
PBS	Pharmaceutical Benefits Scheme
PROM	Patient-Reported Outcome Measure
QALY	Quality Adjusted Life Year
QoL	Quality of Life
QWB	Quality of Well-Being Scale
SD	Standard Deviation
SEIFA	Socio-Economic Indexes for Australia
SRH	Self-Rated Health
VAS	Visual Analogue Scale

CHAPTER 1: INTRODUCTION

1.1 Overview

This chapter introduces the research topic, provides the rationale, and outlines the structure of the thesis. It sets the stage for the subsequent chapters by highlighting the importance of measuring health-related quality of life (HRQoL) in children and the challenges associated with it.

1.2 Economic perspectives

The measurement and valuation of child health is a key component of economic evaluation for interventions targeted at paediatric populations (from birth to 17 years of age) [1]. Regulatory bodies around the world, such as the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia and the National Institute for Health and Clinical Excellence (NICE) in the UK, require evidence from economic evaluations of child health interventions (including new health care technologies, medical devices and pharmaceuticals) to assist in making reimbursement decisions and informing health policy [2, 3]. In addition to evidence of clinical effectiveness, the adoption of child health interventions needs to be justified by value for money considerations, necessitating systematic comparisons that clearly identify, analyse, and evaluate the competing interventions in terms of their relative costs and benefits [4]. Economic evaluations, defined by Drummond et al. (2015) as the “comparative analysis of alternative courses of action in terms of both their costs and consequences”, play a pivotal role in providing a framework to inform decisions concerning health or health-related welfare for both adult and paediatric populations [4]. The PBAC, for example, considers evidence derived from economic evaluations when recommending medicines eligible for government subsidies under the Pharmaceutical Benefits Scheme (PBS) [5].

The economic evaluation process involves the identification, measurement, valuation, and comparison of the costs and consequences of competing alternatives [2, 4]. The selection of a perspective in economic evaluation significantly influences the costs and consequences that are considered and depends on the type of decision-maker for whom the evaluation is performed (e.g., payer, healthcare system or society) [4, 6, 7]. In Australia, economic evaluations typically adopt the perspective of the healthcare system [5], whereas in countries like the Netherlands a broader societal perspective is often recommended [3]. The healthcare system perspective focuses on maximising health outcomes within a healthcare budget. This perspective, therefore, captures only the resource use and costs borne by the healthcare system, excluding non-healthcare-related costs such as travel expenses or productivity loss incurred by patients that may fall outside the scope of the healthcare perspective [8, 9]. Direct costs in this context include medical costs (e.g., costs of vaccines, medications, hospital stays) and non-medical costs that are directly related to healthcare delivery (e.g., administrative costs within the healthcare system).

A broader societal perspective, on the other hand, recognises that health interventions can shift costs between sectors and considers the forgone cost of allocating resources to healthcare over other societal needs [4, 6]. It includes all healthcare-related costs (including costs to access intervention such as travel and accommodation) and may also include relevant non-healthcare-related indirect costs such as those related to caregiving or productivity changes [8, 10]. For instance, in the case of a childhood vaccination program, the healthcare system perspective would include direct healthcare costs including the cost of vaccines and their administration, but a societal perspective would also account for wider cross-sectorial and longer term benefits, e.g., education gains and increased productivity in adult life [11]. Societal perspective, potentially, accounts for the cost-shift between sectors but to mitigate omitted variable bias, it is crucial that the costs and consequences considered are comprehensive, and that the cost of collecting the information is less than the value of the information obtained [12]. However, the guidelines that outline which costs and consequences are relevant with respect to societal perspective vary by the respective authoritative health technology assessment (HTA) agencies [10], making consistent implementation challenging.

Depending on the chosen perspective, identified costs, both direct (medical and non-medical costs) and indirect (time-related or productivity costs) are measured and valued in monetary terms (Australian dollars in this context). The unit of assessment of outcomes, however, varies, leading to different techniques in economic evaluation (e.g., Quality-Adjusted Life Years (QALYs)). There are five commonly employed techniques used in economic evaluations (outlined in Table 1.1) [2, 4].

Table 1-1 Economic evaluation techniques.

Type of Evaluation	Outcome Measurement	Outcome variable
Cost-Minimisation	Assumed equivalent	N/A
Cost-Benefit	Any benefits associated with the alternatives	Australian Dollars ($AUD)
Cost-Consequence	Multiple relevant outcomes associated with the intervention	Units of the consequences (e.g., Carer burden scores, mortality, etc.)
Cost-Effectiveness	Outcome common to alternatives	Natural units (e.g., life-years gained, infection rates reduced)
Cost-Utility	Life years gained and the quality of those years	QALYs, DALYs, healthy years equivalent

QALY= Quality adjusted life years, DALY= Disability adjusted life years

1.3 Economic evaluation techniques

Cost minimisation analysis (CMA) compares the costs of interventions presumed to have equivalent (producing identical health outcomes) or non-inferior outcomes (producing outcomes that are not significantly worse than comparator in terms of effectiveness, within a clinically acceptable margin). This allows the comparison of the relative costs of the alternatives to identify the least costly option. For this analysis to be valid, there must be robust evidence to support that the alternatives considered are equally safe and effective [2, 4, 13]. Cost benefit analysis (CBA) assesses both costs and outcomes in monetary terms (e.g., willingness-to-pay), facilitating comparisons between programs where the net benefits exceed the net costs [2, 4, 13]. As outcomes are assessed in monetary terms CBAs facilitate the inclusion of wider non-health benefits in cost effectiveness calculations thus appealing from a taxpayer's perspective [14]. However, when operating under budget constraints with the goal to maximise health outcomes according to societal preferences, the following techniques may be used: Cost consequence analysis, Cost-effectiveness analysis (CEA) or Cost-utility analysis (CUA) [2, 4, 15].

Cost consequence analysis typically evaluates multiple relevant health outcomes such as patient satisfaction and access to service, rather than focusing solely on a single outcome [2, 13, 15]. It presents all the relevant costs and consequences in a tabular format, without prioritising them by importance or synthesising costs and outcomes to produce cost effectiveness ratios. This allows decision-makers to assign their own values to these metrics, which may differ according to local contexts [15]. CEA and CUA, on the other hand, typically assess a single outcome, of effectiveness or utility, of interventions relative to their costs.

CEA measures outcomes in their natural units, which makes it particularly suitable when a specific clinical endpoint is the focus, rather than a broader health outcome. This approach is advantageous when the clinical endpoint, such as reduction in blood pressure or decrease in mortality rates, is clear and quantifiable [2, 4, 13]. It is also preferred in jurisdictions where QALYs, a generic outcome measure, may not be appropriate or recommended (e.g., Germany) [16, 17]. Outcomes in CEA are typically quantified using clinical outcomes such as infection rates reduced, or units of blood pressure reduced, and the results are presented as cost per unit of health effect. This facilitates the calculation of the Incremental Cost-Effectiveness Ratio (ICER), given by a ratio of the difference in total costs (incremental costs) and difference in total benefits (incremental effect) where benefits are measured using the main clinical outcome of interest [2, 4, 13, 18]. However, a limitation of CEA is its inability to effectively compare interventions that produce different health outcomes, thus often requiring a more generic measure of outcome to provide a comprehensive evaluation.

A variant of CEA, CUA quantifies benefits using a measure of outcome designed to be applicable across different health conditions, which facilitates the evaluation of trade-offs between the length and quality of life, which is critical in the assessment of healthcare interventions [2, 4, 13]. This technique is widely preferred by the HTA agencies [3] and predominantly employs QALYs, a composite measure of outcome that combines both gains in the quantity and quality of life [2-4, 13, 19]. This allows for comprehensive comparisons across interventions that affect various health dimensions e.g. physical health and/or mental health related dimensions. Results are typically reported as incremental cost per QALY gained [2, 4]. The integration of diverse health outcomes into a unified measure, the QALY, will be discussed in more detail in subsequent sections.

1.4 Measuring health outcomes

As highlighted previously, the QALY metric integrates two primary elements: the quantity of life, generally measured as life expectancy or life-years gained (i.e., reduced mortality), and the quality of those life-years (i.e., reduced morbidity), which is adjusted using a set of quality weights or utilities or value sets, ranging from 0 to 1, indicating a health state equivalent to being dead (0) to representing full health (1) [2, 4, 20, 21]. This approach to measurement was significantly advanced by the seminal works of Klarman and colleagues in 1968 [6], followed by Bush et al. [20] and Torrance et al. [22] in 1972, Loomes and Mckenzie in 1989 [23] and others [24, 25]. These foundational studies led to the formalisation of the term QALY, which has since become the cornerstone of HRQoL outcome measurement for economic evaluation. QALY not only facilitates comparisons across a wide range of interventions but also aligns the value of health interventions with societal preferences when the preference weights are elicited from representative members of society [2, 4]. It has been proposed as an outcome metric that reflects shifts in a population's health status, not just on an individual level, thereby enabling the assessment of the overall performance of a healthcare system [26]. However, to operationalise the concept of QALY, it is necessary to measure and value HRQoL. The measurement of HRQoL involves a comprehensive assessment of all components that constitute HRQoL, typically using validated preference-based instruments such as the EQ-5D or the SF-6D, to describe the health states. The valuation of quality of life involves the assignment of preference weights through the application of a value set for different health states defined by the instrument [2, 4].

1.5 Measurement of HRQoL

The World Health Organisation defines health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" [27]. The widespread adoption of this definition of health has resulted in a paradigm shift in healthcare to encompass outcomes beyond the traditional clinical measures. This is particularly important as many health issues extend beyond mere life or death situations, with the burden of chronic diseases being high in many countries [28], necessitating a more comprehensive measure [29-31]. This led to the introduction of the concept of QoL incorporating the patient's viewpoint into clinical and health-related decisions, thereby complementing objective assessments with a subjective dimension that reflects an individual’s overall experience [29, 30]. This implies that an individual’s perception of their quality of life will be influenced by their environment and personal experiences [32]. However, QoL is a broad term that can encompass non-health-related aspects such as material, economic and political evaluations, making it less specific for health-related assessments [33].

In the context of health economics, the maximisation of HRQoL, a subset of QoL, from the perspective of patients and the general public is particularly important. HRQoL focuses specifically on a person’s well-being across different attributes that are influenced by or can influence one’s health status directly [33]. HRQoL is a multi-dimensional measure that includes mental and social aspects, in addition to physical well-being and functioning. As defined by Patrick and Erickson (1993), HRQoL is "… the value assigned to duration of life as modified by the impairments, functional states, perceptions, and social opportunities that are influenced by disease, injury, treatment, or policy” [26]. It represents a key component of evaluating health outcomes to determine the value of health technologies, and a crucial indicator for appraising their quality [2, 34].

In health economics, HRQoL can be measured using direct or indirect methods. Direct methods involve directly asking individuals to assess their own health on the QALY scale through application of techniques such as the standard gamble (SG) or time trade-off (TTO) [4, 13, 35]. These methods are often complex and require substantial time to administer (discussed in more detail below). A frequently employed alternative involves using indirect methods that utilise established instruments, such as the EQ-5D [36] or the Health Utilities Index (HUI) [37, 38], to measure HRQoL. A preference-weighted scoring algorithm (value set) is then applied based upon general population values attached to health states in the descriptive system pertaining to the instrument facilitating the calculation of QALYs (discussed in more detail below) [2, 4].

Patient-Reported Outcome Measures (PROMs) which evaluate HRQoL, offer a comprehensive classification of health states across multiple dimensions or attributes. Each dimension includes distinct levels of severity that are both mutually exclusive and collectively exhaustive, offering detailed health profiles for individuals [4, 21]. Respondents indicate their health status by selecting the appropriate level within each dimension. The aggregation of these selected levels across all dimensions defines the respondent's current health state. In addition, the dimensions also capture disaggregated details about patient experiences, including specific areas of improvement and decrement. As previously mentioned, measures that facilitate the calculation of preference-weights associated with each unique health state are referred to as preference-based measures (PBMs). With the PBMs, the health states described in the descriptive system are synthesised into a single numerical value using pre-existing (or off-the-shelf) preference-weights or value sets. These value sets are typically developed at a country specific level to represent the preferences of the general population (usually adult) for different health states using one or more of the direct valuation methods outlined below, such as the SG, TTO, VAS or DCE [2]. Preference weights are allocated to each level across the dimensions of the classification system and are subsequently aggregated as specified to derive a utility value for each potential health state [35].

1.6 Valuation of HRQoL

The valuation of HRQoL is generally rooted in the concept of utility, prevalent in both economics and philosophy [2, 4, 13, 21, 39]. It is defined as ‘a cardinal measure of the preference for, or desirability of, a specific level of health status or specific health outcome’ [40], thus, emphasising individual’s desirability of a particular health outcome. In health economics, this concept is operationalised through a utility framework that accounts for individual preferences, thereby ensuring that health outcomes are assessed not just in terms of clinical outcomes but also in terms of their impact on quality of life. Within this framework, any variation in health or function is mirrored by corresponding changes in an individual's health state utility, reflecting both positive outcomes (utility) and negative outcomes (disutility) over a defined intervention and follow up period [4, 21]. To facilitate quantitative assessment, health state utility or utility is indexed on an interval scale of 0-1 (where 0 is equivalent to being dead and 1 is equivalent to full health). Health states considered worse than being dead can be assigned negative values (and are often unbound or restricted to -1 for meaningful statistical analysis) [2, 4].

1.6.1 Valuation methods

Commonly used valuation approaches for health state utility assessments are Standard Gamble (SG) [41], Time-Trade Off (TTO) [42], Visual Analogue Scale (VAS) [2, 4] and Discrete Choice Experiments (DCE) [43, 44].

1.6.1.1 Standard Gamble (SG)

The theoretical background of SG is based on Von Neumann and Morgenstern utility theory, commonly known as expected utility theory, which aligns with economic theories of decision-making under uncertainty. By involving choices under uncertainty, SG reflects real-world decisions where individuals often face risks. SG captures individual preferences and risk attitudes, providing an individual measure of utility [2, 41]. SG requires respondents to choose between a certain (intermediate) health state and a gamble between full health (or state better than the intermediate state) and death (or state worse than the intermediate state). The idea is to determine the respondent's willingness to take a risk (the gamble) to avoid a particular health state [2, 41].

Respondents are given a detailed description of a health state they may experience. This can be a chronic condition or a temporary health state. Simply put, respondents are asked to choose between living in the given health state for a certain period or taking a gamble with two possible outcomes: 1) Full Health: A probability p of achieving full health or 2) Death: A probability 1−p of dying immediately. The probability p is varied until the respondent is indifferent between the certain health state and the gamble. This point of indifference is used to derive the utility value for the health state [2].

However, SG is cognitively demanding, requiring respondents to understand probabilities and make complex decisions involving life and death. Asking respondents to consider death as a possible outcome raises ethical concerns, especially in vulnerable populations like children. Consequently, proxies, such as parents or caregivers, are often used to assess child health state utilities with the SG method [45].

1.6.1.2 Time Trade-Off (TTO)

TTO is a widely used method for valuing health states based on decisions made under certainty. This method was developed primarily for healthcare as a simpler and more user-friendly alternative to the SG technique [46]. TTO involves asking respondents to trade off a period of time in a less desirable health state for a shorter period of time in a more desirable health state, typically full health. Respondents are given a detailed description of a specific health state, which may include its impact on physical, emotional, and social functioning. Respondents are asked to choose between two options: 1) Living a specified number of years (T) in the given health state or 2) Living a shorter number of years (X) in full health. The length of time in perfect health (X) is varied until the respondent is indifferent between the two options. The point of indifference is used to derive the utility value for the health state. The utility value (U) of the health state is calculated as the ratio of the number of years in perfect health (X) to the number of years in the less desirable health state (T), i.e., U= X/T [2, 4].

The composite Time Trade-Off (c-TTO) method is an advanced approach used to elicit health state utility values. It integrates two methods: the conventional Time Trade-Off (TTO), and the ‘lead-time’ or ‘lag-time’ TTO, to elicit health state utilities on a single scale [47]. In the conventional TTO, respondents are asked to trade a specified duration of life (e.g., T=10 years in EQ-5D valuation studies [48]) in a given impaired health state for a shorter duration in perfect health, thus determining the utility value for health states that are preferable to death. However, this approach faces challenges when eliciting values for health states that are considered worse than dead, i.e., having values less than 0. In these cases, the respondent would have to indicate a negative value, which can be conceptually and practically difficult. To address this, the lead-time TTO method is incorporated, where an initial period of full health is added before the impaired health state or lag-time TTO where the additional time in full health is placed after the impaired health state [49, 50]. For health states considered worse than dead, respondents trade off years from this additional ‘trading time’ in full health for the same duration of the impaired health state to make more meaningful comparisons [47].

TTO is relatively straightforward for respondents to understand, as it involves making direct trade-offs between longevity and health quality [2, 42]. However, TTO requires respondents to think about abstract concepts such as future time preferences and trade-offs [2], which can be challenging for some individuals, particularly children. And again, asking individuals to trade off life expectancy can raise ethical issues, especially when dealing with vulnerable populations like children or individuals with severe health conditions.

1.6.1.3 Visual Analogue Scale (VAS)

The VAS, or a rating scale, is a line of a specified length with interval properties and clearly defined endpoints. These endpoints are anchored with verbal descriptors that describe the extremes of the attribute being measured, such as “best imaginable health” and “worst imaginable health”. Respondents mark their position on the line to indicate their perceived value of a given health state, typically with numerical values assigned to different points on the scale [2, 4]. Unlike SG, the VAS does not yield preference-based utility values. Instead, it provides a more intuitive understanding of perceived health states by capturing how individuals feel about their health without necessitating the quantification of preferences in terms of trade-offs or probabilities. This results in a value function rather than a utility function, i.e., measures health gains rather than satisfaction (utility) from the gains [51]. Consequently, the VAS values are often mapped onto SG or TTO value using specified algorithms to generate utilities [51, 52].

The simplicity of the VAS, compared to the more complex SG and Time TTO methods, makes it less time-consuming and cognitively demanding, leading to higher completion rates among respondents [2]. Furthermore, the VAS has been shown to be reliable for use with children aged 5 and above, making it a versatile tool in both paediatric and adult populations [53].

1.6.1.4 Discrete Choice Experiments (DCE)

Unlike the aforementioned approaches (SG, TTO and VAS), DCE is a method used to elicit preferences and value health states through ordinal responses, capturing stated preferences. DCEs are based on the theory of random utility, which posits that the utility an individual derives from a particular option is composed of an observable component (derived from attributes) and a random component (derived from unexplained factors) [43, 44]. Respondents are presented with a series of hypothetical scenarios, each involving a choice between two or more alternatives, each characterised by several attributes. Attributes are the key characteristics or dimensions identified of the health state or healthcare service being evaluated (e.g., symptom severity, treatment side effects, cost). Each attribute has different levels (e.g., mild, moderate, severe). Hypothetical choice sets are created by combining different levels of attributes. Each choice set typically includes two or more alternatives, and respondents must choose their preferred alternative. Each respondent completes several choice tasks, providing data on the relative importance of each attribute and the trade-offs respondents are willing to make [2]. Econometric models (e.g., multinomial logit models, conditional logit models) are used to analyse the ordinal choice data revealing the latent preference weights, typically expressed on an interval scale. To estimate meaningful utilities associated with each attribute level derived from the relative importance of each attribute, these weights need to be normalised or rescaled to a QALY scale (0 to 1). Common methods for this include using exogenously defined values, such as the observed mean value from TTO or SG for specific health states like the ‘worst health state (or the ‘pits’ state)’ or ‘dead’ that can be used as reference points for anchoring and transforming the utility values for all health states on the QALY scale. Additionally, including the ‘dead’ state in DCE and incorporating a dummy variable for it in regression models allows for direct rescaling of coefficients to the QALY scale. Alternatively, as suggested by Rowen et al. [54], using a mapping approach or hybrid models combining data from both ordinal (e.g., ranking or rating scales) and cardinal (e.g., TTO or SG) elicitation techniques may be employed. These methods, among others, ensure that the utility values are standardised and interpretable within the context of the QALY framework [2, 54, 55].

Another technique, within the DCE family, that uses ordinal data for preference elicitation is the profile case Best-Worst Scaling (BWS) which allows respondents to identify the most (best) and least (worst) preferred options within a set of alternatives. This method involves presenting participants with a series of scenarios, each containing a single health profile characterised by various attributes. Participants are asked to select the best and worst attribute of the health state in each scenario, providing clear preferences that can be used to infer the relative importance of different health attributes [2]. The data collected through BWS is then rescaled using one of the methods mentioned above by including a survival attribute and transformation of values to align with the QALY scale, thereby facilitating comparison with other health outcomes [2]. Due to the face-validity and reliability of this approach, Ratcliffe et al. successfully utilised the BWS technique in a sample of Australian adolescents aged 11-17 years, to generate health state valuations for the Child Health Utility 9D (CHU9D) [56].

1.6.2 Challenges in HRQoL valuation

In general, QALYs are considered to be a good approximation of utilities (i.e., individual preferences) [4], but may depend on the valuation technique as there is some debate in the literature regarding the extent to which the preference elicitation techniques (SG or TTO) accurately measure utility [51, 57, 58]. These methods help to quantify how changes in health contributes to overall well-being by asking the respondents to make a choice or trade-offs between health and other factors that contribute to their utility (i.e., opportunity costs) or VAS values [2]. For example, the SG method values health improvements based on the degree of risk (usually the risk of immediate death) a person is willing to take. It assumes that utility decreases as the risk increases. The TTO method values health improvements based on how many years of life expectancy a person is willing to give up, assuming that utility increases with longer life expectancy. While DCE involves asking participants to choose between multiple scenarios that describe different health states with various attributes, and then make decisions based on their preferences. However, none of these techniques can be considered as a 'gold standard' [2]. The preference-elicitation methods such as SG and TTO can be influenced by probability weighting (overweighting small probabilities or underweighting large ones) and loss aversion (preference for avoiding losses over acquiring gains) [59]. In addition, the TTO method may be contaminated by scale compatibility (tendency to respond in a way that fits the scale, i.e., more weight is attached to duration than health state being valued) and utility curvature (the decreasing value of additional life years as more years are added), and often fails to account for time preference (preference for immediate benefits over future ones). Moreover, for the VAS, if the end points are not well defined it may lead to a lack of comparability between individual responses. It is also prone to end-point bias (where health states at the extremes of the scale are perceived as more disparate than they are) and context effects (where the rating of one item is influenced by the level of other items being valued). DCEs, while avoiding the influence of risk aversion or time preferences, can be complicated by the complexity of the choice tasks and the framing of questions, potentially leading to inconsistent preferences. Given these biases, the results from these methods may not align consistently with one another [2, 60]. Therefore, they are often used alongside other valuation methods to provide complementary information, offering a more complete picture of individual health state valuations [61-63]. Regardless, QALYs reflect health gains and serve as a valuable summary measure of health status [51, 64].

The valuation of HRQoL, as expected, is underpinned by value judgments, necessitating decisions about what qualifies as a health benefit or an increase in utility. This leads to the critical question of whose values should be applied to value health states: those experiencing the health states themselves (individual values) or those of the general public for hypothetical health states (social values) [65]. Additionally, the valuations may also depend on age and ethnicity [66], the cultural context [67] or the country of residence [68] of the individual performing the valuations. Although this discussion is essential, it is beyond the scope of this thesis. There is a significant issue in using adult values versus child-specific values in the valuation of child HRQoL. Evidence indicates that these values differ markedly, leading to variations in the child QALY calculations [69, 70]. This can lead to discrepancies in the child health states reflecting societal preferences, thereby potentially misguiding resource allocation and policy decisions related to child health. However, to the extent that regulatory authorities tend to stipulate that (adult) general population values are applied to value HRQoL using established validated preference-based measures, this thesis also considers the country-specific general population perspective adopted for the purposes of health state valuation.

1.7 HRQoL measures

Measures focused specifically on the HRQoL impacted by a particular disease or condition are known as disease-specific or condition-specific measures (e.g., Asthma Quality of Life Questionnaire (AQL-5D) [71], Quality of Life Utility Measure-Core 10 dimensions (QLU-C10D) [72], and the Paediatric Asthma Health Outcome Measure (PAHOM) [73]). Conversely, generic measures (e.g., EQ-5D-Y [36]) assess the impact on HRQoL from any disease or condition and are not limited to a specific one. Thus, these measures evaluate the overall impact on HRQoL and can be applied to any patient group or even to samples from the general population. These measures can be self-completed by individuals to describe their own HRQoL or by using a proxy informant [2, 4]. For any meaningful measurement of HRQoL, the measures must be practical, valid, responsive, and reliable [2, 4, 21, 35]. Practical meaning that the measures are easy to use and can be efficiently administered without requiring too much time or resources [2]. Validity implies that measures accurately capture what they are intended to measure, assessed through measurement properties such as content, construct, and criterion validity [74]. Responsiveness indicates that the measures are sensitive enough to detect changes in health status over time, even if the changes are small [74]. Reliability means that the measures yield consistent results over time (test-retest reliability) or across different observers (inter-rater reliability). An important aspect of reliability considered in this thesis is the inter-rater reliability, which is the degree to which different raters provide consistent estimates of HRQoL ratings [2, 74]. These qualities are collectively referred to as the ‘psychometric properties’ of the measures [35].

1.8 HRQoL measures in children

Child-specific PBMs are designed to capture the HRQoL of children. These measures use societal valuations of different health states to assess the impact of a health condition on the child’s health-related well-being [2, 75]. The application of child-specific PBMs enables the derivation of utilities (preference-weights) for incorporating into CUA of health technologies targeted for paediatric populations [76]. The descriptive systems of child-specific measures are tailored to reflect the unique physical, emotional, and social developmental stages and challenges that children experience, distinguishing them from adult measures [75, 77]. Such age-appropriate concepts may enable self-completion by children/adolescents, allowing them to report their own HRQoL. Typically, the preference-weights for child HRQoL values are elicited from adult samples within the general population [77]. However, there are several issues noted in literature regarding the valuation of child HRQoL [78]. There remains an ongoing debate about whose perspective should be used—whether it should be the children/adolescents themselves [56, 79] or adults valuing the hypothetical child health states [80]. Additionally, questions persist about whether the perspective adopted (e.g., age of the hypothetical child) affects how adults value child health states [69, 81]. These considerations are crucial as they may significantly influence the resulting HRQoL values, and subsequently QALYs, and their application in CEAs/CUAs.

Described below are the validated generic PBMs, identified in a previous review by Chen and Ratcliffe [77], that have been applied to measure and value HRQoL in children and adolescents (see Table 1.2):

1.8.1 Quality of Well-Being Scale (QWB)

The QWB scale is one of the earliest generic preference-based measures, initially developed for adults and was later adapted for use in children [82]. The QWB-SA, a self-administered version, comprises four sections in its descriptive system: chronic symptoms/problems, acute physical symptoms, mental health symptoms/behaviours, and the functional dimensions of mobility, physical activity, and social/self-care activities. The items define a total of 945 health states with preference-weights derived from a sample of US adults (N=430) [77, 82].

1.8.2 Health Utilities Index Mark 2 and 3 (HUI 2 and HUI 3)

The HUI comprises two versions: HUI 2 and HUI 3. HUI 2 was initially designed to assess long-term outcomes in childhood cancer patients but has since been broadly applied across both adult and child populations [38]. HUI 3 is an enhanced version suitable for evaluation in both patient and general population health [37]. HUI 2 evaluates seven dimensions: sensation, mobility, emotion, cognition, self-care, pain, and fertility (with the fertility dimension specifically addressing reproductive health impacts for childhood cancer survivors in their adult life), defining 24,000 health states [38]. HUI 3 assesses eight dimensions: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain encompassing 972,000 potential health states [37]. The preference weights for HUI 2 were derived from parents acting as proxies for school-aged children (N=194) [38]. In contrast, the preference weights for HUI 3 were obtained from a sample of adults from the general population aged 16 years and older (N=256) [37].These measures are appropriate for self-reporting in children as young as 8 years old, while proxy reporting is recommended for children aged 5-8 years [77].

1.8.3 Sixteen-dimensional Measure of HRQoL (16D) and Seventeen-dimensional Measure of HRQoL (17D)

The 16D and 17D are extensions of the adult 15D, specifically designed to measure health-related quality of life in child populations. The 16D targets adolescents aged 12-15 years, while the 17D is designed for children aged 8-11 years. Both instruments encompass 14 shared dimensions: mobility, vision, hearing, breathing, sleeping, eating, excretion, speech, discomfort/symptoms, school/hobbies, friends, physical appearance, depression, and vitality [77, 83, 84]. Additionally, the 16D includes mental function and distress, resulting in a total of 1.5×10^{11} possible health states [83], whereas the 17D incorporates dimensions of anxiety, concentration, and learning/memory, defining 7.6×10^{11} possible health states in all [84]. These additional dimensions are tailored to address the unique aspects of HRQoL pertinent to younger children [84]. The preference weights for 16D were derived from school-aged adolescents aged 12-15 years (N=213) [83]. In contrast, the preference weights for 17D were obtained from a sample of parents from the general population valuing for child health states (N=115) [84].

1.8.4 Adolescent Quality of Life-6D (AQoL-6D)

The AQoL-6D measure, adapted from the adult Assessment of Quality of Life (AQoL) measure, builds upon the original four dimensions of the AQoL. The adolescent version was developed with contributions from adolescents in Australia, New Zealand, Fiji, and Tonga to ensure it is culturally and semantically relevant within the Pacific region. It assesses six dimensions: independent living, relationships, psychological health, senses, pain, and coping, comprising a total of 20 items and defining 7.8×10^{13} possible health states. A sample of adolescents (N=279) from the four countries were utilised to obtain the value set for this measure [77, 85].

1.8.5 EQ-5D Youth Version (EQ-5D-Y)

The EuroQol group adapted and validated two child-specific measures, the EQ-5D-Y three-level (EQ-5D-Y-3L) and five-level (EQ-5D-Y-5L) from the corresponding adult EQ-5D versions. As with the EQ-5D (3L and 5L respectively) for adults, the EQ-5D-Y includes five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, with wording adapted for use in child populations [36, 86-88]. The EQ-5D-Y-3L has three response levels per dimension, describing a total of 243 possible health states [36]. In contrast, the EQ-5D-Y-5L has five response levels per dimension, describing a total of 3,125 possible health states [87]. In addition, the EQ-5D-Y (3L and 5L) 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 [36, 77, 87]. Although a preference-based Australian scoring algorithm for EQ-5D-Y (3L and 5L) is under development, value sets have been published for nine countries since the International Valuation Protocol for the EQ-5D-Y-3L was published (see section 8.4.2 for more details) [89]. The recommended age for self-completion for the EQ-5D-Y (3L and 5L) is 8 years and older, while proxy completion is advised for children aged 4-7 years [36, 77, 87].

1.8.6 Child Health Utility 9D (CHU9D)

Unlike the other measures, the CHU9D was specifically designed for young people to self-report. Initially developed for children aged 7-11 years [90], it was later validated for adolescents aged 11-17 years [91]. It includes nine dimensions: worried, sad, pain, tired, annoyed, schoolwork/homework, sleep, daily routine, and ability to join in activities. The CHU9D defines 1,953,125 potential health states [90]. An adolescent-specific valuation of the health states obtained using a sample of Australian adolescents aged 11-17 years (N=590) is available [56].

1.8.7 Adolescent Health Utility Measure (AHUM)

The AHUM is a recently developed generic PBM developed for adolescents, with its dimensions based on concepts from EQ-5D and SF-6D, literature reviews, and feedback from children with Hunter syndrome and their caregivers. It assesses six dimensions: self-care, pain, mobility, strenuous activities, self-image, and health perceptions, constituting 16,800 potential health states in total. The health state valuation was derived from a sample of UK adults (N=312) from the general population. Although designed for adolescents, AHUM can also be applied with adults, providing a flexible tool HRQoL assessment across a wider age range [77, 92].

Table 1-2 Characteristics of the generic child-specific preference-based measures.

Measure	Country of origin	Age range	Respondent	Mode of administration	Recall period	Dimensions/Items	Items	Response levels	Health states defined	Valuation sample example*
Quality of Well-Being Scale (QWB)	USA	Not applicable	Self or proxy-reported	Self or interviewer-administered	Past 3 days, not including today	3 + 58 symptoms	76	2 to 4	945	US adults-self from general population (N=430)
Health Utilities Index Mark 2 (HUI 2)	Canada	5 and above	Proxy for 5-8 years, Self for 8+ years	Self or interviewer-administered	Current (past 1 week, past 2 weeks, past 4 weeks) or usual	7	7	3 to 5	24,000	Parents as proxies for school-aged children (N=194)

Measure	Country of origin	Age range	Respondent	Mode of administration	Recall period	Dimensions/Items	Items	Response levels	Health states defined	Valuation sample example*
Health Utilities Index Mark 3 (HUI 3)	Canada	5 and above	Proxy for 5-8 years, Self for 8+ years	Self or interviewer-administered	Current (past 1 week, past 2 weeks, past 4 weeks) or usual	8	8	5 to 6	972,000	Adults-self from general population aged 16+ years (N=256)
Sixteen-dimensional Measure of HRQoL (16D)	Finland	12-15 years	Self or proxy-reported	Self or interviewer-administered	Today	16	16	5	1.5 x 10^{11}	Adolescents-self aged 12-15 years (N=213)
Seventeen-dimensional Measure of	Finland	8-11 years	Proxy for below 8	Interviewer-administered	Today	17	17	5	7.6 x 10^{11}	Parents rating for

Measure	Country of origin	Age range	Respondent	Mode of administration	Recall period	Dimensions/Items	Items	Response levels	Health states defined	Valuation sample example*
HRQoL (17D)			years, Self for 8+ years							child (N=115)
Adolescent Quality of Life-6D (AQoL-6D)	Australia	Adolescents	Self-reported	Self-administered	Not specified	6	20	4 to 6	7.8 x 10^{13}	Adolescents-self from Australia, New Zealand, Fiji, and Tonga (N=279)
EQ-5D Youth Version (EQ-5D-Y-3L)	Europe	8-15 years	Self-reported (8+), Proxy (4-7 years)	Self-administered	Today	5	5	3	243	Various countries (international child value sets based

Measure	Country of origin	Age range	Respondent	Mode of administration	Recall period	Dimensions/Items	Items	Response levels	Health states defined	Valuation sample example*
										on published protocol)
EQ-5D Youth Version (EQ-5D-Y-5L)	Europe	8-15 years	Self-reported (8+), Proxy (4-7 years)	Self-administered	Today	5	5	5	3,125	Valuation protocol currently under development
Child Health Utility 9D (CHU9D)	UK	7-17 years	Self or proxy-reported	Self-administered	Last night/today	9	9	5	1,953,125	Australian adolescents-self aged 11-17 years (N=590)

Measure	Country of origin	Age range	Respondent	Mode of administration	Recall period	Dimensions/Items	Items	Response levels	Health states defined	Valuation sample example*
Adolescent Health Utility Measure (AHUM)	UK	Children, adolescents, adults	Self-reported	Self-administered	Not specified	6	6	4 to 7	16,800	UK adults-self from general population (N=312)

Table adapted from Chen and Ratcliffe [77], *The valuation samples presented in this table are illustrative examples and are not exhaustive. Additional value sets, such as the UK value set for the CHU9D, may be available for many of these measures.

1.9 Challenges in HRQoL measurement in children

A common theme in the discussion of QoL is consumer sovereignty, a fundamental concept in economics that suggests consumers are the best judges of their own welfare [13]. HRQoL measures apply this principle to evaluate health programs by enabling individuals to assess their own health status. This principle is increasingly applied to children through child-specific HRQoL measures, recognising the importance of considering their perspectives in the measurement and valuation of their own HRQoL [93, 94]. Whilst self-reported child HRQoL measures are important, implementing this approach in children presents significant challenges and, as a result, guidelines for measuring and valuing children's HRQoL remain ambiguous [95].

In addition, most of the currently available generic child-specific PBMs for self-report have been adapted from existing adult HRQoL measures, except the CHU9D and the AHUM [77]. This adaptation process often involves modifying language and simplifying content to make it more understandable for children. However, the core descriptive systems of these measures typically retain an adult-centric perspective, which may lead to methodological issues when applied to younger populations. These adult-focused measures may overlook or inadequately capture dimensions that are most valuable or relevant to children and adolescents, such as aspects of social interaction or school life [96]. Consequently, the resultant HRQoL data may not fully reflect the unique experiences and health-related concerns of children, potentially compromising the accuracy and validity of these assessments in paediatric contexts.

1.9.1 The child-proxy dilemma

As discussed previously, CUA, the most prevalent form of economic evaluation, involves the comparative analysis of alternative programs to determine the most cost-effective interventions where effectiveness is measured and valued according to QALYs as the main measure of outcome [4]. In general, for child health interventions the HRQoL values generated, and subsequently the QALYs, can vary significantly due to several factors such as the specific measure used due to difference in the dimensions or attributes they encompass, the mode of administration, the valuation method employed, the type of respondent and perspective involved in the measurement and valuation process, and the age thresholds specified for the measure [97]. The primary focus of this thesis is to explore the methodological issues related to the type of respondent utilised in measuring child HRQoL and the resulting problem of inter-rater agreement between these respondents.

Economic evaluations rely on standardised processes to ensure that the comparisons between different programs are meaningful and reliable [2, 4]. In the context of child HRQoL, the choice between child self-reports and proxy reports (typically provided by parents or caregivers) presents a significant methodological challenge. When HRQoL values derived from child and proxy reports are not aligned, the standardisation of these evaluations is compromised. Currently, there is no universally accepted guideline indicating whether child or proxy values should be preferred in such evaluations [98]. The lack of guidance is concerning as it introduces variability in HRQoL measurements, thereby impacting the reliability of these measures. In the presence of disagreement between child self-reports and proxy reports, the resulting QALYs, may not accurately represent the child's actual HRQoL. Inconsistent QALYs calculated between child and proxy reports can undermine the validity of economic evaluations, affecting the comparability of programs and potentially leading to suboptimal policy decisions [97].

Investigating the inter-rater agreement between child self-reports and proxy reports in measuring HRQoL is essential to address these issues. By understanding the degree of agreement or disagreement, we can better assess the reliability of the values used in economic evaluations. This investigation will contribute to the development of guidelines on the appropriate use of child versus proxy HRQoL values, thereby enhancing the reliability and, therefore, comparability of interventions targeted at child populations.

1.9.2 The case for proxy-report

According to the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Good Research Practices Patient-Reported Outcomes (PRO) task force report, there is insufficient evidence to determine whether self-reporting of HRQoL by children under eight years of age is reliable or valid [98]. A recent systematic review reported that children with cognitive processing challenges, particularly ADHD, learning disability, speech impairments or special health care needs are more likely to have limited self-report capacity [99]. Children may also be unable to self-report their own HRQoL due to illness or if they lack capacity [100, 101]. In such situation, parents, caregivers, teachers and/or health professionals may be required to act as proxies to provide an informed estimate of the child's HRQoL on their behalf [94, 98]. Proxy-reports typically reflect their own perspectives to report the child's HRQoL (proxy-proxy perspective) or represent that of the child (proxy-child perspective) [77]. In the proxy-proxy perspective, the proxy reports based on their own judgment and observations of the child's health and well-being. In contrast, the proxy-child perspective involves the proxy attempting to represent the child's views. These reports may be completed without necessarily consulting or interacting with the child to gather their input [77].

1.9.3 The case for child-self report

Although proxy-reports of child HRQoL are useful when children are too young to self-report [98], it is important that children are encouraged to self-report their own HRQoL wherever possible [102]. It has been shown that children, as young as six years, are able to report on their own health when age-appropriate questionnaires are used [103]. More generally, there is an emerging consensus that children aged 8 and above are able to self-report their own HRQoL [77, 99]. It is well documented that proxy assessments of HRQoL in any population group tend to differ from self-assessments with proxy assessors typically reporting lower HRQoL than the person themselves [2, 98, 104, 105]. While proxies can provide valuable information about their child's HRQoL, it is important to note that they may not always have the same perception as the child [106]. The challenges in assessing child-self and proxy reported HRQoL are nuanced and distinct from those in adult populations. In comparison to 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 [98].

Among the potential proxies for assessing a child's health-related quality of life (HRQoL), such as clinicians, parents, and teachers, parents are often considered the most appropriate due to their intimate familiarity with their child's health. However, evidence suggests that parental assessment of the child’s HRQoL may be influenced by external factors, e.g. parent’s assessment of the child’s HRQoL may be influenced by their own HRQoL [107, 108]. While clinicians possess detailed knowledge of a child's physical health conditions and can offer valuable medical insights, their limited contact with children outside clinical settings restricts their ability to capture the full spectrum of the child's day-to-day experiences and emotional states [109, 110]. Teachers, on the other hand, may offer valuable perspectives on a child's physical and emotional functioning, particularly within the school environment. Despite this, their observations may be limited to the school setting, potentially missing out on aspects of the child's home life and personal health [111].

1.9.4 Current practices

A review by Wolstenholme and colleagues found that the majority of studies applying PBMs used proxy-reports of child HRQoL, frequently obtained from parents, even for children up to

18 years of age [112]. Another systematic review that examined the economic evaluation of medical devices in paediatric populations identified seven CUAs that included prospectively measured child health state utilities; however, all of these utilities were derived solely from proxy reports [113]. Bailey et al. examined the use of child-specific HRQoL in submissions to the PBAC for decisions concerning funding of medicines and vaccines. In one of the public summary documents, their review noted the PBAC's concern regarding the use of parent proxies to interpret HUI 2 values for children, which created uncertainty in determining the cost-effectiveness of interventions [114].

1.10 Research Questions

In the light of these challenges, this thesis seeks to address the following research questions:

1.10.1 What is the level of inter-rater agreement found in existing literature between self-reports by children and proxy-reports of child HRQoL?

Two previous systematic reviews by Khadka et al. and Jiang et al of child-self and proxy reported child utilities found that utilities tended to differ with proxies often underestimating the child's HRQoL [115, 116]. Child HRQoL ratings obtained by two different observers, the child-self, and the proxy, are likely to differ owing to the differences in their perspectives. Therefore, it is also important to determine the extent to which the two raters agree or assign the same rating for an item being measured. That is, to report inter-rater agreement measures which estimate the strength of agreement between raters [117, 118]. In their review of parent-child reports of HRQoL predominantly using the generic non-preference based PedsQL™, Eiser and Varni [106] reported that the level of agreement between parents and children may be influenced by several variables. Potential factors identified as contributing to limited parent-child agreement included the type of dimension assessed [106]. Similar to the findings in the studies assessing self and proxy concordance in the reporting of HRQoL within the adult population [119, 120], dimensions associated with objective aspects of health typically showed higher agreement as compared to the more subjective (emotional or social) dimensions [106, 115]. Given the discrepancies between child and proxy reports, this thesis seeks to explore the inter-rater agreement between children's self-reports and proxy reports of HRQoL. This investigation will focus on overall HRQoL, as indicated by utility scores for PBMs and summary scores for non-preference-based measures, as well as HRQoL at the dimension level.

Jiang et al examined the difference in self and proxy reported utilities in their systematic review. However, it is also crucial to evaluate how closely the report provided by the parents aligns with the child's self-report, i.e., inter-rater agreement, to determine the extent to which the parental-report is representative of the child’s own HRQoL. The systematic review in Chapter 2 sought to add to the existing systematic evidence [115, 116] by focusing on reported measures of agreement in child and proxy assessments of paediatric HRQoL using established generic preference-based measures, highlighting individual dimension level differences in agreement in addition to overall utilities.

1.10.2 How does the proxy perspective influence inter-rater agreement in the measurement of child HRQoL?

The perspective from which HRQoL is valued is an important consideration in the context of health technology assessments [88, 121]. Likewise, in the measurement of HRQoL, the additional information obtained from various perspectives can contribute to a comprehensive understanding of the HRQoL [122] and offer additional methodological rigor by offering opportunities for triangulation of perspectives [123]. 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 [124]. 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 [125]. 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) [122]. This is identified through the difference between the HRQoL ratings produced using the two proxy versions, namely, the intra-proxy gap. The inter-rater agreement and the intra-proxy gap using the EQ-5D-Y-3L in the community sample is explored in Chapter 4.

1.10.3 How does the age of the child influence the inter-rater agreement in the measurement of child HRQoL?

The age of the child is another important factor that may impact the child-parent agreement in the assessment of child HRQoL. However, the role of age is not yet clearly understood with inconsistent results reported for different age-groups [106]. A study by Cremeens, Eiser and Blades suggested that the age of the child may influence the level of agreement for the PedsQL™ and may interact with the specific dimension being assessed [126]. In a sample of healthy children aged 5.5-8.5 years, they reported a significant agreement between older children (7.5-8.5 years) and parents for overall HRQoL. However, at the dimension level, a significant agreement was observed for the younger children (5.5-6.5 years) within the physical health dimensions and for the older children within the psychosocial dimensions (7.5-8.5 years) [126]. To date the differential effect of age on agreement remains largely unexplored, particularly using preference-based measures. Chapter 5 investigates the inter-rater agreement using two measures, the CHU9D (PBM) and the PedsQL™ (non-preference based) across various age-groups in the community sample. The primary distinction between the two measures is that the former generates health state utilities [127]. Although, it is not possible to calculate utilities for the purposes of applying PedsQL™ in CUAs, the instrument has been widely applied and recognised as a valuable tool for measuring HRQoL in a variety of paediatric populations in both clinical and research settings [128].

1.10.4 How well do children understand the dimensions of HRQoL measures?

Traditionally, less emphasis has been placed on how children interpret and respond to self-report dimensions within HRQoL measures, with greater attention given to establishing the psychometric properties of newly developed measures. Typically, when evaluating the psychometric properties of a measure, its validity, reliability, responsiveness, acceptability, feasibility, and consistency are assessed [129-131]. While these properties can provide critical insights, they may not offer a complete picture of the self-report reliability in young children. Furthermore, a wide range of studies, including those focussing on psychometric validation [132-135] and population health assessments [136-138], often include predominantly healthy children from the general population either as sole participants or as comparators. This may also be an important consideration in longitudinal studies, where it is crucial that changes in health status reflect true changes rather than due to changes in the understanding of the HRQoL dimension presented [139]. Therefore, to ensure that the child's self-reports are valid it is essential to examine whether children, especially those from the general population, can understand the concepts conveyed by HRQoL dimensions and provide meaningful responses [140]. This is examined for the two measures, EQ-5D-Y-3L and the CHU9D, using a mixed-method study design in the community sample in Chapter 6 and 7 respectively.

1.10.5 What is the minimum age at which children can reliably self-report their HRQoL?

There is considerable variability in the recommended cutoff age for children to complete self-report measures of HRQoL. Unlike several other generic non-preference-based measures (e.g., KIDSCREEN) [141], which have a lower age limit for self-report starting at 6 years, the EQ-5D-Y measures (3L and 5L) have been recommended for self-completion by children aged 8 years and above whilst for the CHU9D the recommended age is 7 years and older [77]. 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 able to reliably self-report using HRQoL measures [9] and a number of studies have successfully administered the EQ-5D-Y-3L in cohorts of children aged 8-18 years [142-145]. 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 [146]. 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 [147]. Bray and colleagues also utilised a sub-sample of children aged 6-7 years to assess HRQoL in children with impaired mobility using measures including the EQ-5D-Y [148]. Therefore, to investigate the minimum age at which children can meaningfully self-report their HRQoL, the results from a community sample of children aged 6-12 years and their parents comprising the first empirical (Study 1) will be analysed. The findings from Study Sample 1 will be detailed in Chapters 4, 5, 6, and 7.

1.10.6 How does the use of different value sets impact the inter-rater gap when utilising the EQ-5D-Y-3L?

For measures such as the EQ-5D-Y-3L, Australian child-specific preference-weights are not yet available. Therefore, researchers and practitioners need to make the decision about whether to apply child-specific weights from a different country or whether to apply Australian adult weights, thereby maintaining the cultural representation of the sample. Presently, there is little evidence available to guide this decision or its impact upon the resulting values and the level of inter-rater agreement between child parent dyads. That is, the extent to which cultural differences versus the perspective used to value health states (i.e., adult versus child perspectives) influence HRQoL assessments is yet to be explored. To address this, data from a larger community sample from across Australia, designated as

Study 2, will be analysed. This study will examine the impact of employing different cross-cultural value sets on the consistency of child-proxy reports of child preference-weighted HRQoL using the EQ-5D-Y-3L within an Australian setting. Through the analysis of data from the second empirical study (Study 2) comprising Australian child-parent dyads using diverse international value sets, Chapter 7 intends to explore the cultural adaptability of the preference-weights associated with the EQ-5D-Y-3L measure.

The thesis concludes with a discussion of the implications of the findings of both, Study 1 and 2, in Chapter 9.

CHAPTER 2: ARE WE AGREED? SELF-VERSUS PROXY-REPORTING OF PAEDIATRIC HEALTH-RELATED QUALITY OF LIFE (HRQOL) USING GENERIC PREFERENCE-BASED MEASURES: A SYSTEMATIC REVIEW & META-ANALYSIS

This chapter contains material from:

“Khanna, D., Khadka, J., Mpundu-Kaambwa, C., Lay, K., Russo, R., Ratcliffe, J., & Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team (2022). 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, 40(11), 1043–1067. https://doi.org/10.1007/s40273-022-01177-z”

2.1 Overview

This chapter synthesises existing evidence examining the child-proxy agreement in the assessment of child HRQoL. It includes the methods and findings from a meta-analysis of reported agreement statistics to provide an overall indication of the extent of agreement in child-self and proxy assessments of paediatric HRQoL according to the available evidence.

2.2 Summary

Objective: To provide a comprehensive review of existing literature on inter-rater agreement between child-self (under 18 years) and proxy-reported HRQoL using generic preference-based measures.

Methods: A systematic review of primary studies which reported agreement statistics for self and proxy assessments of overall and/or dimension level child HRQoL using generic preference-based measures was conducted. Where available, data on agreement measures such as the Intraclass Correlation Coefficients (ICCs) were extracted to summarise overall HRQoL agreement levels. Dimension level agreement was described using measures such as the Cohen’s Kappa. A meta-analysis was also performed to synthesise studies and estimate the level of agreement between self and proxy reported child HRQoL at both overall and dimension level.

Results: Of the 35 studies included, 29 reported inter-rater agreement for overall HRQoL, whilst 19 reported dimension-specific agreement. Seven generic preference-based measures were identified as having been applied: Health Utilities Index (HUI) Mark 2 and 3, EQ-5D (EQ-5D-Y-3L, EQ-5D-Y-5L and EQ-5D-3L) measures, Child Health Utility 9 Dimensions (CHU9D), and the Quality of Well-Being (QWB) scale. A total of 92 dyad samples were included with a total pooled sample of 4929 children and 5156 proxies. Most of the identified studies reported a low (ICC=<0.5) inter-rater agreement for the overall HRQoL. In contrast to more observable HRQoL dimensions relating to physical health and functioning, the inter-rater agreement was lower for psychosocial related dimensions e.g., “emotion” and “cognition” attributes of both HUI 2 and HUI 3 and “feeling worried, sad or unhappy” and “having pain or discomfort” dimensions of the EQ-5D-Y measures. Parents demonstrated a higher level of agreement with children relative to health professionals. Child-self and proxy reports of HRQoL showed lower agreement in cancer-related studies than in non-cancer-related studies. The overall ICC from the meta-analysis was estimated to be 0.50 (95% CI 0.36 to 0.62) with moderate inter-rater agreement.

Conclusions: This study provides evidence from a systematic review of studies reporting dyad assessments to demonstrate the discrepancies in inter-rater agreement between child and proxy reporting of overall and dimension level paediatric HRQoL using generic preference-based measures. Further research to drive the inclusion of children in self-

reporting their own HRQoL wherever possible and limiting the reliance on proxy reporting of children's HRQoL is warranted.

2.3 Key Points

- The application of child-specific preference-based measures enables the calculation of utilities for cost utility analysis of health technologies targeted for paediatric populations.
- Proxy-reports (e.g., parent/guardian or a health professional), used in lieu of child-self reports in circumstances when self-reports are not feasible, can often diverge from the child's assessment of their own HRQoL.
- This review examined the agreement between the child-self and proxy reported overall and dimension level HRQoL using generic preference-based measures.
- The inter-rater agreement was, in general, low (ICC=0.5) for overall HRQoL across the measure/s applied and/or the context of the application. In addition, the agreement between children and proxy respondents was lower for psychosocial-related dimensions than for physical dimensions within the respective measures.

2.4 Methods

2.4.1 Search Strategy

The literature search strategy was adapted from a previous study undertaken by Khadka et al., and the search keywords were reproduced [115]. The time frame covered by the previous search was from inception to 30th July 2017. To reflect the latest publications during the four-year period since the initial search undertaken by Khadka and colleagues, this review incorporated peer-reviewed articles published in electronic journals between 30th June 2017 and 19th May 2021. A second iteration of the search was conducted to update this systematic review on 7th March 2024. The online databases searched included

PubMed, the Cochrane Library, Web of Science, EconLit, Embase, PsycINFO and CINAHL (via EBSCOhost). Key words such as “utility”, “quality-adjusted life years”, “children”, “adolescents”, “preference-based measure of HRQoL” and related Medical Subject Headings (MeSH) terms were used for the systematic literature search. A detailed account of the search terms and the strategy is presented in Appendix Table 10.1. The identified studies were screened using the web-based systematic review software Covidence [149]. This review is registered with the International Prospective Register of Systematic Reviews (PROSPERO registration number CRD42021256815, see Appendix Fig 10.1). The Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) Statement guidelines were used for reporting this review (Appendix Table 10.2) [150].

2.4.2 Inclusion and exclusion criteria

All studies published in English with full text availability were included. Eligible studies included primary studies applying generic preference-based measures to report child HRQoL, which could potentially be used to derive health utilities amenable to QALY calculations in a paediatric population as assessed by the child (*child* or *children*, from hereon, refer to all school-age children and adolescents, i.e., between 5-18 years of age unless stated otherwise) and proxy dyads. Inclusion criteria were studies reporting the agreement level for overall and/or dimension level child HRQoL by both children and the proxies reporting on behalf of children. Those studies that reported the child-self and proxy assessed overall and/or dimension level HRQoL but did not include the agreement statistics were excluded. Additionally, as this systematic review focused on studies applying generic preference-based HRQoL measures, primary studies conducted among the child populations were excluded if the health utilities were obtained (i) directly using SG, TTO and VAS (except EQ VAS) or (ii) indirectly using condition-specific (as opposed to generic) HRQoL measures.

2.4.3 Article screening

Article screening was carried out in three steps. In the first step, two independent reviewers (DK and KL) screened the titles and abstracts based on the inclusion and exclusion criteria. Records with conflicting decisions were deferred to a third reviewer to reach a consensus. Articles selected at the screening stage were then included for a full-text review in the second step. The same two reviewers reviewed all the articles included in this stage. Simultaneously, two other reviewers (JK and CMK) independently assessed 10% of the articles in total to confirm the decisions of the former pair of reviewers. Following a discussion with the initial reviewing pair and the other reviewers (JR, JK, CMK) to reach a consensus, full-text articles that met the criteria were included. In the final step of this process, all the eligible articles were subsequently consolidated and information relevant to the study was extracted.

2.4.4 Data extraction

Data extraction was performed by the first author (DK). Each article was assessed to retrieve the following information: bibliographic details, geographic setting, study design, health state experienced, the generic preference-based measure used, target sample size, age range of children included, sample gender composition, proxy type and sample size, mode of administration for both individuals in the dyad, statistical test(s) that report the overall and/or dimension level agreement between self and proxy reported HRQoL, and any reported methodological concerns. A Microsoft Excel (Version 2019) database was used to enter and store the extracted data.

2.4.5 Extraction and interpretation of agreement statistics

Inter-rater agreement is the degree to which the assessments of two or more individuals (raters) are identical using the same measure and assessing the same subject. There are multiple methods to measure inter-rater agreement based on the type of variable (continuous or categorical) and the number of raters. Agreement measures such as the intraclass correlation coefficient (ICC), Concordance Correlation Coefficient (CCC), Cohen's Kappa (κ), Bland-Altman plots, percent agreement and Gwet's agreement coefficient (AC) (unweighted= AC_1) assess the degree to which the assessments by the individual raters are identical or in agreement based on the type of data (e.g., nominal or continuous) [118, 151, 152]. Correlation coefficients, also commonly reported to indicate agreement, determine the linear relationship between two continuous variables (Pearson's product-moment correlation or Pearson's *r*) or two ranked variables (Spearman's rho) [153].

It is important to note that in statistical analysis, correlation coefficients (e.g., Pearson's *r*) are considered as sub-optimal measures of inter-rater agreement. They only provide a measure of the strength of a linear association between scores by raters and may indicate strong correlations even in the presence of a significant difference between the HRQoL assessments if the scores by both raters vary similarly. As a result, correlation coefficients may over or underestimate the true level of agreement and inaccurately reflect the degree of agreement between raters [118, 153-155]. Inter-rater agreement is also often estimated using the percent agreement approach [155]. However, percent agreement does not correct for the level of agreement resulting from a random decision made by the raters. Cohen's Kappa and Gwet's AC account for this random agreement and are more robust [156]. Therefore, percent agreement is excluded from this review as a measure of child and proxy agreement. Only two studies reported the inter-rater agreement using the Bland-Altman plot and were, thus, not included in this review.

Thus, in the present study, to examine the concordance in the paediatric HRQoL obtained by self and proxy reports, ICC/CCC and Kappa/Gwet's AC values were treated as primary evidence. In addition, the results of the correlations coefficients, both Pearson's *r* and Spearman's rho, were presented as supplementary evidence.

ICCs/CCCs can take a value between 0 to 1, whereas Kappa and correlation coefficient statistics range from -1 to 1. Values for ICCs/CCCs less than 0.5 indicate poor agreement between raters, whereas values between 0.5 and 0.75, 0.75 and 0.9, and greater than 0.9 indicate moderate, good, and excellent agreement respectively [157]. Spearman's correlation coefficients with a value less than 0.20 represent no correlation, values between 0.20 and 0.35 represent weak correlation; values between 0.35 and 0.50 represent moderate correlation, and values greater than or equal to 0.50 represent strong correlation [158]. Pearson's *r* coefficients are interpreted using Cohen's conventions. The correlation is small if the coefficient is 0.30 or less, medium if itis 0.50 or less, and large if it is greater than 0.50 [159]. Cohen's Kappa and Gwet's AC_1 have similarly defined thresholds: Landis and Koch [160] and Altman's scale (in brackets) [161] with classifications defined as slight (poor), fair, moderate, substantial (good) and almost perfect (very good) correlation for values less than or equal to 0.2, 0.4, 0.6, 0.8 and 1 respectively. For consistency across the analysis, we interpret the agreement coefficients based on Altman's scale.

2.4.6 Data synthesis and analysis

The estimates of the agreement level between child-self and proxy reported HRQoL were described using a textual approach in the form of a narrative synthesis [162, 163]. Several studies did not report the mean age of participating children in the dyad, and hence only the age range was analysed. Studies which included children with cancer along with other chronic illnesses were identified as non-cancer-related studies. Caregivers reporting as proxies on behalf of children were grouped under parents. When the type of correlation was not mentioned in the study, it was assumed to be Pearson's *r*, as this is the most commonly used correlation coefficient in similar research [115].

A meta-analysis was performed on a sub-set of the studies to synthesise the quantitative information and estimate the overall and dimension level agreement between child-self and proxy reported HRQoL. In this meta-analysis, we distinguish between the overall HRQoL and utilities as we only include those HRQoL scores that are obtained by applying the respective value sets to derive preference-weighted HRQoL values. For instance, the studies reporting the HRQoL scores derived from the EQ VAS would be excluded from the meta-analysis as there is some debate in the literature about VAS scores and the extent to which they can be interpreted as utilities [2]. To obtain an average estimate of inter-rater agreement, the ICCs/CCCs for overall utilities were synthesised as they are reported on a continuous scale. Similarly, considering the ordinal nature of the responses within the attributes, Kappa statistic was used to estimate the dimension level inter-rater agreement. Studies reporting only the correlation coefficients were excluded from the meta-analysis. Furthermore, studies reporting Gwet's AC_1 were also excluded from the analysis because: 1) the inherent differences between the two statistics, Kappa and Gwet's AC_1, prevent their combination, as it remains unclear whether they yield comparable outcomes [151, 164]; and 2) to date, there are no published guidelines currently available on the pooling of Gwet's AC_1 estimates.

The meta-analysis was conducted using Stata (16.1, Stata Corp LLC, College Station, TX). Since the assumption of homogeneity is not reasonable for the present data due to the diverse nature of the target samples in consideration, a random-effects model was used to allow for between study variability in effect sizes. The weights were estimated using a restricted maximum likelihood (REML) method [165]. A Fisher's z-transformation was applied to obtain an approximately normal sampling distribution in order to calculate the 95% confidence intervals (CI) for each ICC/CCC for the overall utilities. The z-scores were then transformed back into correlations for ease of interpretation [166].

For the dimension level meta-analysis, the standard errors (se) for Kappa values ($\hat{\kappa}$) were calculated using the following formula:

$$se_{\kappa} = \sqrt{\frac{p(1-p)}{n(1-p_c)^2}}$$

Where p is the observed percent agreement, n is the number of rater pairs and p_c is the agreement expected by chance. However, since no study reported the values for p_c, but did report p and $\hat{\kappa}$, p_c was calculated as [167]:

$$p_c = \sqrt{\frac{p - \hat{\kappa}}{1 - \hat{\kappa}}}$$

A forest plot was used to depict the results of the meta-analysis of the overall agreement. Heterogeneity was assessed using a forest plot as well as Cochran's test of homogeneity (Q statistic) and the I^2 statistic. Each sample was considered unique if any of the following variables relevant to the analysis were unique: type of proxy, measure, health condition, or if children below 8 years were included in the sample. An exploratory meta-analysis assuming a random-effects model was conducted to estimate the moderation by these variables. A random effect meta-regression was used to supplement the findings of the meta-analysis, as the studies were not considered sufficiently similar for a fixed-effects model [168]. The sample was also considered to be unique if the same sample was examined in a different time period for longitudinal studies. Publication bias was evaluated using funnel-plots and a regression-based funnel plot asymmetry test.

2.4.7 Risk of bias and quality assessment

Two independent reviewers (DK and JK) appraised the quality and suitability of the included studies. The overall reporting quality score was calculated using a checklist for quantitative studies as given by Kmet et al., 2004 and was used to assess the risk of bias [169]. From each of the selected articles that met the inclusion criteria, information for fourteen quality indicator variables was extracted (details provided in Appendix Table 10.3). Two points were assigned to each of these variables if they were appropriately reported in the article, one if the item was incompletely reported and none if not reported at all. The sum of all the points indicated the overall reporting quality score of the article with twenty-eight being the maximum. The summary scores were rescaled between 0 and 1 with 1 denoting the highest quality. If the item was not applicable to a particular study, scores were adjusted by excluding the total possible scores of those items from the summary score. The minimum threshold for inclusion of studies based on quality scores was set at 0.6. A sensitivity analysis was carried out using the criteria by Papaioannou and colleagues to confirm the conclusions from the former appraisal, with the results reported in Appendix Table 10.4 [170].

2.5 Results

2.5.1 Search results

A PRISMA flow diagram illustrates the selection process (Figure 2.1). An extensive literature search of seven databases was conducted using the search strategy described above. 46,180 records published between 30th June 2017 and 7th March 2024 (updated date) were identified and were subsequently imported into Covidence. 21,324 records were deduplicated by Covidence, leaving 24,856 records for title and abstract screening. Of these, the vast majority (23,107) were excluded. Reasons for exclusion were: (i) non-primary studies, (ii) non-paediatric target population, (iii) no health state utilities reported (iv) inaccessible articles and (v) English was not the main language of publication. Subsequently, 749 records were included in the full text review stage. At this stage, in addition to the previously specified exclusion criteria, studies were excluded if agreement statistics between child-self and the proxy reported health state utilities and/or at dimension level were not reported. 20[1] studies were identified from the previous systematic review by Khadka et al. [115]. In the two iterations of the searches, ten studies were initially identified followed by

---

[1] 21 studies were included from the previous review by Khadka et al. in the final review. The two papers by Glaser, A., et al. (1999), ‘Standardized quantitative assessment of brain tumor survivors treated within clinical trials in childhood’ and ‘Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the U.K.’ were published in two different journals but used the same sample to report different results. To prevent double counting, these two papers were considered as one.

four[2] new studies in the updated literature search. In total, 34 studies fully met the inclusion criteria and were, thus, included in the final review.

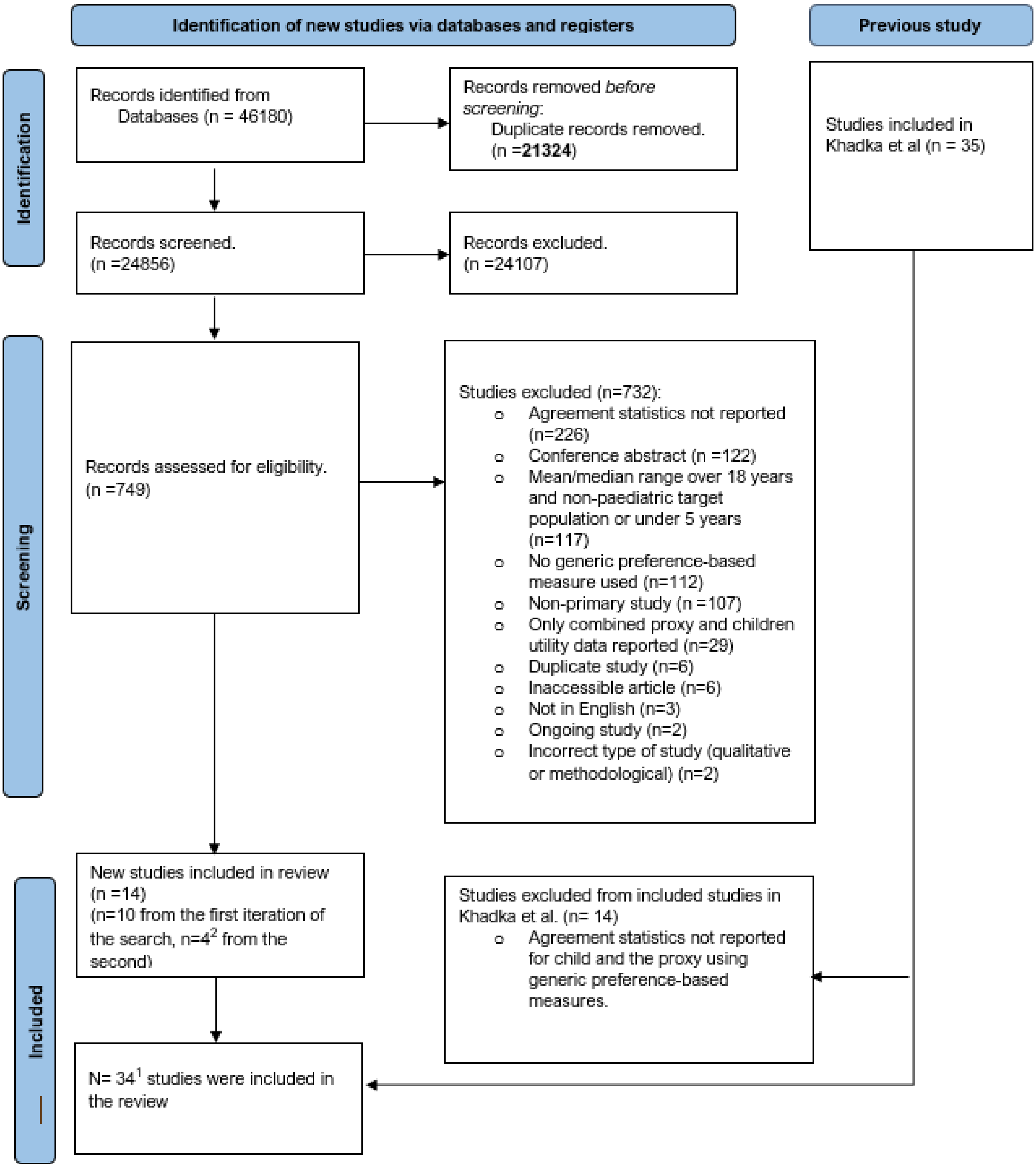


Figure 2.1 Literature search flow diagram using Preferred Reporting of Items for Systematic review and Meta-analysis (PRISMA).

---

[2] Six new studies were identified in the second iteration of the search, including two studies published as part of this PhD project by Khanna et al., ‘Parent-proxy and child self-reported health-related quality of life: using qualitative methods to explain the discordance (2024)’ and ‘Child-Parent Agreement in the Assessment of Health-Related Quality of Life Using the CHU9D and the PedsQL(TM) (2023)’ which were excluded from this systematic review.

2.5.2 Main characteristics of the studies

Table 2.1 presents an overview of the studies included in this systematic review. All the studies appraised for quality of reporting were of a high quality, scoring 0.7 and over. The following study designs were employed: cross-sectional (77%), longitudinal (20%), and case-control (3%). In all the studies, HRQoL measures used either independently or in combination with other measures included the HUI 3 (49%), EQ-5D measures (EQ-5D-Y-3L (40%), EQ-5D-Y-5L (6%), EQ-5D-3L (3%), and the EQ VAS) (49%), HUI 2 (29%), CHU9D (8%), and the QWB scale (3%). Cancer or history of cancer was the most common condition for which HRQoL was assessed (29%), predominantly blood and brain malignancies. Some studies (23%) also included children from the general population as the target sample or as the comparator/control group. The proxy respondent was exclusively a parent (mother, father, or a caregiver) in most of the identified studies (86%). Several studies (14%) used health professionals (nurses, physicians, and physiotherapists) or teachers as proxies, together with parents. The only exception was the study by Barr et al., which used only health professionals (nurses and physicians) for proxy-reported HRQoL using HUI 2 and 3 in cancer survivors [171]. Each study administered the proxy version of the measures adopting a proxy-proxy perspective, except one [172], which used a proxy-patient perspective (asking the proxy to rate the child's HRQoL from the child's perspective) for the EQ-5D-Y-3L.The measures were either administered by a trained interviewer (49%) or self-completed by the children (51%).

The majority of the studies (83%) reported the inter-rater agreement for overall HRQoL. Six studies only reported the dimension level agreement [147, 172-176]. When reported, ICCs were slightly more commonly represented (46%) than correlation coefficients in measuring the overall child-proxy agreement level. At the dimension level, Cohen's Kappa (43%) was the most frequently used measure of agreement, followed by ICC (14%), Gwet's AC_1 (14%) and Spearman's rho (14%).

Table 2-1 An overview of the included studies.

Description	Number of studies
Total studies included	35
Child-specific preference-based measures used:	
HUI 2	10
HUI 3	17
EQ-5D-Youth, EQ-5D and the EQ VAS	17
CHU9D	3
QWB	1
Health conditioned studied:	
Cancer or history of cancer	10
Other health conditions (including general health)	24
Child proxy pairs (with some studies using more than one proxy type)	
Child-Parent	33

Description	Number of studies
Child-Health professionals (Nurses, Physicians, Physiotherapists) or teachers	5
Self-mode of administration for child in the age range	
6-7 years	3
8 and above	14
Interviewer mode of administration for child in the age range	
6-7 years	8
8 and above	9
Level of agreement statistics reported	
For overall utilities	28
For attribute level utilities	19

HUI 2 = Health Utilities Index (HUI) Mark 2, HUI 3= Health Utilities Index (HUI) Mark 3, VAS= Visual Analogue Scale, CHU9D= Child Health Utility 9 Dimensions, QWB= Quality of Well-Being scale

Table 2-2 Details of the cancer studies which reported dyad self and proxy HRQoL using preference-based measures.

AUTHOR, COUNTRY, YEAR	HEALTH STATE EXPERIENCED	MEAN/MEDIAN AGE (RANGE) OF THE CHILD IN THE DYAD (WHERE AVAILABLE) OR STUDY	CHILD SAMPLE (MALE %) INCLUDED IN THE DYAD (WHERE AVAILABLE) OR STUDY	PROXY TYPE (N)	MEASURE	ADMINISTRATION MODE CHILD	QUALITY SCORE
Barr et al. Canada 1999 [171]	Cancer survivors: CNS tumours	13.5 (9.5-17.9)	15 (46.3)	Nurses (15), Physicians (12)	HUI 2/3	Self-administered	0.7
Glaser et al. UK 1999 [177, 178]	Cancer survivors: CNS tumours	10.5 (6.0-16.0)	28	Physiotherapist (30), Parents (30), Physicians (27)	HUI 2/3	Self and Interviewer administered	0.85
Sung et al. Canada 2004 [179]	Cancer patients, Rheumatic diseases, haemophilia, Conditions requiring Bone marrow transplant	13.7 (12.0-18.0)	22 (55)	Parents	HUI 2	Self-administered	0.9
Fu et al. El Salvador, Honduras,	Cancer Survivors: Leukemia, Lymphoma, Renal tumours, Germ cell tumours, Retinoblastoma,	12.8 (5-25.8)	211 (52.6)	Parents (180), Physicians (201)	HUI 2/3	Interviewer administered	0.8

AUTHOR, COUNTRY, YEAR	HEALTH STATE EXPERIENCED	MEAN/MEDIAN AGE (RANGE) OF THE CHILD IN THE DYAD (WHERE AVAILABLE) OR STUDY	CHILD SAMPLE (MALE %) INCLUDED IN THE DYAD (WHERE AVAILABLE) OR STUDY	PROXY TYPE (N)	MEASURE	ADMINISTRATION MODE CHILD	QUALITY SCORE
Nicaragua, Panama 2006 [180]	Malignant bone tumours, CNS tumours, Sympathetic nervous system tumours, soft tissue sarcomas, Carcinomas, Others						
Banks et al. Canada 2008 [181]	Cancer including leukemia, lymphoma, and brain tumour	9.5 (10.0-18.0)	11 (65)	Parents (22)	HUI 2/3	Self-administered	0.85
Fluchel et al. Uruguay 2008 [182]	Cancer survivors: ALL, brain tumours, Wilms tumour, retinoblastoma, Hodgkin disease, non-Hodgkin lymphoma, acute myeloid leukemia, rhabdomyosarcoma, neuroblastoma, Ewing sarcoma, ovarian sarcoma, osteogenic sarcoma	13.6 (7.0-28.0)	95 (49.5)	Parents (95)	HUI 3	Interviewer administered	0.95

AUTHOR, COUNTRY, YEAR	HEALTH STATE EXPERIENCED	MEAN/MEDIAN AGE (RANGE) OF THE CHILD IN THE DYAD (WHERE AVAILABLE) OR STUDY	CHILD SAMPLE (MALE %) INCLUDED IN THE DYAD (WHERE AVAILABLE) OR STUDY	PROXY TYPE (N)	MEASURE	ADMINISTRATION MODE CHILD	QUALITY SCORE
	General health (Control)	12.2 (8.0-17.0)	96 (33.3)	Parents (91)	HUI 3	Interviewer administered	
Penn et al. UK 2011 [183]	Cancer patients: brain tumour	12.4 (8.0-17.6)	29 (48.3)	Parents (29)	HUI 3	Interviewer administered	0.85
	General health (Control)	10.7 (8.0-18.9)	32 (50)	Parents (32)	HUI 3	Interviewer administered	
Zhou et al. China 2021 [144]	Haematological malignancies	10.5 (8.0 -17.0)	96 (64.6)	Caregiver (96)	EQ-5D-3L-Y/ EQ VAS, EQ-5D-5L-Y	Interviewer administered	0.95
Hetherington et al. Australia 2022 [184]	Childhood cancers	14.8 (12.0-17.0)	9 (39)	Parents (23)	EQ-5D-Y-3L/ EQ VAS	Self-administered	0.95

HUI 2 = Health Utilities Index (HUI) Mark 2, HUI 3= Health Utilities Index (HUI) Mark 3, VAS= Visual Analogue Scale, CHU9D= Child Health Utility 9 Dimensions, QWB= Quality of Well-Being scale,

CNS= Central Nervous System, ALL= Acute lymphoblastic leukaemia.

Table 2-3 Details of the studies with health conditions other than cancer which reported dyad self and proxy HRQoL using preference-based measures.

AUTHOR, COUNTRY, YEAR	HEALTH STATE EXPERIENCED	MEAN/MEDIAN AGE (RANGE) OF THE CHILD IN THE DYAD (WHERE AVAILABLE) OR STUDY	CHILD SAMPLE (MALE %) INCLUDED IN THE DYAD (WHERE AVAILABLE) OR STUDY	PROXY TYPE (N)	MEASURE (PROXY MEASURE)	ADMINISTRATION MODE CHILD	QUALITY SCORE
Czyzewski et al. USA 1994 [185]	Cystic Fibrosis	(12-17.9)	55	Parents (199)	QWB	Self-administered	0.8
Verrips et al. Netherlands 2001 [173]	Very Low Birth Weight (VLBW): Mail	14.2 (14.0)	486 (49)	Parents (481)	HUI 3	Self-administered	0.85
	Telephone	14.3 (14.0)	100 (54)	Parents (100)	HUI 3	Self-administered	
	Face-to-face	14.3 (14.0)	103 (51)	Parents (103)	HUI 3	Self-administered	
	Repeat mail	14.2 (14.0)	203 (52)	Parents (203)	HUI 3	Self-administered	
Brunner et al. Canada 2003 [186]	Musculoskeletal Disorders	9 (8.0-18.0)	55	Parents (68)	HUI 3	Interviewer administered	0.8

AUTHOR, **COUNTRY,** **YEAR**	**HEALTH STATE EXPERIENCED**	**MEAN/MEDIAN AGE (RANGE) OF THE CHILD IN THE DYAD (WHERE AVAILABLE) OR STUDY**	**CHILD SAMPLE (MALE %) INCLUDED IN THE DYAD (WHERE AVAILABLE) OR STUDY**	**PROXY TYPE (N)**	**MEASURE (PROXY MEASURE)**	**ADMINISTRATION MODE CHILD**	**QUALITY SCORE**
Jelsma and Ramma South Africa 2010 [172]	Children with functional impairment	(7.0-12.0)	61 (74)	Mother (57)	EQ-5D-Y/ VAS (EQ-5D-Y Proxy 2)	Self-administered	0.85
	General health (Control)	(7.0-12.0)	567 (45)	Mother (530)	EQ-5D-Y/ VAS	Self-administered	
Belfort et al. Germany 2016 [187]	Overweight or obese	10.3 (8.0-17.0)	76 (52.6)	Parents (63)	HUI 3	Interviewer administered	0.95
	General health (Control)	11.5 (8.0-18.0)					
Lee et al. USA 2011 [188]	Type 1 Diabetes Mellitus; Complications: Hypertension, Hypercholesterolemia, Cardiovascular Disease, Renal disease, Neurological disease, Retinopathy	13.7 (8.0-18.0)	231 (48.5)	Parents (223)	HUI 3	Interviewer administered	0.95
Morrow et al. Australia 2012 [174]	Chronic illness: any cancer, cystic fibrosis, type 1 diabetes, cerebral palsy [Gross Motor Function Classification System	12.2 (12.0-18.0)	69 (54.2)	Parents (129), Physicians (34)	HUI 2/3	Self-administered	0.85

AUTHOR, COUNTRY, YEAR	HEALTH STATE EXPERIENCED	MEAN/MEDIAN AGE (RANGE) OF THE CHILD IN THE DYAD (WHERE AVAILABLE) OR STUDY	CHILD SAMPLE (MALE %) INCLUDED IN THE DYAD (WHERE AVAILABLE) OR STUDY	PROXY TYPE (N)	MEASURE (PROXY MEASURE)	ADMINISTRATION MODE CHILD	QUALITY SCORE
	(GMFCS) V], any chronic neurological condition, liver transplant, inflammatory bowel disease, chronic kidney disease, autism.						
Rhodes et al. USA 2012 [189]	Obesity; Type 2 Diabetes Mellitus; prediabetes; insulin resistance	15.5 (12.0-18.0)	108	Parents (108)	HUI 3	Interviewer administered	0.85
Ungar et al. Canada 2012 [190]	Asthma	10.9 (8.0-17.0)	91 (55)	Parents (91)	HUI 2/3	Interviewer administered	1
Kulpeng et al. Thailand 2013 [191]	Common pneumococcal infections and sequelae: meningitis, bacteraemia, pneumonia, acute otitis media (AOM), hearing loss, chronic lung disease, epilepsy, mild mental retardation (MMR), severe mental retardation (SMR), and	10 (7.0-14.0)	74	Caregiver (74)	HUI 2/3, EQ-5D	Interviewer administered	0.85

AUTHOR, COUNTRY, YEAR	HEALTH STATE EXPERIENCED	MEAN/MEDIAN AGE (RANGE) OF THE CHILD IN THE DYAD (WHERE AVAILABLE) OR STUDY	CHILD SAMPLE (MALE %) INCLUDED IN THE DYAD (WHERE AVAILABLE) OR STUDY	PROXY TYPE (N)	MEASURE (PROXY MEASURE)	ADMINISTRATION MODE CHILD	QUALITY SCORE
	mental retardation (MR) combined with epilepsy.						
Wolke et al. Germany 2013 [175]	Very Low Birth Weight (VLBW)/ Very Preterm (VP)	13 (13.0)	260 (52)	Parents (260)	HUI 3	Self-administered	0.85
	General health (Control)	13 (13.0)	282 (49)	Parents (282)	HUI 3	Self-administered	
Gusi et al. Spain 2014 [147]	General health	(6.0-17.0)	442	Mother (442), Father (266)	EQ-5D-Y	Self and Interviewer administered	0.9
Sims-Williams Uganda 2017 [192]	Open Spina bifida; Associated complications	(10.0-14.0)	66 (56)	Caregiver (66)	HUI 3	Interviewer administered	0.95
Bharij et al. USA 2017 [193]	Paediatric liver transplant recipients	13.6 (12.0-21.7)	108 (44.4)	Parents (108)	HUI 2/3, CHU9D	Interviewer administered	0.9
Bray et al.	Long term mobility impairment: Cerebral	(6.0-18.0)	13 (61.5)	Parents (13)	HUI 2/3, EQ-5D-Y/ VAS	Self-administered	0.9

AUTHOR, COUNTRY, YEAR	HEALTH STATE EXPERIENCED	MEAN/MEDIAN AGE (RANGE) OF THE CHILD IN THE DYAD (WHERE AVAILABLE) OR STUDY	CHILD SAMPLE (MALE %) INCLUDED IN THE DYAD (WHERE AVAILABLE) OR STUDY	PROXY TYPE (N)	MEASURE (PROXY MEASURE)	ADMINISTRATION MODE CHILD	QUALITY SCORE
UK 2017 [148]	palsy, hemiplegia, muscular dystrophy						
Perez Sousa et al. Spain 2017 [194]	Cerebral palsy	10.9 (6.0-17.0)	62 (65.4)	Mother (62)	EQ-5D-Y/ VAS	Interviewer administered	0.95
Perez Sousa et al. (b) Spain 2018 [195]	Obesity: Exercise	9.6 (6.0-14.0)	106 (55)	Parents (106)	EQ-5D-Y/ VAS	Interviewer administered	0.95
	Obesity: Control	8.7 (6.0-13.0)	45 (47)	Parents (45)	EQ-5D-Y/ VAS	Interviewer administered	
van Summeren et al. Netherlands 2018 [196]	Functional constipation	10 (8.0-18.0)	56 (43)	Parents (56)	EQ VAS	Self-administered	0.95
Rogers et al. Netherlands 2019 [197]	Dental caries	11 (11.0)	486 (48)	Parents (486)	CHU9D (NL)	Self-administered	1

AUTHOR, COUNTRY, YEAR	HEALTH STATE EXPERIENCED	MEAN/MEDIAN AGE (RANGE) OF THE CHILD IN THE DYAD (WHERE AVAILABLE) OR STUDY	CHILD SAMPLE (MALE %) INCLUDED IN THE DYAD (WHERE AVAILABLE) OR STUDY	PROXY TYPE (N)	MEASURE (PROXY MEASURE)	ADMINISTRATION MODE CHILD	QUALITY SCORE
Shiroiwa et al. Japan 2019 [145]	General health	11 (8.0-15.0)	654 (50)	Parents (654)	EQ-5D-Y/ VAS	Self-administered	0.9
Sinlapamongkolkul et al. Thailand 2020 [198]	Thalassemia	9.1 (8.0-18.0)	85 (54)	Parents (85)	EQ VAS	Self and Interviewer administered	0.95
Lin et al. Hong Kong 2020	Adolescent idiopathic scoliosis (AIS)/ Juvenile idiopathic scoliosis (JIS)	14 (10.0-12.0)	125 (9.4)	Caregivers (125)	EQ-5D-Y-3L/ EQ VAS	Self-administered	0.95
Ralph et al. UK 2022 [199]	Congenital cytomegalovirus	10 (5.0-18.0)	8 (40)	Parents (20)	CHU9D, EQ-5D-Y-3L/ EQ VAS	Self-administered	0.9
Abraham et al. UK	Conduct disorders	7.6 (5.0-10.0)		Caregivers (21)	EQ-5D-Y-3L/ EQ VAS	Interviewer administered	0.85

AUTHOR, COUNTRY, YEAR	HEALTH STATE EXPERIENCED	MEAN/MEDIAN AGE (RANGE) OF THE CHILD IN THE DYAD (WHERE AVAILABLE) OR STUDY	CHILD SAMPLE (MALE %) INCLUDED IN THE DYAD (WHERE AVAILABLE) OR STUDY	PROXY TYPE (N)	MEASURE (PROXY MEASURE)	ADMINISTRATION MODE CHILD	QUALITY SCORE
2022 [200]							
Fitriana et al. Indonesia 2022 [176]	Major beta-thalassemia, Acute Lymphoblastic Leukemia, Haemophilia, and acute illness	11.2 (8.0-16.0)		Caregivers (286)	EQ-5D-Y-3L, EQ-5D-Y-5L, EQ VAS	Self-administered	0.95

HUI 2 = Health Utilities Index (HUI) Mark 2, HUI 3= Health Utilities Index (HUI) Mark 3, EQ-5D-Y= EQ-5D Youth version, VAS= Visual Analogue Scale, CHU9D= Child Health Utility 9 Dimensions,

CHU9D (NL)= CHU9D Dutch version, QWB= Quality of Well-Being scale

A summary of the included studies is presented in Tables 2.2 and 2.3 grouped into cancer and non-cancer related conditions, respectively. All the included studies were published between 1994 and 2022 and used primary data to obtain child health state utilities by employing generic preference-based measures. The majority of the studies were published in North America (USA and Canada) (32%) and Europe (UK, Spain, Netherlands, and Germany) (32%), followed by Asia (Thailand, Japan, Hong Kong, and China) (16%), Australia (8%), Africa (Uganda and South Africa) (5%) and Central and South America (including El Salvador, Honduras, Nicaragua, Panama, and Uruguay) (5%). There were 94 unique dyad samples included in the studies, with a total pooled sample of 4929 children and 5156 proxies. The age range for children in the included studies was between 5 and 18 years. 12 studies reported children younger than 8 years of age completing a self-report questionnaire either independently or with some assistance [147, 148, 172, 177, 178, 180, 182, 191, 193-195, 199, 200].

2.5.3 Proxy-child agreement

Table 2.4 presents a summary of reported agreement statistics for overall HRQoL using ICCs or correlation coefficients, i.e., Pearson's *r* and Spearman's rho. The studies used all the identified measures except for the EQ-5D-Y-5L (only dimension level agreement reported with this measure) and employed both caregivers and health professionals as proxies. The sample size of the dyad ranged from 11 [148] to 654 [198]. From a total of 29 studies (64 dyads), 13 studies reported only the ICCs [144, 145, 171, 180-182, 190, 193-196, 199, 201], one reported the CCC [184] and three studies reported ICCs alongside the correlation coefficients [177-179, 186]. Six studies reported only Spearman's rho [148, 183, 187-189, 198] whereas five studies reported only Pearson's *r* [185, 191, 192, 197, 200]. Details of the included studies reporting the dimension level agreement statistics is presented in Table 2.5. The dimension level agreement was reported for 21 studies (27 dyads), of which 11 studies used Cohen's Kappa [145, 147, 171-175, 177, 178, 184, 194, 195], three studies used ICC

[180, 182, 190], three used Gwet's AC_1 [144, 176, 201] and four used either Spearman's rho [183, 187, 189] or Pearson's [200] correlation coefficients. No study reported the dimension level agreement for the CHU9D and QWB measures.

Table 2-4 Details of the included studies of level of agreement by overall HRQoL between self and proxy reports using preference-based measures.

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	SAMPLE SIZE DYAD	CORRELATION TEST	CORRELATION COEFFICIENT (P-VALUE)	95% CONFIDENCE INTERVAL
Barr et al.	HUI 2	Nurses	15	ICC	0.85	
		Physicians	12		0.95	
Glaser et al.	HUI 2	Physiotherapist	25	ICC	0.4	
		Parents	24		0.57	
		Physicians	19		0.15	
Glaser et al. (b)	HUI 2	Physiotherapist	25	Pearson	0.54 (<0.01)	
		Parents	24		0.59 (<0.01)	
		Physicians	19		0.37 (0.12)	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	SAMPLE SIZE DYAD	CORRELATION TEST	CORRELATION COEFFICIENT (P-VALUE)	95% CONFIDENCE INTERVAL
Sung et al.	HUI 2	Parents	19	ICC	0.11 (0.3)	-0.35, 0.53
				Spearman	0.14	-0.34, 0.55
	HUI 3	Parents	19	ICC	-0.01	-0.45, 0.44
				Spearman	0.11	0.35, 0.55
Fu et al.	HUI 2	Parents	120	ICC	0.389	0.227, 0.531
		Physicians	156		0.379	0.237, 0.506
	HUI 3	Parents	156	ICC	0.433	0.297, 0.552
		Physicians	166		0.341	0.200, 0.469
Banks et al.	HUI 2	Parents	11	ICC	0.74	0.29, 0.92
	HUI 3	Parents	11	ICC	0.42	-0.21, 0.80

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	SAMPLE SIZE DYAD	CORRELATION TEST	CORRELATION COEFFICIENT (P-VALUE)	95% CONFIDENCE INTERVAL
Fluchel et al.	HUI 3	Parents	92	ICC	0.3087	0.1125, 0.4818
		Physicians	91		0.066	-0.1402, 0.2669
Fluchel et al. (Control)	HUI 3	Physicians/ Teachers	89	ICC	-0.3103	-0.4857, -0.1106
Penn et al.	HUI 3	Parents	21	Spearman	0.76 (<0.001)	
Penn et al. (Control)	HUI 3	Parents	22	Spearman	0.31	
Zhou et al. (Baseline)	EQ VAS	Caregiver	96	ICC	0.22	
Zhou et al. (Follow-up)	EQ VAS	Caregiver	96	ICC	0.556	

AUTHORS **(INTERVENTION)**	**MEASURE**	**PROXY TYPE**	**SAMPLE SIZE DYAD**	**CORRELATION TEST**	**CORRELATION COEFFICIENT** **(P-VALUE)**	**95% CONFIDENCE INTERVAL**
Czyzewski et al.	QWB	Parents	55	Pearson	0.39	
Brunner et al.	HUI 3	Parents	45	ICC	0.43	
				Pearson	0.57	
Belfort et al. (Overall)	HUI 3	Parents	63	Spearman	0.47 (0.0002)	
Lee et al.	HUI 3	Parents	223	Spearman	0.34	0.22, 0.45
Rhodes et al.	HUI 3	Parents	96	Spearman	0.24 (<0.05)	
Ungar et al.	HUI 2	Parents	72	ICC	0.021	-0.22, 0.262
	HUI 3	Parents	75	ICC	0.169	-0.070, 0.389

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	SAMPLE SIZE DYAD	CORRELATION TEST	CORRELATION COEFFICIENT (P-VALUE)	95% CONFIDENCE INTERVAL
Ungar et al. (Dyad)	HUI 2	Parent with child	72	ICC	0.545 (<0.0001)	0.360, 0.689
	HUI 3	Parent with child	75	ICC	0.735 (<0.0001)	0.611, 0.824
Kulpeng et al.	HUI 2	Caregiver	74	Pearson	0.58 (<0.05)	
	HUI 3	Caregiver	74	Pearson	0.67 (<0.05)	
	EQ-5D	Caregiver	74	Pearson	0.77 (<0.05)	
	EQ VAS	Caregiver	74	Pearson	0.5 (<0.05)	
Sims-Williams et al.	HUI 3	Caregiver	62	Pearson	0.848	
Bharij et al.	HUI 2	Parents	61	ICC	0.9 (<0.001)	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	SAMPLE SIZE DYAD	CORRELATION TEST	CORRELATION COEFFICIENT (P-VALUE)	95% CONFIDENCE INTERVAL
	HUI 3	Parents	60	ICC	0.75 (<0.001)	
	CHU9D	Parents	96	ICC	0.69 (<0.001)	
Bray et al.	HUI 2	Parents	13	Spearman	0.728 (0.005)	
	HUI 3	Parents	13	Spearman	0.842 (<0.001)	
	EQ-5D-Y	Parents	11	Spearman	0.665 (0.026)	
	EQ VAS	Parents	13	Spearman	0.545 (0.054)	
Perez Sousa et al.	EQ VAS	Mother	62	ICC	0.389 (0.029)	
		Father	62		0.581 (0.962)	
Perez Sousa et al. (b)	EQ VAS	Parents	151	ICC	0.5 (<0.0001)	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	SAMPLE SIZE DYAD	CORRELATION TEST	CORRELATION COEFFICIENT (P-VALUE)	95% CONFIDENCE INTERVAL
(Overall: Baseline)						
Perez Sousa et al. (b)	EQ VAS	Parents	151	ICC	0.7 (<0.0001)	
(Overall: Follow-up: Post treatment)						
van Summeren et al.	EQ VAS	Parents	56	ICC	0.78	0.65, 0.87
Rogers et al.	CHU9D	Parents	184	Pearson	0.156 (0.02)	
Rogers et al. (Control)	CHU9D	Parents	302	Pearson	0.183 (0.01)	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	SAMPLE SIZE DYAD	CORRELATION TEST	CORRELATION COEFFICIENT (P-VALUE)	95% CONFIDENCE INTERVAL
Rogers et al. (Overall)	CHU9D	Parents	386	Pearson	0.183 (<0.001)	
Shiroiwa et al.	EQ VAS	Parents	654	ICC	0.06	
Sinlapamongkolkul et al.	EQ VAS	Caregiver	85	Spearman	0.334 (0.001)	
Lin et al. (Overall)	EQ VAS	Caregiver	125	ICC	0.29	
Ralph et al.	CHU9D	Parents	20	ICC	0.8	0.44, 0.93
Abraham et al.	EQ VAS	Caregivers	11	Pearson	0.94	
Hetherington et al.	EQ VAS	Mothers	18	CCC	0.61	0.26, 0.81
Hetherington et al.	EQ VAS	Fathers	11	CCC	0.32	-0.15, 0.67

HUI 2 = Health Utilities Index (HUI) Mark 2, HUI 3= Health Utilities Index (HUI) Mark 3, EQ-5D-Y= EQ-5D Youth version, VAS= Visual Analogue Scale, CHU9D= Child Health Utility 9 Dimensions, QWB= Quality of Well-Being scale, ICC= Intraclass Correlation Coefficient, CCC= Concordance Correlation Coefficient; Altman's scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good.

Table 2-5 Details of the included studies of level of agreement by dimension between self and proxy reports using preference-based measures.

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
Barr et al.	HUI 2	Nurses	Sensation	Cohen's Kappa	0.05	
			Emotion		0.13	
			Cognition		0.54	
			Pain		0.71	
	HUI 2	Physicians	Sensation	Cohen's Kappa	0.42	
			Emotion		0.13	
			Cognition		0.37	
			Pain		0.73	
Fu et al.	HUI 2	Parents	Sensation	ICC	0.773	0.706, 0.826
			Mobility		0.67	0.584, 0.742

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Emotion		0.104	-0.058, 0.262
			Cognition		0.121	-0.026, 0.263
			Self-care		0.422	0.298, 0.532
			Pain		0.14	-0.002, 0.277
	HUI 2	Physicians	Sensation	ICC	0.829	0.778, 0.870
			Mobility		0.569	0.465, 0.657
			Emotion		0	-0.143, 0.143
			Cognition		0.102	-0.045, 0.245
			Self-care		0.754	0.686, 0.810
			Pain		0.08	-0.063, 0.219
Fluchel et al.: (Cancer survivors)	HUI 3	Parents	Vision	ICC	1.00	1, 1
			Hearing		0.77	0.68, 0.84
			Speech		0.56	0.41, 0.69
			Ambulation		0.60	0.46, 0.72
			Dexterity		1.00	0, 0

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Emotion		-0.07	-0.27, 0.13
			Cognition		0.31	0.12, 0.48
			Pain		0.41	0.22, 0.56
Fluchel et al.: (Cancer survivors)	HUI 3	Physicians	Vision	ICC	0.81	0.73, 0.87
			Hearing		0.56	0.4, 0.68
			Speech		0.28	0.08, 0.46
			Ambulation		0.43	0.25, 0.58
			Dexterity		0.42	0.24, 0.58
			Emotion		-0.25	-0.43, -0.05
			Cognition		0.06	-0.14, 0.26
			Pain		0.12	-0.09, 0.31
Fluchel et al.: (General health)	HUI 3	Physicians/ Teachers	Vision	ICC	0.26	0.06, 0.44
			Hearing		1.00	0, 0
			Speech		-0.03	-0.23, 0.18
			Ambulation		1.00	0, 0

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Dexterity		1.00	0, 0
			Emotion		-0.30	-0.47, -0.1
			Cognition		0.00	-0.21, 0.21
			Pain		-0.08	-0.28, 0.13
Morrow et al.	HUI 2	Parents	Sensation	Cohen's Kappa	0.51	0.23, 0.78
			Mobility		0.59	0.31, 0.86
			Emotion		0.32	0.10, 0.53
			Cognition		0.29	0.35, 0.54
			Pain		0.44	0.23, 0.64
	HUI 2	Physicians	Sensation	Cohen's Kappa	0.27	-0.26, 0.56
			Mobility		0.62	0.37, 0.88
			Emotion		0.18	-0.03, 0.88
			Cognition		0.07	-0.16, 0.30
			Pain		0.11	-0.11, 0.34
	HUI 3	Parents	Ambulation	Cohen's Kappa	0.52	0.29, 0.77
			Dexterity		0.12	-0.11, 0.34

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Emotion		0.27	0.04, 0.51
			Cognition		0.32	0.09, 0.55
			Pain		0.43	0.25, 0.62
	HUI 3	Physicians	Ambulation	Cohen's Kappa	0.56	0.31, 0.82
			Dexterity		0.11	-0.12, 0.33
			Emotion		0.16	-0.05, 0.37
			Cognition		0.05	-0.11, 0.20
			Pain		0.36	0.17, 0.55
Glaser et al.	HUI 2	Physiotherapist	Sensation	Cohen's Kappa	0.32	
			Mobility		NS	
			Emotion		0.37	
			Cognition		0.7	
			Self-care		0.43	
			Pain		NS	
	HUI 2	Parents	Sensation	Cohen's Kappa	0.54	
			Mobility		0.72	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Emotion		0.37	
			Cognition		NS	
			Self-care		0.47	
			Pain		0.62	
	HUI 2	Physicians	Sensation	Cohen's Kappa	0.38	
			Mobility		0.77	
			Emotion		NS	
			Cognition		NS	
			Self-care		0.78	
			Pain		NS	
Glaser et al. (b)	HUI 3	Physiotherapist	Vision	Cohen's Kappa	0.62	
			Hearing		0.12	
			Speech		0.64	
			Ambulation		0.19	
			Dexterity		0.77	
			Emotion		0.4	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Pain		0.33	
	HUI 3	Parents	Vision	Cohen's Kappa	0.62	
			Hearing		0.49	
			Speech		0.47	
			Ambulation		0.73	
			Dexterity		0.82	
			Emotion		0.28	
			Pain		0.56	
	HUI 3	Physicians	Vision	Cohen's Kappa	0.6	
			Hearing		0.67	
			Speech		0.14	
			Ambulation		0.77	
			Dexterity		0.48	
			Emotion		0.14	
			Pain		0.14	
Ungar et al.	HUI 2	Parents	Mobility	ICC	0.108	-0.101, 0.308

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Emotion		0.065	-0.155, 0.278
	HUI 2	Parent with child	Mobility	ICC	0.713	0.593, 0.802
			Emotion		0.468	0.281, 0.621
Verrips et al.: Mail	HUI 3	Parents	Vision	Cohen's Kappa	0.87	
			Hearing		0.33	
			Speech		0.23	
			Ambulation		0.66	
			Dexterity		0.63	
			Emotion		0.29	
			Cognition		0.36	
			Pain		0.43	
Verrips et al.: Telephone	HUI 3	Parents	Vision	Cohen's Kappa	0.69	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Speech		0.21	
			Ambulation		0.73	
			Dexterity		0.61	
			Emotion		0.2	
			Cognition		0.17	
			Pain		0.22	
Verrips et al.: Face-to-face	HUI 3	Parents	Vision	Cohen's Kappa	0.75	
			Hearing		0	
			Speech		0.19	
			Ambulation		0.39	
			Dexterity		0.8	
			Emotion		0.07	
			Cognition		0.09	
			Pain		0.08	
Wolke et al.	HUI 3	Parents	Vision	Cohen's Kappa	0.87	0.88, 0.86

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Hearing		0.59	0.59, 0.59
			Speech		0.22	0.22, 0.22
			Ambulation		0.78	0.78, 0.78
			Dexterity		0.67	0.68, 0.66
			Emotion		0.41	0.42, 0.4
			Cognition		0.32	0.32, 0.32
			Pain		0.48	0.49, 0.47
Wolke et al.: General health (Control)	HUI 3	Parents	Vision	Cohen's Kappa	0.82	0.81, 0.83
			Hearing		1	0.99, 1.01
			Speech		0.23	0.23, 0.23
			Dexterity		0.67	0.66, 0.68
			Emotion		0.37	0.36, 0.38
			Cognition		0.2	0.2, 0.2
			Pain		0.46	0.45, 0.47
Gusi et al.			Pain or discomfort		0.68 (<0.05)	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Worried, sad or unhappy		0.221 (<0.05)	
Jelsma and Ramma	EQ-5D-Y	Mother	Mobility	Cohen's Kappa	0.15	
			Self-care		0.08	
			Doing usual. activities		0.01	
			Pain or discomfort		0.2	
			Worried, sad or unhappy		0.21	
Jelsma and Ramma: General health (Control)	EQ-5D-Y	Mother	Mobility	Cohen's Kappa	0.6	
			Self-care		0.33	
			Doing usual activities		0.34	
			Pain or discomfort		0.41	
			Worried, sad or unhappy		0.22	
Rhodes et al.	HUI 3	Parents	Vision	Spearman	0.56 (<0.05)	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Dexterity		0.49 (<0.05)	
Penn et al.	HUI 3	Parents	Vision	Spearman	0.73 (<0.01)	
			Hearing		1.00	
			Speech		1.00	
			Ambulation		0.82 (<0.01)	
			Dexterity		0.73 (<0.01)	
			Emotion		0.30 (0.19)	
			Cognition		0.75 (<0.01)	
			Pain		0.20 (0.39)	
Belfort et al.	HUI 3	Parents	Emotion	Spearman	0.45 (<0.01)	
			Cognition		0.30 (<0.05)	
			Pain		0.14 (0.30)	
Perez Sousa et al.	EQ-5D-Y	Mother	Mobility	Cohen's Kappa	0.713 (<0.001)	
			Self-care		0.057 (0.536)	
			Doing usual activities		0.436 (<0.001)	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Pain or discomfort		0.128 (0.183)	
			Worried, sad or unhappy		0.165 (0.14)	
	EQ-5D-Y	Father	Mobility	Cohen's Kappa	0.042 (0.653)	
			Self-care		0.044 (0.622)	
			Doing usual activities		0.019 (0.841)	
			Pain or discomfort		0.067 (0.469)	
			Worried, sad or unhappy		0.016 (0.854)	
Perez Sousa et al. (b)	EQ-5D-Y	Parents	Mobility	Cohen's Kappa	0.51 (<0.001)	
			Self-care		0.36 (<0.001)	
			Doing usual activities		0.22 (<0.001)	
			Pain or discomfort		0.27 (<0.001)	
			Worried, sad or unhappy		0.42 (<0.001)	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
Perez Sousa (b): Control	EQ-5D-Y	Parents	Mobility	Cohen's Kappa	0.15 (0.03)	
			Self-care		0.13 (0.04)	
			Doing usual activities		0.09 (0.19)	
			Pain or discomfort		0.26 (<0.001)	
			Worried, sad or unhappy		0.37 (<0.001)	
Shiroiwa et al.	EQ-5D-Y	Parents	Mobility	Cohen's Kappa	0.5	
			Self-care		0.91	
			Doing usual activities		0.78	
			Pain or discomfort		0.15	
			Worried, sad or unhappy		0.12	
Hetherington et al.	EQ-5D-Y-3L	Mother	Mobility	Weighted Kappa	0.94	0.81, 1
			Self-care		0.70	0.39, 1
			Usual activities		0.68	0.29, 1

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Pain/discomfort		0.00	-0.56, 0.56
			Anxiety/depression		0.13	-0.45, 0.7
	EQ-5D-Y-3L	Father	Mobility	Weighted Kappa	0.65	0.23, 1
			Self-care		0.37	-0.04, 0.77
			Usual activities		0.25	-0.28, 0.79
			Pain/discomfort		0.42	-0.17, 1
			Anxiety/depression		0.26	-0.21, 0.72
Abraham et al. (Baseline)	EQ-5D-Y-3L	Parents	Mobility	Pearson		
			Self-care		0.43	
			Usual activities		0.09	
			Pain/discomfort		0.63	
			Anxiety/depression		0.31	
Abraham et al. (Follow-up)	EQ-5D-Y-3L	Parents	Mobility	Pearson		
			Self-care		0.54	
			Usual activities		0.23	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Pain/discomfort		0.76	
			Anxiety/depression		0.91	
Fitriana et al.: Thalassemia (Baseline)	EQ-5D-Y-3L	Parents	Mobility	Gwet's AC1	0.86	
			Self-care		0.87	
			Usual activities		0.72	
			Pain/discomfort		0.65	
			Anxiety/depression		0.80	
	EQ-5D-Y-5L	Parents	Mobility	Gwet's AC1	0.87	
			Self-care		0.95	
			Usual activities		0.81	
			Pain/discomfort		0.70	
			Anxiety/depression		0.77	
Fitriana et al.: Acute disease (Baseline)	EQ-5D-Y-3L	Parents	Mobility	Gwet's AC1	0.73	
			Self-care		0.30	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Usual activities		0.03	
			Pain/discomfort		0.57	
			Anxiety/depression		0.47	
	EQ-5D-Y-5L	Parents	Mobility	Gwet's AC1	0.84	
			Self-care		0.29	
			Usual activities		0.07	
			Pain/discomfort		0.53	
			Anxiety/depression		0.64	
Fitriana et al.: Acute lymphoblastic leukemia (Baseline)	EQ-5D-Y-3L	Parents	Mobility	Gwet's AC1	0.81	
			Self-care		0.87	
			Usual activities		0.66	
			Pain/discomfort		0.72	
			Anxiety/depression		0.67	
	EQ-5D-Y-5L	Parents	Mobility	Gwet's AC1	0.91	
			Self-care		0.77	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Usual activities		0.67	
			Pain/discomfort		0.81	
			Anxiety/depression		0.65	
Fitriana et al.: Haemophilia (Baseline)	EQ-5D-Y-3L	Parents	Mobility	Gwet's AC1	0.63	
			Self-care		0.82	
			Usual activities		0.66	
			Pain/discomfort		0.55	
			Anxiety/depression		0.58	
	EQ-5D-Y-5L	Parents	Mobility	Gwet's AC1	0.73	
			Self-care		0.89	
			Usual activities		0.76	
			Pain/discomfort		0.62	
			Anxiety/depression		0.62	
Fitriana et al.: Thalassemia (Follow-up)	EQ-5D-Y-3L	Parents	Mobility	Gwet's AC1	0.96	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Self-care		0.93	
			Usual activities		0.80	
			Pain/discomfort		0.80	
			Anxiety/depression		0.90	
	EQ-5D-Y-5L	Parents	Mobility	Gwet's AC1	0.95	
			Self-care		0.94	
			Usual activities		0.87	
			Pain/discomfort		0.79	
			Anxiety/depression		0.93	
Fitriana et al.: Acute disease (Follow-up)	EQ-5D-Y-3L	Parents	Mobility	Gwet's AC1	0.96	
			Self-care		0.97	
			Usual activities		0.93	
			Pain/discomfort		0.90	
			Anxiety/depression		0.82	
	EQ-5D-Y-5L	Parents	Mobility	Gwet's AC1	0.96	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
			Self-care		0.95	
			Usual activities		0.91	
			Pain/discomfort		0.93	
			Anxiety/depression		0.93	
Fitriana et al.: Acute lymphoblastic leukemia (Follow-up)	EQ-5D-Y-3L	Parents	Mobility	Gwet's AC1	0.95	
			Self-care		0.85	
			Usual activities		0.83	
			Pain/discomfort		0.73	
			Anxiety/depression		0.74	
	EQ-5D-Y-5L	Parents	Mobility	Gwet's AC1	0.97	
			Self-care		0.91	
			Usual activities		0.91	
			Pain/discomfort		0.76	
			Anxiety/depression		0.76	

AUTHORS (INTERVENTION)	MEASURE	PROXY TYPE	ATTRIBUTE	STATISTIC REPORTED	AGREEMENT STATISTIC (P-VALUE)	95% CONFIDENCE INTERVAL
Fitriana et al.: Haemophilia (Follow-up))	EQ-5D-Y-3L	Parents	Mobility	Gwet’s AC1	0.96	
			Self-care		0.98	
			Usual activities		0.87	
			Pain/discomfort		0.46	
			Anxiety/depression		0.84	
	EQ-5D-Y-5L	Parents	Mobility	Gwet’s AC1	0.96	
			Self-care		0.98	
			Usual activities		0.84	
			Pain/discomfort		0.79	
			Anxiety/depression		0.89	

HUI 2 = Health Utilities Index (HUI) Mark 2, HUI 3= Health Utilities Index (HUI) Mark 3, EQ-5D-Y= EQ-5D Youth version, VAS= Visual Analogue Scale, CHU9D= Child Health Utility 9 Dimensions, QWB= Quality of Well-Being scale, ICC= Intraclass Correlation Coefficient; Altman’s scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good.

2.5.4 Inter-rater agreement based on the type of measure

2.5.4.1 HUI 2 and 3

The inter-rater agreement between children and proxies for nine studies as indicated by the ICCs was low (ICC=<0.5) for overall HRQoL [171, 177-182, 186, 190, 193]. The overall ICC for HUI 2 was higher than that of HUI 3. In contrast to HUI 2, which showed good to very good agreement for the overall HRQoL for a quarter of the samples in the studies, the agreement using HUI 3 was moderate at best. The correlation coefficients obtained from ten studies indicated moderate associations between child-self and proxy reports [148, 177-179, 183, 186-189, 191, 192].

Across the HUI 2 attributes of “emotion”, “cognition” and “pain”, the overall agreement was the lowest. In comparison, the kappa values suggested good to moderate agreement for “mobility”, “self-care” and “sensation” [171, 174, 177, 178]. Similar levels of agreement were observed for HUI 3 attributes, where the kappa values indicated fair agreement for “cognition”, “emotion”, “pain” and for the “speech” attribute. The agreement was only moderate for “hearing”, but between good for “dexterity” and “ambulation”. The highest agreement (substantial/very good) was observed for “vision” [173-175, 177, 178].

The ICC values demonstrated a poor (ICC<0.2) agreement for subjective dimensions (“emotion”, “cognition”, and “pain”) of HUI 2/3, with some even reporting negative values. For the observable HUI 2/3 attributes, the agreement was generally between good to very good for the dimensions of “sensation”, “vision”, “self-care”, and “hearing”. However, relatively lower agreement (moderate) was observed for “speech”, “mobility”, “ambulation” and “dexterity” [180, 182, 190].

2.5.4.2 EQ-5D measures and the EQ VAS

None of the studies reported the ICCs for the overall utilities or the summary scores using EQ-5D measures. However, one study reported good association between the child-caregiver EQ-5D utilities [191]. Of the six studies reporting the ICCs [144, 145, 194-196, 201] and one reporting the CCC [184] for the EQ VAS scores, the majority showed low agreement (<0.5) between child-proxy dyads. However, an improvement in the inter-rater agreement was noted from baseline to follow-up [144, 195].

Kappa coefficients reported for six studies varied across the EQ-5D-Y-3L dimensions, ranging between fair to very good [145, 147, 172, 184, 194, 195]. As with the HUI 2/3 attributes, dimensions associated with psychosocial health− “feeling worried, sad or unhappy” and “having pain or discomfort”− demonstrated lower agreement varying from poor to moderate compared to the more observable dimensions of “walking about” and “looking after myself” where agreement was between good to very good. Lower agreement (poor to moderate) was also observed for the “doing usual activities” dimension, with instances of good agreement being less frequent.

The inter-rater agreement between children and proxies within the EQ-5D-Y dimensions reported using Gwet’s AC_1 ranged from good to very good [144, 176, 201]. The agreement for the “walking about” and “looking after myself” dimensions was consistently high (very good) for both the 3L and the 5L versions. Despite the lower agreement levels for the “feeling worried, sad or unhappy”, the “having pain or discomfort” and the “doing usual activities” dimensions, the agreement was still categorised as good. Importantly, no significant differences were observed in the agreement levels between the 3L and 5L versions, suggesting that the choice of version does not impact the agreement within the dimensions [144]. Moreover, the studies using longitudinal analysis reported an improvement in agreement from baseline to follow-up [144, 176].

2.5.4.3 CHU9D and QWB

Two studies reported the ICC using CHU9D which showed good to very good inter-rater agreement [193, 199]. However, in a large sample of child-parent dyads, Rogers et al. [197] reported statistically significant but almost no correlation (<0.2) between the child-self and proxy reports using CHU9D. In their study, Czyzewski et al. reported a moderate correlation between the self and proxy reported HRQoL using QWB [185].

2.5.5 Inter-rater agreement based on the type of proxy

Both types of proxies (parents and health professionals) showed moderate (0.4 to 0.6) inter-rater agreement and correlation, although parents showed slightly higher agreement overall, regardless of measures and/or health conditions. All studies using health professionals as proxies assessed the HRQoL of children with cancer or child cancer survivors. Among these, Fluchel and colleagues used physicians and teachers as proxies for the children in the control group with no health condition. A negative ICC (-0.31, 95% CI: -0.22 to 0.262) was noted indicating poor inter-rater agreement between these pairs [182]. Only one study showed good to excellent agreement between cancer survivors and health professionals (nurses and physicians) using HUI 2 [171]. Glaser and colleagues compared the inter-rater agreement between children with a history of cancer and their parents, physicians, and physiotherapists. Both the agreement (ICC) and correlation (Pearson's *r*) values were better for parents, closely followed by physiotherapists and worst for physicians [177, 178]. Among the parents, Perez-Souza et al. observed higher agreement levels in the EQ VAS assessments between father-child pairs compared to mother-child pairs [194]. The study by Ungar et al. reported a poor (approximately 0.2) inter-rater agreement when children and parents reported paediatric HRQoL separately using the HUI 2 and 3. However, the agreement was found to be significant and moderate (>0.5) using a consensus based dyad approach [190].

The agreement between children and physiotherapists was generally low with the exception of one study where physiotherapists reported higher agreement than parents and physicians within the HUI 3 attributes of "vision" and "speech" [177, 178]. Overall, physicians reported excellent agreement when assessing the functional attributes, e.g., "mobility" and "ambulation", whereas the subjective attributes of "emotion", "pain" and "cognition" lacked sufficient agreement [174, 177, 178, 180, 182]. The worst agreement was observed between physicians/teachers and children from the general population. Perfect agreement was observed within the "hearing", "ambulation", and "dexterity" attributes whilst the remaining attributes showed poor or no agreement [182].

Parents followed a similar suit and reported slight to fair agreement within the "emotion" and "cognition" attributes of HUI 2 and 3. In the assessment of "emotion", the only exception was reported in a study of children with very low birth weight by Wolke et al., which showed moderate agreement with the parents in the study population [175]. Between parent genders, mother-child pairs demonstrated a higher agreement as compared to father-child dyads [147, 172, 184, 194]. Moreover, father-child pairs generally reported only poor to moderate levels of agreement, except within the "walking about" dimension, where one study [184] indicated good agreement.

2.5.6 Inter-rater agreement based on the type of condition

Overall, the agreement/correlation between dyads within cancer-related studies was generally lower than studies with conditions other than cancer. However, no clear trend emerged in this analysis. For instance, in a longitudinal study of cancer patients, Penn and colleagues found strong associations between the HUI 3 derived overall HRQoL as reported by children and proxies in the study population but weak correlations for those in the control group [183]. Conditions like respiratory (asthma) and musculoskeletal diseases assessed using HUI 2 and 3 showed poor to moderate inter-rater agreement between child-self and proxy reported HRQoL [186, 190]. Using the EQ VAS, van Summeren and colleagues found good inter-rater agreement in children with functional constipation [196]. Additionally, in a longitudinal study of children with obesity, the agreement between children and parents for EQ VAS scores was found to be moderate at the baseline and follow-up [195]. Strong associations (Spearman's rho) were noted between the overall HRQoL reported by children with cerebral palsy, hemiplegia, and/or muscular dystrophy and their parents using both EQ-

5D-Y & EQ VAS [148] whilst the correlation between children with thalassemia and their caregivers with EQ VAS was weak [198]. Kulpeng et al also indicated a large correlation (Pearson's *r*) between self and proxy derived HRQoL using EQ-5D & EQ VAS in children with severe childhood infections [191]. Within the cancer-related studies, unsurprisingly, the agreement/correlations were worse for children with (active) cancer than those with a history of cancer/cancer survivors.

The agreement and the correlation between child-self and proxy reported HRQoL observed in the cohort of healthy children was, in general, low (<0.5) for the overall HRQoL [145, 182, 183, 187, 197] as well as for the dimension-level HRQoL [145, 147, 172, 175, 182, 187]. For the HUI 3 attributes, good to very good agreement between children from the general population and parents was only observed for "hearing", "vision" and "dexterity" [175]. Similarly, higher levels agreement was only observed in this cohort for the "looking after myself" and "doing usual activities" dimensions. In contrast, the less observable dimensions such as the "emotion", "cognition" and "pain" attributes of the HUI 3 and the "feeling sad, worried or unhappy" dimension of the EQ-5D-Y-3L demonstrated poor to fair agreement.

2.5.7 Meta-analysis results

In the following, results for the meta-analysis are provided for studies which reported the ICC (95% CI) for the overall utilities and Cohen's Kappa for the dimension level HRQoL. Ten studies were included in the analysis to estimate the ICC for overall utilities generated using the child-specific generic preference-based measures [171, 177-182, 186, 190, 193, 199]. Seven studies reported the ICCs for EQ VAS scores [144, 145, 184, 194-196, 201] and one reported only Pearson's *r* for the EQ-5D-3L utilities [191], and were, therefore, excluded. Kappa statistics for the dimension level agreement were reported for 10 studies employing HUI 2 and 3 (5 studies) [171, 173-175, 177, 178] and EQ-5D-Y (3L/5L) (5 studies) [145, 147, 172, 194, 195]. However, only one [145] of the five studies using EQ-5D-Y (3L/5L) reported the percent agreement values. Three studies reported the Gwet's AC_1 [144, 154, 176, 200] and one reported Pearson's *r* [200]. Consequently, the EQ-5D measures were excluded from the dimension level meta-analysis of agreement because standard errors could not be estimated for the former studies and the agreement statistics were not suitable for pooling for the others.

2.5.7.1 Inter-rater agreement for overall utilities

The overall ICC for all 25 samples for the HUI 2, HUI 3 and the CHU9D was 0.45 (0.36, 0.62). Figure 2.2 depicts the study-specific and overall estimates of ICC, their respective 95% CIs and the study weights (%). The test for homogeneity resulted in a Q test statistic of 203.13 (p<0.001). The heterogeneity in the studies was high (I^2=91%) due to the presence of high variability between studies.

Exploratory moderators such as the type of measure, health condition, proxy, age-group (below 8 years vs above) of the children in the sample were used to potentially explain this heterogeneity. The moderators were categorised according to the 1) type of measure used- HUI 2 (12 samples) or HUI 3 (11 samples) or CHU9D (2 samples), 2) health condition assessed- cancer (15 samples) or non-cancer (10 samples) related, 3) Type of proxy used - parent/caregiver (17 samples) or health professional/teacher (8 samples), and 4) the lower age limit of the sample- below 8 years (11 samples) or 8 and above years (14 samples).

HUI 3 had an estimated ICC of 0.37 (0.18, 0.53), much lower than HUI 2, which had an estimated ICC of 0.58 (0.34, 0.75). The overall ICC for cancer-related samples was 0.43 (0.27, 0.57) whereas for samples with conditions other than cancer, including general health, it was 0.57 (0.33, 0.74). The ICC estimate for parent proxies was 0.53 (0.37, 0.66) whereas for health professionals it was only marginally lower at 0.47 (0.11, 0.72). Samples that also included younger children had an ICC of 0.42 (0.29, 0.54) which was lower than the ICC of 0.57 (0.33, 0.74) with older children. However, none of the group differences were statistically significant and, therefore, did not suggest moderation by any of the included variables.

The results of the meta-regression showed that none of the explanatory variables were statistically significant, thus, showing no significant differences in child and proxy agreement according to the type of measure, health condition experienced, proxy type and the inclusion of children below 8 years in the sample. The funnel plot and the funnel-plot test for asymmetry (p=0.07) did not suggest any publication bias.

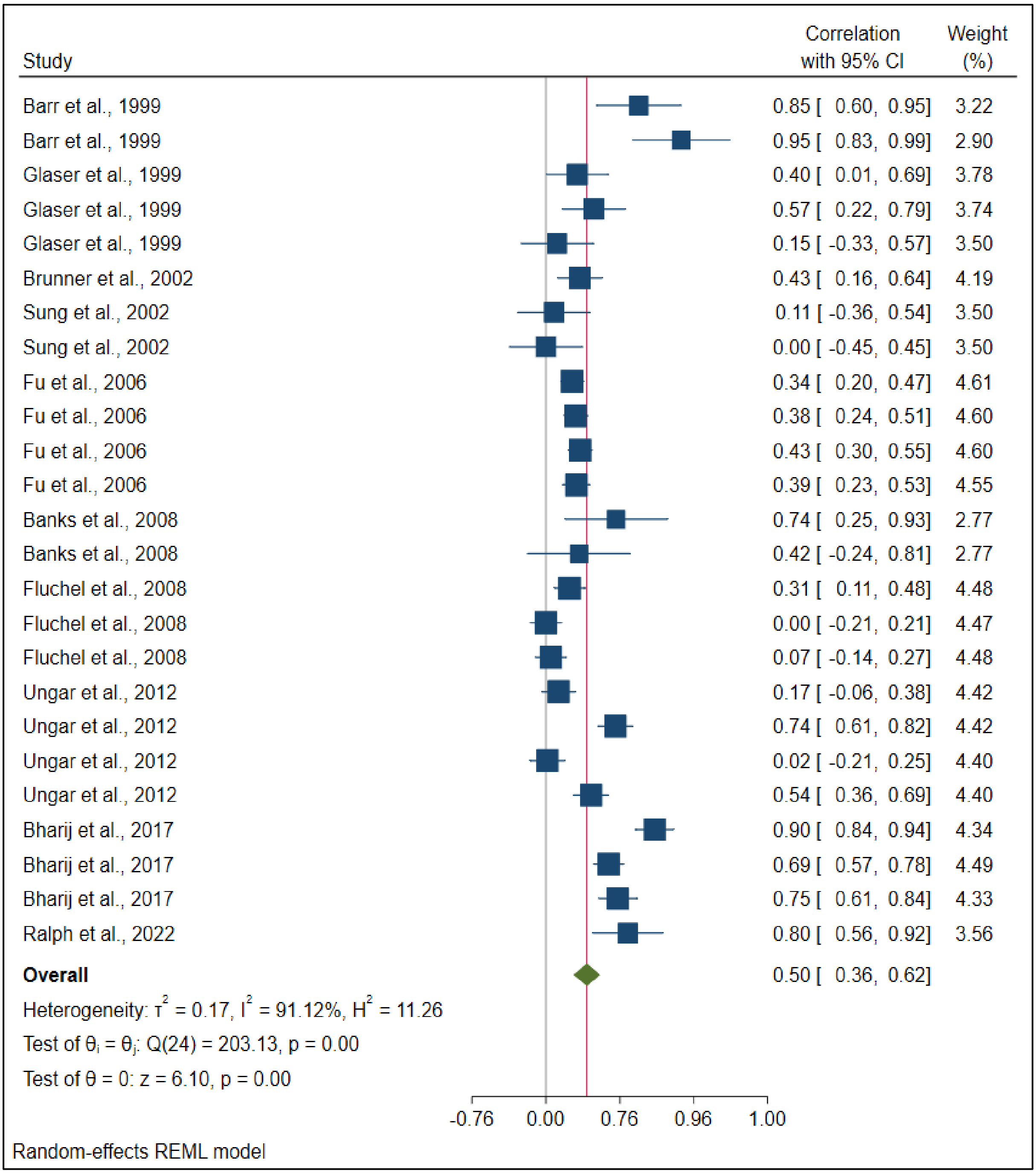


Figure 2.2 Summary of the interrater reliability across studies. The forest plot depicts the study-specific and overall estimates of Intraclass Correlation Coefficients (ICC), their respective 95% confidence intervals (CI) and the study weight (%) for 25 studies.

2.5.7.2 Inter-rater agreement for dimension level HRQoL

The estimated kappa and its 95% CI for HUI 2 and 3 attributes is summarised in Table 2.6. In total, 36 samples for HUI 2 and 68 samples for HUI 3 were synthesised, respectively, for the meta-analysis. The estimated Kappa values for HUI 2 attributes of “emotion” (0.25), “cognition” (0.3) and “pain” (0.38) and the HUI 3 attributes of “cognition” (0.23), “emotion” (0.27), “speech” (0.3) and “pain” (0.36) were the lowest. In contrast, there was higher agreement for the more easily observable physical or function related attributes such as “mobility” (0.61) for HUI 2 and “ambulation” (0.64), “dexterity” (0.65) and “vision” (0.78) for HUI 3. The heterogeneity was lower for HUI 2 studies (I^2=75%) than for HUI 3 studies (I^2=90%). Although no small-study bias was present in the analysis of HUI 3 samples (p=0.327), there was a possibility of such a bias using the HUI 2 samples (p=0.003).

Table 2-6 Dimension level overall Kappa estimates with their 95% confidence intervals (CI) for HUI 2 and 3.

MEASURE	ATTRIBUTE	AGREEMENT ($\hat{\kappa}$)	LOWER 95% CI	UPPER 95% CI
HUI 2	Self-care	0.576	0.347	0.806
	Cognition	0.296	0.088	0.505
	Emotion	0.250	0.158	0.342
	Mobility	0.615	0.463	0.767
	Pain	0.385	0.148	0.622
	Sensation	0.409	0.306	0.512
HUI 3	Ambulation	0.641	0.535	0.747
	Cognition	0.229	0.145	0.313
	Dexterity	0.646	0.541	0.751
	Emotion	0.272	0.190	0.353
	Hearing	0.497	0.232	0.762
	Pain	0.361	0.265	0.457
	Speech	0.300	0.174	0.427
	Vision	0.782	0.713	0.850

HUI 2 = Health Utilities Index (HUI) Mark 2, HUI 3= Health Utilities Index (HUI) Mark 3, $\hat{\kappa}$ = estimated Kappa value; Altman's scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good.

2.6 Discussion

This systematic review extends the existing evidence on child and proxy-reported health state utilities, such as those by Khadka et al.[115] and Jiang et al.[116], by focusing on the additional dimension of inter-rater agreement between child-proxy dyads. To date, this is the first study to comprehensively examine the inter-rater agreement between child-self and proxy reported HRQoL using generic preference-based measures across health conditions. Of the thirty studies included, most reported a poor to moderate agreement level (ICC=<0.5) between the child and proxy dyads for overall utilities. At the dimension level, there were some important differences common to all the generic preference-based measures. In particular, the agreement between children and proxy respondents was weaker for psychosocial-related HRQoL (e.g., “emotion”, “cognition” and “feeling worried, sad or unhappy”) dimensions and stronger for physical HRQoL dimensions (e.g., “vision”, “hearing”, “looking after myself” and “walking about”).

It is unclear how the preference-based measure/s applied in the identified studies influence on the level of agreement between self and proxy reported child HRQoL. In this review, greater agreement with HUI 2 was observed than HUI 3. There are two main differences between the measures. First, the two measures differ in their response levels. HUI 3 has 5-6 response levels whereas HUI 2 has 3-5 [202]. Intuitively, a higher inter-rater agreement would be expected with measures with fewer response levels, assuming that fewer choices may reduce ambiguity and ease the rating process. Second, HUI 2 and HUI 3 have different underlying constructs for the attributes with the same name. For example, in HUI 2 the “emotion” attribute assesses distress and anxiety, whilst the HUI 3 frames “emotion” in terms of happiness rather than depression [202]. Currently, there is insufficient evidence to investigate whether the discrepancy reflects this difference or is a coincidental finding.

In contrast to the above, the study evaluating the child and proxy agreement using the EQ-5D-Y 3L and 5L versions found a higher agreement, although not significant, with the five-response level version than with three [176]. This counterintuitive finding suggests that for the EQ measures, a greater number of response options may provide a more precise and nuanced recording of HRQoL [203], thereby, potentially, enhancing agreement. Similarly, the agreement for the CHU9D, which also has five response options, was found to be good to very good. [193, 199]. However, this does not imply that more response levels always result in better agreement. There may a practical limit to the number of response levels that can be effectively used before they no longer contribute to, or may even reduce, the dyadic agreement. For example, both agreement and correlations for the EQ VAS generally ranged from fair to moderate, which was lower compared to the dimension level agreement for the corresponding descriptive system. This discrepancy may be attributed to several factors, including the use of different test statistics to measure agreement/correlation and the variations in response scales. Specifically, the EQ VAS has a response scale from 0 to 100, whereas each of the five dimensions of the EQ-5D measures are described using a 3 or 5 level response scale [77]. Hence, a higher discrepancy may be expected with EQ VAS due to the much larger range for its response scale.

Proxy type used was found to have some influence on the level of agreement between self and proxy reported paediatric HRQoL. The findings of HRQoL studies conducted in a paediatric oncology setting suggest that the information obtained from the child, the parent and the health professional are generally complementary and valid [122]. However, Sprangers and Aronson concluded that health professionals generally tend to underestimate the pain and also, conversely, the overall HRQoL of the individual [204]. Whilst able to accurately assess the patient’s physical condition, health professionals often failed to consider the emotional and social components of HRQoL [205]. In line with previous studies in adult cancer patients where agreement was higher with close companions, the child-parent agreement in this review was also found to be higher compared to child-health professional agreement [206]. Moreover, mothers demonstrated a higher agreement than fathers. This gender disparity may be associated with their degree of involvement in childcare [207].

The level of inter-rater agreement decreases with more severe conditions [205]. Longitudinal studies in paediatric patients found that the agreement between children and parents/caregivers was higher when their condition improved as compared to periods when they were ill [144, 176, 195, 200]. Additionally, the findings in this review indicate that cancer-related cohorts with active cancer had lower overall agreement than those with a history of cancer or with non-cancer health conditions. This is an important finding since the timing of HRQoL assessments, which can capture the changes in the individual's health status, form critical inputs for the calculation of QALYs for CUA. Divergences in self and proxy reported childhood HRQoL over time and according to the severity of the condition may impact, potentially substantially, upon the results of economic evaluations and regulatory decision-making for the recommendation of new pharmaceuticals/medical technologies.

Interestingly, a low inter-rater agreement was seen between children with no obvious health conditions and their parents/caregivers. One study in this review, for instance, showed worse correlations between parents and healthy children than children with a history of cancer [182]. A similar trend was observed in a cohort of adolescents, where those with chronic health conditions exhibited higher agreement levels on the non-preference-based PedsQL™ compared to their healthy peers [208]. It remains unclear whether this low agreement reflects a true pattern or is the result of systematic differences in how HRQoL dimensions are understood by the child and the adult caregivers. Further exploration through qualitative studies could help clarify whether this is a demonstrable trend and shed light on the underlying reasons.

Self and proxy agreement data in the assessment of mental health conditions remains scarce, highlighting a significant gap in our understanding of the inter-rater agreement within this vulnerable population. Although studies have examined HRQoL in children suffering from mental or behavioural disorders using preference-based measures [209-213], the assessment of the level of child-proxy agreement in this population remains largely unexplored, with only one study to date specifically addressing it [200]. Proxy-child concordance may be crucial in this cohort, especially since children with mental or behavioural disorders may experience several barriers in communicating their feelings and experiences [214]. A pivotal review by De Los Reyes et al. underscores the importance of the choice of proxy in reporting the psychosocial functioning of children, highlighting the discrepancies that exist between children and adult informants [215]. Such discrepancies may lead to misunderstandings about the child's condition and potentially impact the adoption of interventions that are sensitive to the actual needs and experiences of these children.

Self-report using the EQ-5D-Y has been prescribed for children aged 8 and older [77]. The use of HUI 2/3 is not recommended for self-report in children under 12 [202]. However, studies included in this review have reportedly used these measures for self-completion in children in a younger than recommended age group. The minimum age at which children can reliably and accurately self-report has not been conclusively identified yet and is likely to be influenced by a variety of factors (including the reading and comprehension abilities of the child, the measure/s being applied and the mode of completion) [98]. There also remains a gap in the literature exploring the potential for differential levels of agreement between proxies and children by age groups. Previous studies have shown lower agreement among older children than their younger counterparts [216, 217], while others have reported that larger inter-rater gaps with younger children [218, 219]. In this review, one study reported the agreement statistics (Gwet's AC_1) for children (10 to 12 years) and adolescents (13 to 15 years) separately. In both groups, the correlation between child-self and proxy reported dimension level HRQoL was strong and positive, with a marginally stronger association reported between adolescents and caregivers than children and caregivers [201]. Due to these inconsistent findings, further research is needed to determine if an age differential exists in the level of child-proxy agreement.

In this review, 33% of the studies reported only the correlation coefficients which were synthesised to measure the inter-rater agreement. The difference between agreement and correlation has been addressed in the literature [154]. However, until recently, standalone correlation coefficients have been employed to assess agreement between child-self and proxy report [200]. Correlation and agreement, both measure the strength of association between the two variables of interest. However, the key difference is that agreement coefficients, in addition, account for the absolute agreement between the raters. Correlations may be high even if the ratings are not equal but only vary similarly. On the other hand, a perfect agreement would imply that all ratings, by each rater, are the same [118, 153]. Thus, correlation coefficients, if used, presented along with agreement statistics may provide a more comprehensive picture of the level of agreement.

2.7 Limitations

This review has several limitations which are important to highlight. The inter-rater agreement for overall utilities and for the respective dimensions was quantitatively examined for only HUI 2 and 3 for the following reasons: (1) HUI measures were widely used among the studies included in this analysis with HUI 3 being the most dominant (2) Despite its relatively wide application, the majority of the identified studies using the EQ-5D-Y did not report the overall utilities, potentially due to the absence of an established preference-based scoring algorithm for the EQ-5D-Y for the respective countries at the time of publication. When reported, the utilities were calculated using country-specific preference-weights for the adult EQ-5D-3L [191, 200]. Furthermore, these studies only reported the correlation (rather than agreement) using Pearson's *r*. Whilst agreement was reported for the EQ VAS scores, they were not pooled due to paucity of evidence demonstrating the comparability of the VAS scores with the index scores. The EQ VAS scores were therefore not included in the meta-analysis. Additionally, the lack of studies reporting the dimension-level agreement between self and proxy reports of child HRQoL, along with percent agreement and published guidelines for pooling agreement statistics such as Gwet's AC, made a meta-analysis of the EQ-5D-Y dimensions infeasible (3) The analysis for the CHU9D and QWB was also limited due to inadequate reporting of agreement statistics. The interpretation of the results of the meta-analysis is bounded by the presence of a high heterogeneity which could not be explained by the sub-group analysis.

Further, due to practical resource constraints only articles published in the English language were included. Although no publication bias was detected, grey literature was not searched. Additionally, authors were not contacted for unreported data, resulting in the rejection of several papers. Additionally, data extraction was conducted by a single author which may have affected the reliability of the findings. Finally, many of the studies included in the review did not provide detailed descriptions of the nature or extent of assistance provided, nor did they report whether parents completed their own proxy-reports before or after assisting their children. This lack of consistency in reporting limited the ability to systematically examine the influence of assisted self-report on the estimates. Future research should aim to standardise the reporting of assisted self-report practices to enable more meaningful analyses of its potential impact on HRQoL estimates.

2.8 Conclusions

This systematic review summarising the agreement between child-self and proxy rating of HRQoL using established generic preference-based measures generally found a low

(ICC/CCC<0.5) inter-rater agreement. Convergence with child-self rating was more likely in the proxy assessment of paediatric HRQoL within dimensions with observable attributes e.g., physical health dimensions than in those with less-observable attributes e.g., psychosocial dimensions. Further research to drive the inclusion of children in self-reporting their own HRQoL wherever possible and limiting the reliance on proxy reporting of children’s HRQoL is warranted.

CHAPTER 3: METHODS

3.1 Overview

This chapter details the research design, methodology, and analytical approaches used in the first empirical study, i.e., Study 1. The results obtained using these methods are reported upon in Chapters 4, 5, 6, 7. It describes the sample selection, data collection procedures, and the measures employed for assessing HRQoL in children and proxies for the study. (Please refer to Chapter 8 for the methods for Study 2).

3.2 Background

As discussed in Chapter 1, self-report of the person's own HRQoL is considered the gold standard for measurement, given that a person's perceptions of their health state are what constitutes HRQoL. In child health research, information gathered through self-reported measures of a child's HRQoL can offer a more comprehensive perspective of their overall subjective well-being, beyond just their physical health [220, 221]. However, it is common for parents to act as proxy respondents when assessing the HRQoL of children, especially in cases where the child is too young or unwilling to provide their own responses or has a health condition that impacts their ability to self-complete the measure [94, 98]. While parental reports can offer valuable insights, parents may have different perceptions of their child's HRQoL compared to the child's own perceptions [106].

Chapter 2 (systematic review and meta-analysis) provides a comprehensive overview of the evidence in the literature regarding inter-rater agreement between self and proxy-reports for preference weighted generic HRQoL measures in children. Of the 35 studies included, the inter-rater agreement was found to be low for overall HRQoL across the different measures used. Notably, the agreement between children and proxy respondents was lower for psychosocial-related dimensions than for physical dimensions within these measures [222].

This finding was consistent with previous research reporting discrepancies between child self-reports and parental reports of HRQoL [115, 222, 223]. It is, therefore, crucial to not only evaluate how closely the report provided by the parents aligns with the child's self-report, i.e., the extent to which the two raters agree in their ratings, but also to investigate the factors influencing this agreement and the underlying reasons for any discrepancies to gain more comprehensive insights [118].

Agreement statistics are indispensable for determining inter-rater agreement in HRQoL assessments. In the context of child HRQoL, agreement statistics help to identify the degree of alignment between self-reports and proxy reports. High agreement can indicate that parents can closely represent their child's health perceptions, while low agreement suggests discrepancies that warrant further investigation. As mentioned in Chapter 2, statistics such as, ICC and Cohen's Kappa (K), provide a rigorous method to quantify the level of agreement between different raters. These measures provide an assessment of the agreement that accounts for the possibility of agreement occurring by chance alone [118, 151, 152, 222]. On the other hand, correlation coefficients, such as Pearson's r or Spearman's rho, measure the strength and direction of a linear relationship between two variables. However, they do not account for chance agreement and can be misleading when used to assess agreement. For example, a high correlation coefficient does not necessarily imply good agreement; it could simply reflect a consistent pattern of differences between the two sets of ratings. Thus, this thesis employs agreement statistics to better understand the extent to which the parental-report is representative of the child's own HRQoL, thereby informing the interpretation of HRQoL values.

Bevans et al. suggest that cognitive interviews with children can assist in measuring their cognitive capacity and provide evidence of their understanding of HRQoL dimensions [99]. Cognitive interview methods such as the think aloud, which involves articulating one's thoughts while engaging in a task (concurrent) or after task completion (retrospective), are a key tool for assessing the quality of responses and for determining whether the HRQoL measure is generating the intended information from the respondent [224, 225]. Cognitive interviews including think aloud have been employed with children aged 5 years and older in the development and content validation of HRQoL measures [226-229]. Additionally, these methods have been used to adapt established HRQoL measures to different cultural and linguistic settings [230-232]. For example, cognitive interviews were conducted with samples of healthy and chronically ill children aged 8-18 years from Germany, Italy, Spain, and Sweden to adapt the EQ-5D-Y to the respective languages [36].

Cremeens and Blades conducted a qualitative study using think aloud to investigate how children aged 5-9 years understood and responded to questions in the TedQL [233], a generic measure of HRQoL for children aged 3–8 years [140]. Their findings suggested that older children had a better understanding and interpretation of the items than younger children [140]. While cognitive interviews have been conducted with children and adolescents to assess their understanding of non-preference based HRQoL measures [234, 235], there is a limited amount of qualitative research exploring how children comprehend and respond to child-specific preference-based measures for HRQoL. Notably, Amien et al. conducted a study employing cognitive interviews to assess the comprehensibility of the EQ-5D-Y-3L interviewer-administered version among a South African sample of children aged 5-7 and 8-10 years [236]. Their findings indicated that children aged 5-7 years experienced more challenges with comprehension, specifically with the “looking after myself” dimension compared to children aged 8-10 years (5-7 years: 55% vs 8-10 years: 28%) [236]. However, the extent to which these comprehension challenges affect the self-report validity of the HRQoL measurements in younger children, specifically the impact of these challenges on the reported HRQoL, remains underexplored.

Tourangeau’s four-stage response model is a useful approach for identifying and categorising issues related to the four cognitive processes involved in responding to questions in a HRQoL measure. These processes include comprehension (understanding the question), retrieval (recalling relevant information), judgment (evaluating the information), and response (selecting and reporting an answer) [237]. A framework based on Tourangeau’s response model has recently been successfully utilised with a think aloud protocol to assess the self-report reliability of the EQ-5D-5L in older adults with varying levels of cognitive impairment and dementia [238, 239]. Additionally, a study by Nwankwo et al. applied a similar methodology to investigate participant responses when completing the HRQoL measures, including the EQ-5D-5L, among individuals with end-stage organ failure [240]. In conjunction with quantitative methods, the qualitative think aloud approach may thus offer additional insights into the psychometric properties of a measure [241, 242].

This study (Study 1) first, quantitatively, examines the inter-rater agreement (for both, the overall HRQoL and the HRQoL dimensions) between child-parent dyads and then qualitatively analyses the child responses to the think aloud interview to determine if they engage meaningfully with the HRQoL measures. The detailed methods are outlined below.

3.3 Measures

3.3.1 EQ-5D-Y

The EQ-5D-Y-3L and its proxy versions were used to examine inter-rater agreement between self and proxy reported HRQoL [86]. For proxies, in version 1, the proxy is asked to rate their child’s HRQoL according to their opinion (proxy-proxy), whilst 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 (when self-reporting) or their child's overall health status (when proxy-reporting) on a scale from 0 to 100, with 0 indicating the worst possible health state and 100 the best possible health 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 TTO approach was applied to both proxy and self HRQoL ratings to generate the overall HRQoL or HRQoL values (preference-weighted HRQoL). The weights range from 1, indicating full health to -0.217 for the worst health state (33333) [243]. 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 [88]. However, arguably, such differences are of lesser importance for this particular study, since the 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) [244]. 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 [245]. The preferences for the EQ-5D-Y-3L health states were elicited from a German adult population using DCE and composite TTO (c-TTO) methods. The value set was applied to both self and proxy responses.

3.3.2 CHU9D

The CHU9D, a validated generic preference-based measure of children's HRQoL, includes 9 dimensions: "Worried", "Sad", "Pain", "Tired", "Annoyed", "Schoolwork/homework", “Sleep”, “Daily routine”, and “Activities” and each dimension has five response levels. A scoring algorithm can be used to generate individual level preference-weights for all possible response combinations to the CHU9D. The preference-weights range from 1 (full health) to -0.1059 for the most severe (PITS) state [56]. An Australian adolescent-specific preference-based scoring algorithm, derived from Australian adolescents aged 11-17 years, was applied in this study to calculate the CHU9D generated utilities [56].

3.3.3 PedsQL™ 4.0 generic core scales

The PedsQL™ 4.0 Generic Core Scales include 23 items that are grouped into 4 Scales (dimensions): Physical Functioning (8 items), Emotional Functioning (5 items), Social Functioning (5 items), and School Functioning (5 items). The psychosocial dimensions represent the emotional, social, and school functioning subscales of the PedsQL™ whilst the physical dimension represent the physical functioning scale. For children aged 5-7 years, the response levels are simplified to three levels: “Not at all”, “Sometimes”, and “A lot”. For children aged 8 years and above, the response levels include five options: “Never”, “Almost never”, “Sometimes”, “Often”, and “Almost always” [246].

Since the PedsQL™ is not a preference-based measure, equal weights were assigned for each of its 23 items when calculating the total score. Items were scored in reverse and transformed into a 0-100 continuous scale (0 = 100, 1 = 75, 2 = 50, 3 = 25, 4 = 0), such that higher scores represented better HRQoL. To calculate the mean for individual scale scores, the items were summed across and divided by the number of items answered. The average individual scale scores were used to compute a total summary score [246].

3.3.4 EQ-5D-3L

The EQ-5D-3L was used to measure the HRQoL of the parent participants. The EQ-5D-3L measures HRQoL across five dimensions: “mobility”, “self-care”, “usual activities”, “pain/discomfort”, and “anxiety/depression”. Each dimension has three different response options, ranging from no problems to severe problems [243]. An Australian adult scoring algorithm was applied to calculate the adult utilities (as mentioned above) [243].

3.4 Participant recruitment

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 a pre-specified 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 or any condition that would impact their ability to self-complete the measures 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[3]. 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). Participants received an e-gift card worth AUD $90 as a token of appreciation for their participation from the social research company.

Information about the child participant's age, gender, and household income bracket (less than $50,000 or greater than or equal to $50,000[4][247]) was collected from parents following

---

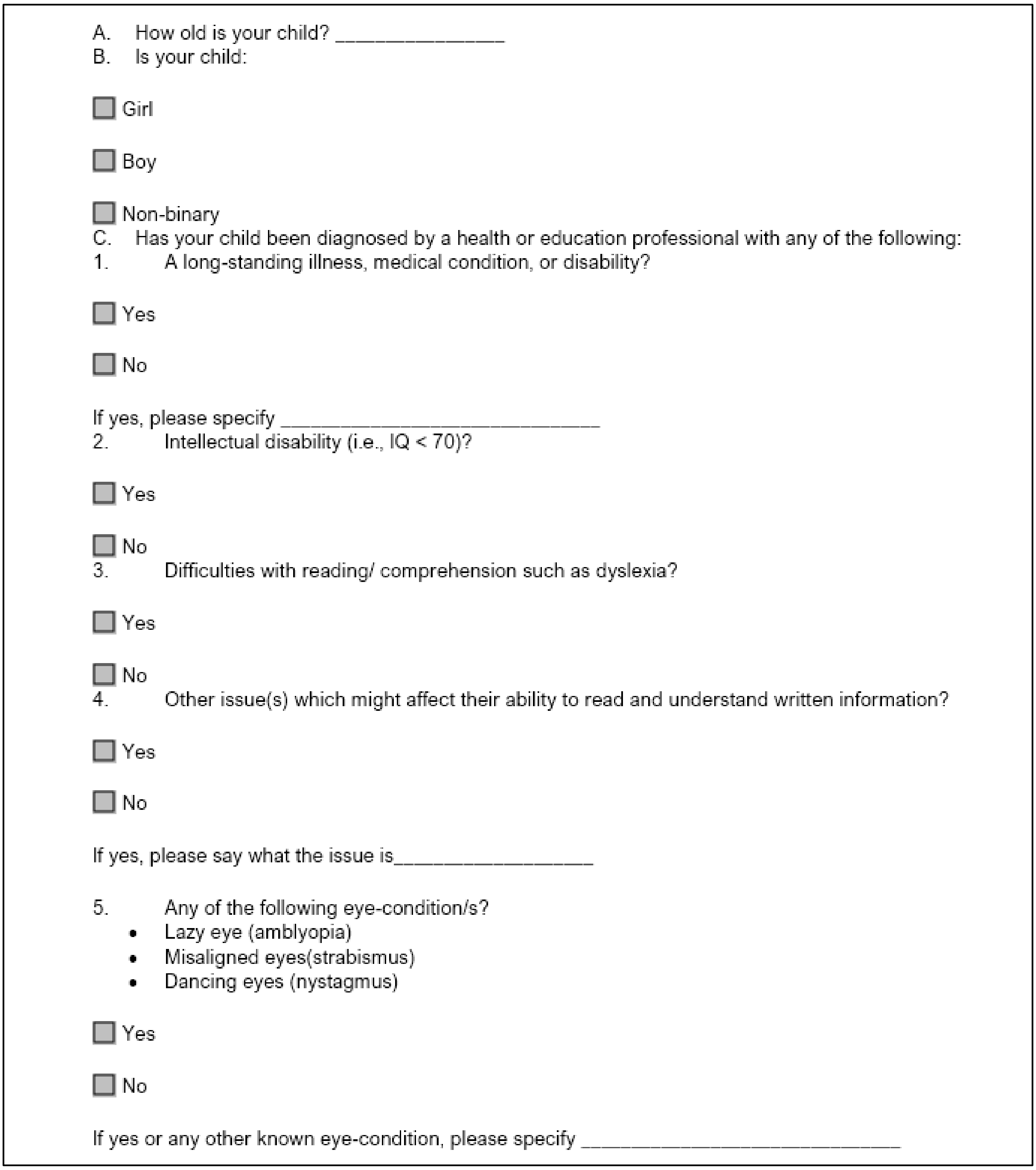
[3] The eye tracking data, focussing on the analysis of children's gaze patterns and other metrics, will be explored as the subject of subsequent research.

[4] This income threshold was roughly calculated based on the average Australian household weekly income of $866, which falls within the 20th percentile of income distribution and is categorised as part of the low-income bracket according to the Australian Bureau of Statistics (2019-20), Survey of Income and Housing, User Guide, Australia. Available from https://www.abs.gov.au/statistics/detailed-methodology-information/concepts-sources-methods/survey-income-and-housing-user-guide-australia/2019-20/income#low-and-lower-income-households

informed consent to participate from both the parent and child at the pre-interview stage via email/ mail correspondence. 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 (Fig 3.1).

A broad representation in relation to key socio-demographic characteristics and common health conditions affecting children in the general population such as asthma, anxiety disorders, conduct disorders, depressive disorders, autism spectrum disorders and dental caries was achieved using a stratified random sampling method [248]. This method involved dividing the population into distinct subgroups, or strata, based on gender (girl vs boy), socio-economic status (income brackets) and health condition. Random samples were then recruited from each stratum in proportion to their presence in the overall population, ensuring that each subgroup is adequately represented in the final sample. This approach enhances the generalisability of the findings by reducing sampling bias [249].

Previous studies suggest a high percent agreement (approximately 60%) can be expected between proxies and children from the general population [147, 250]. In this thesis, to assess inter-rater agreement, Gwet’s AC was preferred over Cohen’s Kappa for several reasons. First, Gwet’s AC is less affected by the presence high percent agreement which can cause Cohen’s Kappa to underestimate agreement. Second, Gwet's AC provides a more stable measure of agreement when the marginal distribution of data is imbalanced, meaning the distribution of ratings is heavily skewed towards certain categories. Cohen’s Kappa can produce lower agreement estimates in such cases where the distribution of ratings is unevenly skewed. Therefore, given that the sample in Study 1 is from the community and generally healthy, these advantages make Gwet’s AC better suited to measure inter-rater agreement compared to Cohen’s Kappa [151, 161]. Gwet recommends a sample size of 25 to estimate the agreement coefficient with an error margin of 20% under high percent agreement [161]; however, this study aimed for a larger sample size (N>25) to increase the statistical power of the agreement analysis.



A. How old is your child? _______________

B. Is your child:

☐ Girl

☐ Boy

☐ Non-binary

C. Has your child been diagnosed by a health or education professional with any of the following:

1. A long-standing illness, medical condition, or disability?

☐ Yes

☐ No

If yes, please specify ______________________________

2. Intellectual disability (i.e., IQ < 70)?

☐ Yes

☐ No

3. Difficulties with reading/ comprehension such as dyslexia?

☐ Yes

☐ No

4. Other issue(s) which might affect their ability to read and understand written information?

☐ Yes

☐ No

If yes, please say what the issue is__________________

5. Any of the following eye-condition/s?

- Lazy eye (amblyopia)
- Misaligned eyes(strabismus)
- Dancing eyes (nystagmus)

☐ Yes

☐ No

If yes or any other known eye-condition, please specify ______________________________

Figure 3.1 Pre-interview screening questions for parents.

3.5 Study design

A convergent mixed method design was used. The intent of the convergent design was to gain a comprehensive understanding of child-self reporting of HRQoL by drawing on the strengths of both quantitative and qualitative methodologies with equal emphasis given to both [251]. The quantitative component involved the independent completion of the validated HRQoL measures (see Table 3.1), administered to both children and their parents. The qualitative component included the think aloud interview with only the children to explore their understanding and reasoning behind their responses.

Integration was achieved at the data analysis stage by using a data transformation model, which quantifies qualitative data for incorporation into the quantitative analyses [251]. This approach allowed for a holistic view of the data, capturing both the numerical extent of inter-rater agreement and the underlying reasons for any discrepancies. Additionally, the qualitative findings were reported in detail to further elaborate on and provide context to the integrated results.

3.6 Ethics

Study 1 was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. The study was conducted in South Australia and complied with the ethical guidelines of the Flinders University Human Research Ethics Committee (Project ID 4178). Written informed consent to participate in the study was sought from the parent for themselves and on behalf of the child prior to commencing the interview.

3.7 Procedure

Child-parent dyads were invited to attend a semi-structured, face-to-face interview with a researcher at Flinders University to complete a series of validated HRQoL measures (Table 3.1). Each interview was scheduled for a maximum duration of 45 minutes. A team of three interviewers, including DK, was assigned to conduct the interviews. Participants were offered two interview slots, which were conducted simultaneously by DK and another interviewer. To ensure that the interviewers were well-prepared for conducting interviews with children, they underwent training sessions that included mock interviews with consenting colleagues and their children. These training sessions were supervised by the Principal Supervisor, who has adequate experience in conducting interviews with children (outlined in the ethics application). An interview protocol was designed and provided to all interviewers to ensure consistency and reliability in the interview process. To minimise any psychological harm in the interview, which included questions about the parent's and the child's health, participants were given the opportunity to withdraw from the interview at any time if distress occurred (see materials provided under Appendix 10.2 for more details).

The interview with the child comprised two stages: quantitative and qualitative. In stage one, namely, the quantitative stage, the child was invited to self-complete their own HRQoL assessment using the EQ-5D-Y-3L, the CHU9D, the PedsQL^{TM}, and a self-rated general health (SRH) item question- "In general, would you say your health is poor, fair, good, very good, or excellent?" [252], administered online via the REDCap, a secure web-based platform for data capture and management [253, 254]. The order of the generic preference-based measures, i.e., the EQ-5D-Y-3L and the CHU9D was randomised to appear either first or last, with PedsQL^{TM} positioned in the middle. This was done using a stratified randomisation method based upon the child's chronological age (6-7 years vs 8-12 years). The on-screen process of children completing the measure was recorded in a video (screen recording) to later replay in the qualitative stage of the interview.

Following the completion of the questionnaires, in stage two, face-to-face retrospective think aloud interviews were conducted using an age-appropriate interview guide specifically developed for this study. The retrospective think aloud method was chosen as it allows uninterrupted completion of the questionnaire by the child and reduce the workload during the task [255, 256]. The interviewer asked the children to retrospectively think aloud whilst recording the conversation on a voice recorder. The recorded video of the child completing the EQ-5D-Y-3L on the laptop screen was played back to them to assist with the process of think aloud and prompt recollection of their responses. The interviewers explained the task to the participants and asked them to verbalise their thoughts and reasons for their choices. Only minimal verbal probes were used to encourage responses. Probes used were open-ended such as “can you explain why you chose that answer?” and “was there anything that confused you?” (please see Appendix 10.2 interview protocol for more details). At the conclusion of each interview, the interviewer assessed the child’s understanding and engagement with the questionnaire. This evaluation was categorised into five distinct levels, adapted from a study by Guerriero at al. [257], to determine the child’s ability to comprehend and complete the task:

- Category 1: The child did not complete the task due to fatigue.
- Category 2: The child did not complete the task, but there is insufficient evidence to conclude that a lack of understanding prevented completion.
- Category 3: The child did not complete the task, and there is sufficient evidence to assume that a lack of understanding prevented completion.
- Category 4: The child completed the task, but comments and questions indicated a poor understanding of the task.
- Category 5: The child completed the task, and comments and questions indicated a good to excellent understanding of the task.

Simultaneously, the parent respondent was asked to self-complete the corresponding proxy versions of the measures. The proxy versions of the EQ-5D-Y-3L included two perspectives, proxy-proxy and proxy-child whereas the CHU9D and the PedsQL^{TM} were assessed from proxy-proxy perspective only. These measures were provided as hard copies whilst using noise-cancelling headphones such 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 may otherwise occur in an interviewer-led mode of administration [258]. The parent first completed the (traditional) proxy-proxy version on behalf of the child for all the three measures. 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 SRH item about themselves, and a series of socio-demographic questions including their age, gender, and postcode.

There is some evidence to indicate that online and paper-pen administrations are equivalent [259] as long as they are consistent for each rater [260]. The respective method for each rater type was chosen as a matter of convenience and resource availability.

Table 3-1 List of HRQoL measures to be completed by the participants.

Child 6-7	Child 8-12	Parent
CHU9D self-report	CHU9D self-report	CHU9D proxy-report (version 1)

PedsQL self-report (aged 5-7 version)	PedsQL self-report (aged 8-12 version)	PedsQL proxy-report (version 1)
EQ-5D-Y self-report	EQ-5D-Y self-report	EQ-5D-Y proxy-report (versions 1 & 2)
		Socio-demographics (including EQ-5D self-report)

3.8 Data Analysis

3.8.1 Quantitative analysis

Sociodemographic data were analysed using descriptive statistics. The relative socio-economic disadvantage of the postal area in which the child-parent dyad resided was determined from the Socio-Economic Indexes for Australia (SEIFA) area-decile number[5] [261]. The first six decile numbers were classified as disadvantaged quintiles (quintiles 1 to 3) and the last four as advantaged quintiles (quintiles 4 and 5) [261]. Sub-groups were based on age, the presence of long-term health condition/s (Y/N), and child gender (girl, boy, non-binary) 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 the aim to contrast the responses from younger children, under the age of 8, with those in the older age group for which the EQ-5D-Y-3L is typically recommended for self-completion. Additionally, the 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-

---

[5] Income bracket was not used as a variable to denote socio-economic status because the recruitment company was unable to recruit a sufficient number of dyads from the lower income thresholds; only 10 participating dyads reported an income of less than $50,000.

rater differences and agreement were analysed for the overall sample and by sub-groups for the overall HRQoL utilities (EQ-5D-Y-3L and CHU9D) and the HRQoL scores (EQ VAS and PedsQL^{TM}) and the dimension level HRQoL. Further intra-proxy agreement was estimated for the overall and dimension level HRQoL for the two proxy versions of the EQ-5D-Y-3L.

Medians and interquartile range (IQR) were primarily used to describe the summary statistics for the HRQoL values and the HRQoL 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. However, means and standard deviations (SD) were also reported in the tables for completeness. Further, agreement was assessed using Concordance Correlation Coefficient (CCC) for the HRQoL values and EQ VAS scores [152, 262]. The CCC is frequently used to evaluate agreement between two raters and does not rely on the analysis of variance (ANOVA) model assumptions unlike ICC (see Chapter 2, section 2.4.5 for more details) [263]. Gwet's agreement coefficient (AC_1) was used to analyse the dimension level HRQoL, as mentioned previously [161]. Agreement estimates were reported along with their corresponding 95% confidence intervals (CI).

For the dimension level agreement, the unweighted AC_1 was chosen because: 1) a predominance of healthy children in the study sample, and 2) 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 sample (see section 4.4.5). 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 [161, 264]. The Wilcoxon matched-pairs signed-ranked test was used to compare group differences for continuous variables (and non-paired for when necessary) and Fisher's exact test for categorical variables. In this study (Study 1), the statistical significance level was set at 0.05. Analysis was carried out using Stata 16.1 (Stata Corp LLC, College Station, TX, USA) [265].

3.8.2 Qualitative analysis

Interviews were transcribed verbatim, and the data was analysed using NVivo qualitative analysis software (release 1.3). The Tourangeau four-stage response model framework was used to identify response issues at each of the four stages of comprehension, recall, judgement, and response mapping (Table 3.2) [237]. This model, grounded in the cognitive theory of survey response, recognises that each stage presents opportunities for cognitive errors, potentially leading to response biases and inaccuracies. By facilitating a systematic examination of these stages, the framework provides a structured approach to identifying and understanding the sources and nature of response issues, thereby validating survey responses [237].

Two coders (DK and KL) independently coded the transcripts and identified response issues in accordance with the Tourangeau response model framework. The identified issues were then categorised into one or more type of response issues, with potential overlap, depending on the stage at which the response process was deemed erroneous. Codes were created based on the perceived source of response issue within these categories. The EQ VAS was not coded for response issues under the framework; however, an inductive thematic analysis was undertaken. Open coding was conducted followed by the iterative development and refinement of themes to understand perceptions and interpretations of the EQ VAS [266]. During the analysis, the codes and findings were thoroughly discussed, and any discrepancies were addressed through discussion with the supervisory panel (JR and JK) until a consensus was reached.

3.8.3 Integration

Qualitative data, comprising response issues using the Tourangeau four-stage response model framework were integrated into the quantitative data analysis. Data transformation was achieved by converting the response issues into numerical variables that indicated the type of issue, as detailed in the qualitative analysis section. The sample was described using descriptive statistics based on the response issues of comprehension, judgment, recall, and response mapping for the subgroups: 1) age-group: children 6-7 years old, 8-10 years old and 11-age group and 2) presence of long-term condition: yes or no and 3) gender: girl or boy (or non-binary, if reported). A dichotomous variable representing the presence or absence of response issue was used.

Table 3-2 Tourangeau four-stage response model framework.

Stage	Description
Comprehension	Involves problems with the interpretation of the HRQoL dimensions, where participants encounter problems understanding or have misunderstandings of words, phrases, or response options. Issues may arise due to ambiguous wording or unfamiliar terms.
Recall	Involves problems with the retrieval process, encompassing challenges such as using an incorrect time frame for information retrieval and difficulties in recalling relevant information that aligns with the measure's specified time frame.
Judgment	Involves problems with assessing the information retrieved, including drawing upon irrelevant information or inadequate evaluation, potentially leading to under or over-reporting of health states. This stage is often closely linked to comprehension, as it entails the evaluation of recalled information to determine its relevance to the specific question and is influenced by how the question was understood in the first place.

Stage	Description
Response Mapping	Addresses issues with the response categories, such as dissatisfaction with the provided options or inappropriate application of them (e.g., selecting multiple options when only one is permissible, or reluctance to choose any). It also includes cases where the stated answer (verbal protocol data) is misaligned with the chosen answer (survey response data), reflecting a disconnection in the final stage of the survey.

(Adapted from Lay, K. et al.[239])

3.9 Secondary Analysis

Chapter 8 utilises the existing P-MIC dataset to address the corresponding research questions. The Paediatric Multi-Instrument Comparison (P-MIC) dataset comprises an online cohort of 845 child-parent dyads from across Australia. Children aged 6–10 years and their parents participated in this study, providing paired self- and proxy-reports of the child's HRQoL. Child-reported HRQoL was assessed using the EQ-5D-Y-3L self-report, while proxy-reports were collected using the EQ-5D-Y-3L Proxy Version 1. HRQoL values were derived using the Australian EQ-5D-3L (adult) value set and published EQ-5D-Y-3L value sets from nine other countries (see section 8.5.2). To explore variations in agreement, analyses were stratified by child age group (6–7 years and 8–10 years), gender, and the presence or absence of a health condition.

For statistical analyses, paired t-tests were employed to examine group differences. Concordance between child self-reported and proxy-reported HRQoL values was assessed using the ICC. At the dimension level, agreement between child and proxy responses was evaluated using Gwet's AC_1. Further methodological details can be found in Section 8.5 of Chapter 8.

3.10 Conclusions

In this chapter, the methodologies used in the first empirical study (Study 1) were detailed, including participant recruitment and selection, data collection, and analytical methods. A rigorous methodological framework was employed to ensure the robustness of the findings of this study. The subsequent chapters (4-7), present the findings of Study 1 and the respective discussions. Chapter 4 explores the impact of proxy perspectives on inter-rater agreement using the EQ-5D-Y-3L and its proxy versions. Chapter 5 investigates age-related variability in inter-rater agreement using the CHU9D and the PedsQL^{TM}. Finally, Chapters 6 and 7 examine children’s comprehension of HRQoL dimensions, employing mixed methods to assess their interpretation and identify response issues with the EQ-5D-Y-3L and CHU9D measures.

CHAPTER 4: 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

This chapter contains material from:

“Khanna, D., Khadka, J., Mpundu-Kaambwa, C., Chen, G., Dalziel, K., Devlin, N., Ratcliffe, J., & Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team (2024). 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. PharmacoEconomics, 42(Suppl 1), 113–128. https://doi.org/10.1007/s40273-024-01356-0”

4.1 Overview

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 [142]. 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, whilst the Proxy version 2 uses the ‘proxy-child’ approach whereby the proxy is asked to respond as they believe their child would complete the measure if they were self-reporting their own HRQoL [142]. 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 judgment (proxy-child) whilst maintaining the construct validity of the measure (see section 3.3.1 for more details) [122]. The results from the systematic review (Chapter 2) indicate that 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 chapter presents the following results of Study sample 1 for the 85 dyads who completed the respective self and proxy reports using the EQ-5D-Y-3L 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.

4.2 Summary

Objective: Self-reporting of 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 children (aged 6-12 years) and parent dyads were invited to participate in a semi-structured interview (N=85 dyads). 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 CCCs for the overall (preference-weighted) HRQoL whilst agreement at dimension level was evaluated using Gwet's AC_1. To assess the differences between the self and the two proxy reports, Wilcoxon

matched-pair signed-rank test was used (please refer to Chapter 3, section 3.3 onwards for specific details).

Results: 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 to the traditional proxy-proxy perspective (AC_1=0.58, moderate). Although the preference-weighted HRQoL values were 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 measures and for alternative value sets with different properties.

4.3 Key points

- 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 EQ-5D-Y-3L (self, proxy-proxy, and proxy-child), it

was found that proxy perspective influenced the agreement between child and proxy ratings of HRQoL.

- Proxy-child perspective showed a stronger agreement at the dimension level for the psychosocial dimension compared to the traditional proxy-proxy perspective. Whilst 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 proxy-child perspective was adopted, indicating that perspective may influence this aspect of HRQoL measurement.

4.4 Results

4.4.1 Sample characteristics

Children aged 6-12 years who did not have any difficulties with reading/comprehension, a diagnosis such as dyslexia or an intellectual disability (IQ < 70)[267] were included in the study. None of the children included in the study had any eye-condition contraindicated in eye-tracking (not discussed in this study, see section 3.3). 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 4.1 describes the socio-demographic characteristics of the child-parent respondents in the sample. The median age of children in this sample was 9 (IQR=2). The gender distribution was nearly balanced, with a slight overrepresentation of girls (56%), and no non-binary genders were reported. Parents in the sample had a median age of 41 (IQR=5) 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[6]: asthma (42%), autism spectrum disorder (8%),

---

[6] Health condition can vary in nature, ranging from acute disturbances to chronic disorders.

dental caries (15%), attention deficit hyperactivity disorder (4%), anxiety/depression (15%), sleep problems (12%) and congenital heart disease (4%). Based on the SEIFA area-decile numbers, in comparison to the Australian population, the sample had a lower representation of respondents residing in post-codes associated with relatively disadvantaged quintiles (37%) [268]. Finally, in terms of their ability to perform the required tasks, all of the children in the sample were classified under category 5, i.e., the child completed the task, and comments and questions indicated a good to excellent understanding of the task.

Table 4-1 Sociodemographic characteristics for all study participants (children and parent proxies).

Variable	**Child (N=85)**	**Parent (N=85)**
	N (%)	N (%)
Age		
Mean (SD)	9.13 (2)	41.7 (5.6)
Median (IQR)	9 (7-11)	41 (37-46)
Gender		
Female	47 (56)	68 (81)
Male	37 (44)	16 (19)
SRH		
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)	-

Variable	Child (N=85)	Parent (N=85)
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)	-
Socio-economic condition according to post-code		
Relatively advantaged quintile	-	52 (63)
(SEIFA decile 7,8,9,10)		
Relatively disadvantaged quintile	-	31 (37)
(SEIFA decile 1,2,3,4,5,6)		

*Health conditions can vary in nature, ranging from acute disturbances to, predominantly in this sample, chronic disorders.
SEIFA= Socio-Economic Indexes for Australia

4.4.2 Dyad EQ-5D-Y-3L values, EQ VAS scores and dimension level responses

Table 4.2 reports the EQ-5D-Y-3L values and EQ VAS scores of the overall dyad sample and by raters and sub-groups. Of the 85 dyad participants, two children did not report EQ VAS scores. When compared to 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 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 to 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 proxy-child perspective. In view of these findings, the 11–12-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. Interestingly, the self-rated HRQoL values were lower in girls (median= 0.84, IQR= 0.24) as compared to boys (median= 1, IQR= 0.16). However, the proxy-proxy rated HRQoL values were higher for girls (median= 1, IQR= 0.2) than for boys (median= 0.84, IQR= 0.2). None of the self and proxy (from both perspectives) HRQoL values were significantly different.

Table 4-2 Description of EQ-5D-Y-3L (self, proxy-proxy, and proxy-child) preference-weighted health states: overall and based on sub-groups.

Variable	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)
Overall						
N (%)	85	83	85	85	85	85
Mean (SD)	0.89 (0.15)	85.27 (14.01)	0.88 (0.13)	88.45 (9.85)	0.89 (0.14)	90.53 (9.65)
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 (p-value)			0.32	0.19	1	0.01*
Age group						
6 to 7 yrs.:						
N (%)	30 (0.35)	29 (0.34)	30 (0.35)	30 (0.35)	30 (0.35)	30 (0.35)

Variable	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)
Mean (SD)	0.85 (0.16)	90.05 (14.1)	0.88 (0.15)	91.48 (7.43)	0.88 (0.12)	94.3 (6.05)
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 (p-value)			0.78	0.49	0.79	0.48
8 to 10 yrs.:						
N (%)	30 (0.35)	29 (0.34)	30 (0.35)	30 (0.35)	30 (0.35)	30 (0.35)
Mean (SD)	0.87 (0.17)	84.21 (15.51)	0.88 (0.13)	87.67 (10.01)	0.88 (0.15)	87.33 (11.87)
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 (p-value)			0.78	0.49	0.79	0.48

Variable	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)
11 to 12 yrs.:						
N (%)	32 (0.38)	32 (0.38)	32 (0.38)	32 (0.38)	32 (0.38)	32 (0.38)
Mean (SD)	0.93 (0.12)	82.94 (12.02)	0.89 (0.11)	87 (10.98)	0.9 (0.14)	90.81 (8.61)
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 (p-value)			0.09	0.15	0.43	0.01*
Long-term health condition/s						
No:						
N (%)	59 (0.69)	58 (0.68)	59 (0.69)	59 (0.69)	59 (0.69)	59 (0.69)
Mean (SD)	0.89 (0.14)	85.76 (13.54)	0.9 (0.11)	88.78 (9.68)	0.91 (0.11)	91.27 (7.72)

Variable	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)
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 (p-value)			0.84	0.29	0.48	0.03*
Yes:						
N (%)	26 (0.31)	25 (0.29)	26 (0.31)	26 (0.31)	26 (0.31)	26 (0.31)
Mean (SD)	0.87 (0.18)	84.12 (15.27)	0.84 (0.16)	87.69 (10.38)	0.83 (0.17)	88.85 (13.06)
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 (p-value)			0.18	0.47	0.3	0.21
Gender						

Variable	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)
Girls:						
N (%)	47 (0.55)	47 (0.55)	47 (0.55)	47 (0.55)	47 (0.55)	47 (0.55)
Mean (SD)	0.86 (0.17)	85.85 (14.71)	0.88 (0.14)	88.06 (9.89)	0.88 (0.15)	91 (9.48)
Median (IQR)	0.84 (0.24)	92 (25)	1 (0.2)	90 (13)	1 (0.2)	90 (15)
Self vs proxy difference (p-value)			0.81	0.99	0.86	0.14
Boys:						
N (%)	37 (0.44)	36 (0.42)	37 (0.44)	37 (0.44)	37 (0.44)	37 (0.44)
Mean (SD)	0.93 (0.1)	84.5 (13.21)	0.89 (0.12)	88.89 (10.05)	0.9 (0.12)	89.81 (10.05)
Median (IQR)	1 (0.16)	89 (17.5)	0.84 (0.2)	90 (10)	1 (0.2)	90 (5)

Variable	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)
Self vs proxy difference (p-value)			0.07	0.16	0.24	0.07

*p-value significant at alpha=0.05 for 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

Figure 4.1 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 4.3, 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-value=0.02), "doing usual activities" (p-value<0.001) and the "feeling worried, sad or unhappy" (p-value<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, only for the "walking about" dimension, significant differences for self and proxy-proxy perspective (p-value=0.02), but not between self and proxy-child perspective (p-value=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 6–7-year-olds for "walking about" (p-value=0.03) and "doing usual activities" (p-value<0.01), and among the 8–10-year-olds for "doing usual activities" (p-value<0.01). In contrast to children with reported health condition, among the children categorised as relatively healthy (no reported health condition), a significant difference was observed across all the three dimensions: "walking about" (p-value<0.01), "doing usual activities" (p-value<0.001) and the "feeling worried, sad or unhappy" (p-value<0.01). Regarding gender, significant differences were observed in the "walking about" (p-value=0.03) and "doing usual activities" (p-value<0.001) dimension among girls. For boys, significant differences were observed for the "feeling worried, sad or unhappy" (p-value<0.01) dimension.

Figure 4.1 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)

Table 4-3 Distribution of EQ-5D-Y-3L (self, proxy-proxy, and proxy-child) dimension level responses: overall and based on sub-groups.

Variable	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 p-value$^{\#}$
Overall (N=85)	**Walking about**	No problems	76 (0.89)	84 (0.99)	83 (0.98)	0.02*
		Some problems	8 (0.09)	1 (0.01)	2 (0.02)	

Variable	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 p-value$^{\#}$
		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*
		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*
		A little worried, sad or unhappy	11 (0.13)	30 (0.35)	27 (0.32)	

Variable	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 p-value#
		Very worried, sad or unhappy		1 (0.01)	1 (0.01)	
Age group						
6 to 7 yrs. (N=23)	**Walking about**	No problems	19 (0.83)	23 (1)	23 (1)	0.03*
		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*
		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)	

Variable	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 p-value#
		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 yrs. (N=30)	**Walking about**	No problems	27 (0.90)	30 (1)	29 (0.97)	0.32
		Some problems	3 (0.1)		1 (0.03)	
		A lot of problems				
	Looking after myself	No problems	27 (0.90)	28 (0.93)	30 (1)	0.36
		Some problems	3 (0.1)	2 (0.07)		
		A lot of problems				
	Doing usual activities	No problems	23 (0.77)	28 (0.93)	30 (1)	<0.01*
		Some problems	7 (0.23)	1 (0.03)		

Variable	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 p-value[#]
		A lot of problems		1 (0.03)		
	Having pain/discomfort	No pain/discomfort	21 (0.7)	22 (0.73)	23 (0.77)	0.97
		Some pain/discomfort	8 (0.27)	7 (0.23)	7 (0.23)	
		A lot of pain/discomfort	1 (0.03)	1 (0.03)		
	Feeling worried, sad or unhappy	Not worried, sad or unhappy	25 (0.83)	18 (0.6)	18 (0.6)	0.08
		A little worried, sad or unhappy	5 (0.17)	12 (0.4)	12 (0.4)	
		Very worried, sad or unhappy				
11-12 yrs. (N=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)	

Variable	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 p-value#
		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				
Health condition						
No (N=59)	**Walking about**	No problems	52 (0.88)	59 (1)	58 (0.98)	<0.01*

Variable	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 p-value$^{\#}$
		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*
		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*
		A little worried, sad or unhappy	5 (0.08)	18 (0.31)	16 (0.27)	

Variable	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 p-value#
		Very worried, sad or unhappy				
Yes (N=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
		Some pain/discomfort	4 (0.15)	7 (0.27)	9 (0.35)	
		A lot of pain/discomfort	1 (0.04)		1 (0.04)	

Variable	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 p-value#
	Feeling worried, sad or unhappy	Not worried, sad or unhappy	20 (0.77)	13 (0.5)	14 (0.54)	0.19
		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)	
Gender						
Girl (N=47)	**Walking about**	No problems	40 (0.85)	46 (0.98)	46 (0.98)	0.03*
		Some problems	6 (0.13)	1 (0.02)	1 (0.02)	
		A lot of problems	1 (0.02)			
	Looking after myself	No problems	39 (0.83)	44 (0.94)	45 (0.96)	0.14
		Some problems	8 (0.17)	3 (0.06)	2 (0.04)	
		A lot of problems				
	Doing usual activities	No problems	37 (0.79)	44 (0.94)	47 (1)	<0.01*
		Some problems	10 (0.21)	2 (0.04)		
		A lot of problems		1 (0.02)		

Variable	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 p-value$^{\#}$
	Having pain/discomfort	No pain/discomfort	36 (0.77)	37 (0.79)	36 (0.77)	>0.99
		Some pain/discomfort	10 (0.21)	10 (0.21)	10 (0.21)	
		A lot of pain/discomfort	1 (0.02)		1 (0.02)	
	Feeling worried, sad or unhappy	Not worried, sad or unhappy	38 (0.81)	31 (0.66)	30 (0.64)	0.26
		A little worried, sad or unhappy	9 (0.19)	15 (0.32)	16 (0.34)	
		Very worried, sad or unhappy		1 (0.02)	1 (0.02)	
Boy (N=37)	**Walking about**	No problems	36 (0.97)	37 (1)	36 (0.97)	>0.99
		Some problems	1 (0.03)		1 (0.03)	
		A lot of problems				
	Looking after myself	No problems	35 (0.95)	34 (0.92)	35 (0.95)	>0.99
		Some problems	2 (0.05)	3 (0.08)	2 (0.05)	
		A lot of problems				

Variable	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 p-value$^{\#}$
	Doing usual activities	No problems	30 (0.81)	34 (0.92)	35 (0.95)	0.23
		Some problems	7 (0.19)	3 (0.08)	2 (0.05	
		A lot of problems				
	Having pain/discomfort	No pain/discomfort	31 (0.84)	27 (0.73)	32 (0.86)	0.40
		Some pain/discomfort	6 (0.16)	10 (0.27)	5 (0.14)	
		A lot of pain/discomfort				
	Feeling worried, sad or unhappy	Not worried, sad or unhappy	35 (0.95)	23 (0.62)	27 (0.73)	<0.01*
		A little worried, sad or unhappy	2 (0.05)	14 (0.38)	10 (0.27)	
		Very worried, sad or unhappy				

P-Value from omnibus Fisher's Exact Test for comparison among self, proxy-proxy, and proxy-child reported HRQoL; *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

4.4.3 Inter-rater agreement for EQ-5D-Y-3L values and dimensions

Table 4.4 presents the dyad agreement for overall HRQoL and across dimensions by rater and sub-groups along with the 95% 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 proxy-child report was considered. Similarly, a poor self and proxy-child agreement was observed in the dyad comprising children without any health condition. In contrast to the poor agreement noted between boys and parents, from both perspectives, a significant agreement was observed with girls and parent proxy-proxy perspective. However, the proxy-child perspective for girls only showed a non-significant agreement.

The dimension level agreement ranged between good and very good, with AC_1 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, whilst proxy-child report provided a higher (good) agreement estimate (0.7).

Inter-rater agreement was mostly consistent across both sub-groups (as categorised by age-groups, presence of a long-term health condition and gender) 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 “looking after myself” (both=0.54). In comparison, very good agreement was observed within the same dimension in the 11–12-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 8–10-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 proxy-proxy report, and in the dimension "having pain or discomfort" (0.56) with proxy-child report. In contrast, for children without any reported health conditions, agreement levels ranged between good and very good across all dimensions.

For the gender subgroup, the agreement typically ranged from good ("looking after myself", "doing usual activities", "having pain/discomfort", "feeling worried, sad or unhappy") and very good ("walking about") across both perspectives. The only exception was the "feeling worried, sad or unhappy" for the boy-parent dyads, where the proxy-proxy perspective resulted in a lower agreement (0.53, moderate) compared to proxy-child perspective (0.75, good).

Table 4-4 Agreement overall and by subgroup: EQ-5D-Y-3L values and dimension level agreement (Proxy-proxy and Proxy-child).

Variable	Dimension/ HRQoL values	EQ-5D-Y-3L (Self and Proxy-proxy) AC1 (95% CI)	EQ-5D-Y-3L (Self and Proxy-child) AC1 (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)

Variable	Dimension/ HRQoL values	EQ-5D-Y-3L (Self and Proxy-proxy) AC1 (95% CI)	EQ-5D-Y-3L (Self and Proxy-child) AC1 (95% CI)
	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)
Age group			
6 to 7 yrs. (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 yrs. (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 yrs. (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)
Health condition			

Variable	**Dimension/ HRQoL values**	**EQ-5D-Y-3L (Self and Proxy-proxy)** **AC1 (95% CI)**	**EQ-5D-Y-3L (Self and Proxy-child)** **AC1 (95% CI)**
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)
Gender			
Girl (N=47)	**Walking about**	0.82 (0.68, 0.95)	0.82 (0.68, 0.95)
	Looking after myself	0.74 (0.58, 0.89)	0.94 (0.86, 1)
	Doing usual activities	0.71 (0.55, 0.87)	0.76 (0.62, 0.91)
	Having pain/discomfort	0.66 (0.49, 0.84)	0.64 (0.45, 0.82)
	Feeling worried, sad or unhappy	0.63 (0.45, 0.81)	0.68 (0.51, 0.85)
CCC (95% CI)	**EQ-5D-Y-3L values**	0.33 (0.06, 0.56)	0.27 (-0.02, 0.51)
Boy (N=37)	**Walking about**	0.97 (0.92, 1)	0.94 (0.86, 1)
	Looking after myself	0.86 (0.72, 0.99)	0.89 (0.77, 1)
	Doing usual activities	0.69 (0.51, 0.88)	0.73 (0.55, 0.91)

Variable	Dimension/ HRQoL values	EQ-5D-Y-3L (Self and Proxy-proxy) AC1 (95% CI)	EQ-5D-Y-3L (Self and Proxy-child) AC1 (95% CI)
	Having pain/discomfort	0.72 (0.54, 0.9)	0.67 (0.48, 0.87)
	Feeling worried, sad or unhappy	0.53 (0.3, 0.76)	0.75 (0.58, 0.92)
CCC (95% CI)	**EQ-5D-Y-3L values**	0.11 (-0.19, 0.4)	0.17 (-0.15, 0.45)

Altman's scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good; EQ-5D-Y-3L proxy-proxy = Proxy version 1 report, EQ-5D-Y-3L proxy-child = Proxy version 2 report

4.4.4 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 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-value=0.02).

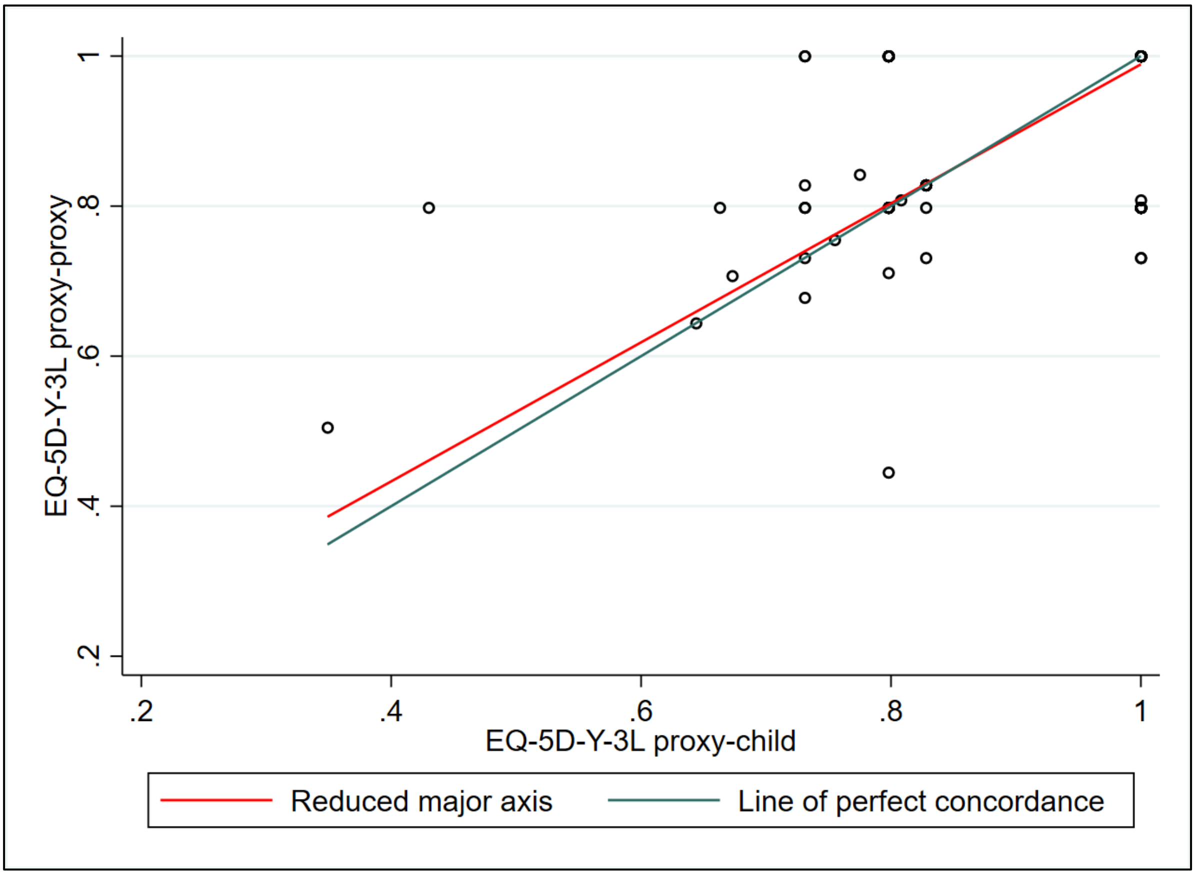


Figure 4.2 Intra-proxy gap in agreement between Proxy-proxy and Proxy-child versions of the EQ-5D-Y-3L.

EQ-5D-Y-3L proxy-proxy = Proxy version 1 report, EQ-5D-Y-3L proxy-child = Proxy version 2 report

Figure 4.2 shows the intra-proxy gap between the two proxy versions of the EQ-5D-Y-3L. As illustrated in Table 4.5, 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).

Table 4-5 Overall and dimension level intra-proxy agreement (Proxy 1 and Proxy 2) EQ-5D-Y-3L.

	EQ-5D-Y-3L Values	**Walking about**	**Looking after myself**	**Doing usual activities**	**Having pain/discomfort**	**Feeling worried, sad or unhappy**

	CCC (95% CI)	AC_1 (95% CI)	AC_1 (95% CI)	AC_1 (95% CI)	AC_1 (95% CI)	AC_1 (95% CI)
Overall (N =85)	0.53 (0.35, 0.66)	0.99 (0.96, 1)	0.95 (0.9, 1)	0.95 (0.9, 1)	0.64 (0.75, 0.93)	0.59 (0.45, 0.72)

Altman's scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good; EQ-5D-Y-3L Proxy 1= proxy-proxy version, EQ-5D-Y-3L Proxy 2= proxy-child version

4.4.5 Sensitivity analysis

(See Chapter 3, section 3.3.1, paragraph 2)

The findings using the German EQ-5D-Y-3L value-set indicate similar inter-rater agreements in terms of overall HRQoL for both the sample and across the subgroups. The agreement was 0.29 (0.08, 0.47) for proxy-proxy report and 0.25 (0.04, 0.44) for proxy-child report in the overall sample. The lowest agreement (poor) was noted for children aged 11-12 years their parents from both perspectives (proxy-proxy=0.04 (-0.29, 0.36), proxy-child=0.06 (-0.26, 0.38). 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 with no significant differences between the child and proxy reports (See Supplementary information, Table 4.6 for more details).

(See Chapter 3, section 3.8.1, paragraph 3)

The dimension level agreement estimates (AC_2), calculated using linear weights, were higher in comparison to the non-weighted estimates (AC_1). This discrepancy arises because linear weighting accounts for the degree of disagreement between categories, assigning less penalty to disagreements between adjacent categories and more penalty to those between distant categories [161]. As a result, the weighted agreement ranged from good to very good across all dimensions and both perspectives (see Supplementary information, Table 4.7 for more details), potentially overestimating the inter-rater agreement.

4.5 Discussion

This chapter 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 date, 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. The results of this 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). This study also found when proxy-child 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 [243, 244]. In this 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 proxy-proxy perspective were lower relative to the 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 [101]. Furthermore, discrepancies between child and parent assessments may 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 [2]. 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 [171, 173, 174, 177, 182], 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 agreement decreases with age in a sample of children aged 8-18 years [216]. The findings in this study 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 of age 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 PedsQL^{TM}, when assessed within the same sample in another study (see Chapter 5) [269]. 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 [161, 201]. On the other hand, AC_1 relies on the percent agreement and chance-corrected agreement between the raters [161] and may be more informative due to its disaggregated approach.

The oldest age-group also yielded a statistically significant difference in the EQ VAS scores when proxy-child perspective was adopted. A statistically significant difference in the self and proxy EQ VAS scores was reported in a study by Jelsma and Ramma involving school children using the EQ-5D-Y-3L proxy-child report [172]. 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 EQ VAS. Research has also indicated that children under the age of 7 may lack the conceptual ability to use the EQ VAS [53]. However, no significant difference in the self and proxy EQ VAS scores was observed in this age group (6-7 years).

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 [270]. This has been attributed to either the inability of young children to accurately self-report or differences in the interpretation of the same construct [271]. For example, children may consider themselves too young to dress themselves 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 (examined in Chapters 6 and 7). 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 [176, 272, 273]. 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 [182, 222]. 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. 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 given the presence of a large 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. This may provide further evidence to inform guidance around the minimum age for child self-reporting of their own HRQoL using the EQ-5D-Y-3L measure as well as guide the choice on self-report versus proxy perspective where either are possible [146].

4.6 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 [148, 274]. Given that the EQ-5D-Y-3L valuation set for Australia is not yet available and the aim is not to assess the HRQoL of children in this sample, the Australian EQ-5D-3L weights were used to calculate self and proxy child values [243]. 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, a child-specific Australian value set for the EQ-5D-Y-3L could weigh the dimensions differently than the adult value set. For instance, child-specific value sets for the EQ-5D-Y-3L have been observed to have different orders of dimension importance compared to corresponding adult value sets in the same country [88]. 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. This limitation will be addressed in detail in Chapter 8. Additionally, despite the distraction task of completing the EQ-5D-3L for themselves in between each proxy task, given that proxy-child report was completed subsequent to 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 [275] 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 underrepresentation 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 and the 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.

4.7 Conclusions

Overall, for 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 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.

4.8 Supplementary information

Table 4-6 Description of EQ-5D-Y-3L (Proxy-proxy and Proxy-child) German value set preference-weighted health states: overall and based on sub-groups.

Variable	EQ-5D-Y-3L (Self-report)	EQ-5D-Y-3L (Proxy-proxy)	EQ-5D-Y-3L (Proxy-child)
Overall			
N (%)	85	85	85
Median (IQR)	1 (0.87-1)	0.98 (0.87-1)	1 (0.87-1)
Self vs proxy difference (p-value*)		0.5	0.66

Variable	**EQ-5D-Y-3L (Self-report)**	**EQ-5D-Y-3L (Proxy-proxy)**	**EQ-5D-Y-3L (Proxy-child)**
CCC		0.29 (0.08, 0.47)	0.25 (0.04, 0.44)
Age group			
6 to 7 yrs.:			
N (%)	23 (0.27)	23 (0.27)	23 (0.27)
Median (IQR)	0.92 (0.89-1)	1 (0.87-1)	0.98 (0.87-1)
Self vs proxy difference (p-value*)		0.86	0.9
CCC		0.23 (-0.17, 0.56)	0.07 (-0.34, 0.46)
8 to 10 yrs.:			
N (%)	30 (0.35)	30 (0.35)	30 (0.35)
Median (IQR)	1 (0.85-1)	0.91 (0.12)	0.9 (0.13)
Self vs proxy difference (p-value*)		0.77	1
CCC		0.42 (0.08, 0.67)	0.41 (0.06, 0.66)
11 to 12 yrs.:			
N (%)	32 (0.38)	32 (0.38)	32 (0.38)
Median (IQR)	1 (0.9-1)	0.96 (0.87-1)	1 (0.87-1)
Self vs proxy difference (p-value*)		0.38	0.45
CCC		0.04 (-0.29, 0.36)	0.06 (-0.26, 0.38)

Variable	EQ-5D-Y-3L (Self-report)	EQ-5D-Y-3L (Proxy-proxy)	EQ-5D-Y-3L (Proxy-child)
Long-term health condition/s			
No:			
N (%)	59 (0.69)	59 (0.69)	59 (0.69)
Median (IQR)	1 (0.87-1)	1 (0.87-1)	1 (0.87-1)
Self vs proxy difference (p-value*)		0.71	0.55
CCC		0.17 (-0.08, 0.4)	0.18 (-0.08, 0.41)
Yes:			
N (%)	26 (0.31)	26 (0.31)	26 (0.31)
Median (IQR)	1 (0.87-1)	0.87 (0.76-1)	0.87 (0.78-1)
Self vs proxy difference (p-value*)		0.1	0.1
CCC		0.34 (-0.04, 0.62)	0.26 (-0.12, 0.57)
Gender			
Girls:			
N (%)	47 (0.55)	47 (0.55)	47 (0.55)
Median (IQR)	0.98 (0.87-1)	1 (0.87-1)	1 (0.87-1)
Self vs proxy difference (p-value*)		0.91	0.92
CCC		0.29 (0.01, 0.53)	0.23 (-0.05, 0.48)

Variable	**EQ-5D-Y-3L (Self-report)**	**EQ-5D-Y-3L (Proxy-proxy)**	**EQ-5D-Y-3L (Proxy-child)**
Boys:			
N (%)	37 (0.44)	37 (0.44)	37 (0.44)
Median (IQR)	1 (0.92-1)	1 (0.87-1)	1 (0.87-1)
Self vs proxy difference (p-value*)		0.18	0.27
CCC		0.25 (-0.05, 0.5)	0.24 (-0.04, 0.49)
CCC (Intra-proxy)		0.52 (0.35, 0.66)	

*p-value of Wilcoxon matched-pair signed-rank test; CCC= Concordance Correlation Coefficient; EQ-5D-Y-3L proxy-proxy = Proxy version 1 report, EQ-5D-Y-3L proxy-child = Proxy version 2 report

Table 4-7 Weighted Gwet's AC2 (linear weights) agreement overall and by subgroup: EQ-5D-Y-3L (Proxy-proxy and Proxy-child) values and dimension level agreement.

Variable	**Dimension**	**EQ-5D-Y-3L (Self and Proxy-proxy) AC1 (95% CI)**	**EQ-5D-Y-3L (Self and Proxy-child) AC1 (95% CI)**
Overall (N=85)	**Walking about**	0.93 (0.88, 0.98)	0.92 (0.87, 0.97)
	Looking after myself	0.88 (0.82, 0.94)	0.9 (0.84, 0.95)
	Doing usual activities	0.83 (0.76, 0.9)	0.86 (0.79, 0.92)
	Having pain/discomfort	0.81 (0.73, 0.89)	0.77 (0.68, 0.87)
	Feeling worried, sad or unhappy	0.75 (0.66, 0.84)	0.82 (0.73, 0.9)

Variable	Dimension	EQ-5D-Y-3L (Self and Proxy-proxy) AC1 (95% CI)	EQ-5D-Y-3L (Self and Proxy-child) AC1 (95% CI)
Age group			
6 to 7 yrs. (N=23)	**Walking about**	0.87 (0.73, 1)	0.87 (0.73, 1)
	Looking after myself	0.73 (0.55, 0.92)	0.73 (0.55, 0.92)
	Doing usual activities	0.66 (0.45, 0.86)	0.7 (0.5, 0.89)
	Having pain/discomfort	0.86 (0.73, 1)	0.79 (0.63, 0.96)
	Feeling worried, sad or unhappy	0.76 (0.58, 0.95)	0.79 (0.63, 0.96)
8-10 yrs. (N=30)	**Walking about**	0.95 (0.88, 1)	0.93 (0.85, 1)
	Looking after myself	0.9 (0.81, 1)	0.95 (0.88, 1)
	Doing usual activities	0.83 (0.7, 0.96)	0.86 (0.74, 0.97)
	Having pain/discomfort	0.78 (0.61, 0.94)	0.77 (0.58, 0.96)
	Feeling worried, sad or unhappy	0.77 (0.63, 0.92)	0.82 (0.69, 0.96)
11-12 yrs. (N=32)	**Walking about**	0.95 (0.89, 1)	0.95 (0.89, 1)
	Looking after myself	0.95 (0.89, 1)	0.95 (0.89, 1)

Variable	**Dimension**	**EQ-5D-Y-3L (Self and Proxy-proxy) AC1 (95% CI)**	**EQ-5D-Y-3L (Self and Proxy-child) AC1 (95% CI)**
	Doing usual activities	0.93 (0.86, 1)	0.95 (0.89, 1)
	Having pain/discomfort	0.8 (0.67, 0.93)	0.76 (0.62, 0.91)
	Feeling worried, sad or unhappy	0.73 (0.57, 0.88)	0.83 (0.68, 0.97)
Health condition			
No (N=59)	**Walking about**	0.93 (0.86, 0.99)	0.91 (0.85, 0.98)
	Looking after myself	0.88 (0.81, 0.95)	0.9 (0.84, 0.97)
	Doing usual activities	0.87 (0.79, 0.94)	0.87 (0.79, 0.94)
	Having pain/discomfort	0.83 (0.74, 0.92)	0.81 (0.72, 0.91)
	Feeling worried, sad or unhappy	0.78 (0.68, 0.88)	0.83 (0.74, 0.92)
Yes (N=26)	**Walking about**	0.94 (0.86, 1)	0.94 (0.86, 1)
	Looking after myself	0.89 (0.78, 1)	0.89 (0.78, 1)
	Doing usual activities	0.73 (0.55, 0.91)	0.83 (0.7, 0.97)
	Having pain/discomfort	0.75 (0.56, 0.94)	0.67 (0.43, 0.91)

Variable	Dimension	EQ-5D-Y-3L (Self and Proxy-proxy) AC1 (95% CI)	EQ-5D-Y-3L (Self and Proxy-child) AC1 (95% CI)
	Feeling worried, sad or unhappy	0.68 (0.49, 0.87)	0.78 (0.59, 0.97)
Gender			
Girl (N=47)	**Walking about**	0.89 (0.8, 0.98)	0.89 (0.8, 0.98)
	Looking after myself	0.86 (0.77, 0.95)	0.94 (0.88, 1)
	Doing usual activities	0.84 (0.74, 0.94)	0.87 (0.79, 0.96)
	Having pain/discomfort	0.79 (0.67, 0.91)	0.75 (0.62, 0.89)
	Feeling worried, sad or unhappy	0.74 (0.59, 0.88)	0.79 (0.67, 0.91)
Boy (N=37)	**Walking about**	0.99 (0.96, 1)	0.97 (0.93, 1)
	Looking after myself	0.92 (0.85, 1)	0.94 (0.88, 1)
	Doing usual activities	0.83 (0.72, 0.94)	0.85 (0.75, 0.96)
	Having pain/discomfort	0.85 (0.74, 0.95)	0.81 (0.69, 0.93)
	Feeling worried, sad or unhappy	0.74 (0.59, 0.88)	0.86 (0.76, 0.96)

Altman's scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good; EQ-5D-Y-3L proxy-proxy = Proxy version 1 report, EQ-5D-Y-3L proxy-child = Proxy version 2 report

CHAPTER 5: CHILD-PARENT AGREEMENT IN THE ASSESSMENT OF HEALTH-RELATED QUALITY OF LIFE USING THE CHU9D AND THE PEDSQL$^{\text{TM}}$.

This chapter contains material from:

“Khanna, D., Khadka, J., Mpundu-Kaambwa, C., Ratcliffe, J., & Quality of Life in Kids: Key Evidence to Strengthen Decisions in 6 Australia (QUOKKA) Project Team (2023). Child-Parent Agreement in the Assessment of Health-Related Quality of Life Using the CHU9D and the PedsQLTM. Applied health economics and health policy, 21(6), 937–947. https://doi.org/10.1007/s40258-023-00831-7”

5.1 Overview

This chapter presents the following results of Study sample 1 for the 85 dyads who completed the respective self and proxy reports using the CHU9D and the PedsQL$^{\text{TM}}$ measures:

1) Parent-child agreement in reporting of children’s HRQoL (aged 6-12 years) using the CHU9D (a preference-based measure of children’s HRQOL) and the PedsQL$^{\text{TM}}$ (a non-preference-based measure of children’s HRQOL),

2) Impact of age on child-parent agreement across the dimensions of the two measures.

5.2 Summary

Objective: This study examined the inter-rater agreement between child-self and parental proxy HRQoL ratings (overall and dimension level) using two different generic child-specific measures, the CHU9D and the PedsQL$^{\text{TM}}$, in a community-based sample of Australian children. A secondary objective was to investigate the impact of age on child-parent agreement across the dimensions of the two measures.

Methods: 85 child-parent dyads (children aged 6-12 years) recruited from the community completed the self and proxy versions (proxy-proxy perspective only) of the CHU9D and the PedsQL^{TM}, respectively. The inter-rater agreement was estimated using CCC and Gwet's AC_1 for the overall sample and across age-groups. To assess the differences between the self and the proxy HRQoL reports, Wilcoxon matched-pair signed-rank test was used (please refer to Chapter 3, section 3.3 onwards for specific details).

Results: Agreement was low for overall HRQoL for both the CHU9D (CCC= 0.28) and the PedsQL^{TM} (CCC= 0.39). Across the CHU9D dimensions, agreement was the highest for "sad" (AC_1= 0.83) and lowest for "tired" (AC_1= 0.31). The PedsQL^{TM} demonstrated stronger agreement (AC_1= 0.41 to 0.6) for the physical health dimension but weaker for the psychosocial dimensions (AC_1<0.4). Except for the "tired" dimension, agreement was consistent across age-groups with the CHU9D, whilst the PedsQL^{TM} showed poor agreement for most of the psychosocial health items among the older age-groups only (8-10 and 11-12 years). No significant differences were reported between self and proxy reported HRQoL.

Conclusions: This study highlights that the agreement between child and parent reported HRQoL may be influenced by both the measure used and the age of the child. These findings may have implications for the economic evaluation of healthcare interventions and services in child populations when both child and proxy perspectives are considered in the assessment of child HRQoL.

5.3 Key Points

- Child-parent agreement at the individual dimension level was higher for CHU9D than for PedsQL™. In contrast, agreement for overall HRQoL was lower for CHU9D relative to the PedsQL™.
- In general, younger children (6-7 years) reported comparable agreement with parental proxies to their older counterparts providing some evidence to indicate that they may be able to meaningfully self-report.

5.4 Results

5.4.1 Sample characteristics

Child-parent participant characteristics are the same as presented in Chapter 4 results (see section 4.4.1).

5.4.2 Child-parent difference in reported HRQoL and overall concordance

Table 5.1 describes the child and parent reported HRQoL scores and the dyad agreement using the CHU9D and the PedsQL™ respectively. Overall, parents underreported child's HRQoL with the CHU9D but overreported with the PedsQL™. Median difference across the age groups was the largest for ages 11–12 years with the CHU9D and ages 6-7 years with the PedsQL™. However, these differences were not found to be statistically significant.

The overall agreement between child-parent dyads for both measures was fair with a lower agreement for CHU9D (0.28) ***(Fig. 5.1a)*** than for the PedsQL™ (0.39) ***(Fig. 5.1b)***. The agreement between parents and 8-10-year-olds was good for both measures. For overall HRQoL, this was the only age group that demonstrated a statistically significant level of agreement across both measures.

Overall, the largest median difference in HRQoL ratings between children and proxies, although not statistically significant, was observed in the subgroup of children with a reported health condition using the CHU9D. The PedsQL™ also demonstrated a notable inter-rater gap within this subgroup. However, the subsequent agreement between child-parent dyads observed in this subgroup was higher with both measures compared to those with no reported health condition. A contrasting pattern of agreement was observed across the gender subgroup for the two measures. Compared to girls, boy-parent dyads exhibited a lower agreement with CHU9D (non-significant, CCC=0.16) but higher agreement with the PedsQL™ (significant, CCC=0.72) (see Supplementary information, Table 5.3 for more details).

Table 5-1 Description of child and proxy reported HRQoL values and agreement using CHU9D and the PedsQL™ 4.0 generic core scales.

	Child			Parent			Agreement	
	N	Mean (SD)	Median (IQR)	N	Mean (SD)	Median (IQR)	N	CCC (95% CI)
CHU9D								
Overall	81	0.82 (0.16)	0.86 (0.22)	84	0.79 (0.16)	0.83 (0.21)	80	0.28 (0.07 ,0.47)
Age group								
6 to 7 yrs.:	22	0.81 (0.18)	0.86 (0.2)	22	0.82 (0.15)	0.87 (0.2)	21	-0.18 (-0.55 ,0.26)
8 to 10 yrs.:	29	0.79 (0.16)	0.81 (0.22)	30	0.77 (0.16)	0.8 (0.21)	29	0.69 (0.43 ,0.84)
11 to 12 yrs.:	30	0.84 (0.14)	0.89 (0.17)	32	0.79 (0.17)	0.84 (0.23)	30	0.19 (-0.16 ,0.5)
PedsQL™								
Overall	83	76.84 (13.49)	78.41 (15.46)	85	78.76 (12.53)	81.82 (15.76)	83	0.39 (0.2, 0.56)
Age group								

	Child			Parent			Agreement	
	N	Mean (SD)	Median (IQR)	N	Mean (SD)	Median (IQR)	N	CCC (95% CI)
6 to 7 yrs.:	22	79.41 (13.6)	82.61 (12.55)	23	86.07 (9.01)	86.96 (13.04)	22	0.02 (-0.31, 0.34)
8 to 10 yrs.:	30	73.11 (13.42)	75.57 (18.86)	30	75.69 (12.37)	78.98 (15.85)	30	0.67 (0.43, 0.83)
11 to 12 yrs.:	31	78.61 (13.11)	81.82 (13.64)	32	76.38 (13.04)	79.55 (18.18)	31	0.29 (-0.06, 0.58)

CCC= Concordance correlation coefficient

Altman’s scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good.

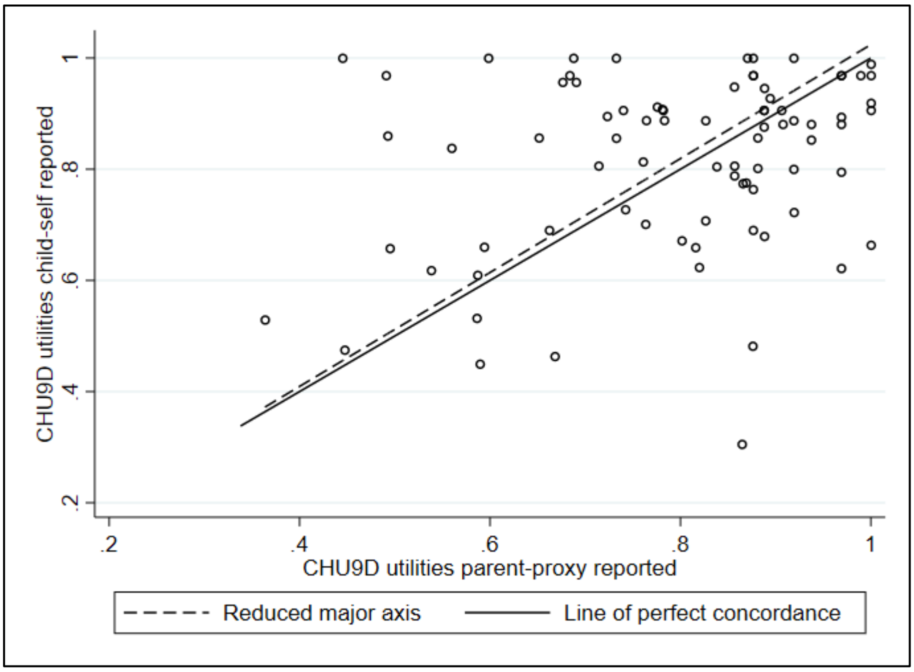


Figure 5.1a. Concordance between child and parent reported HRQoL utilities using the CHU9D.

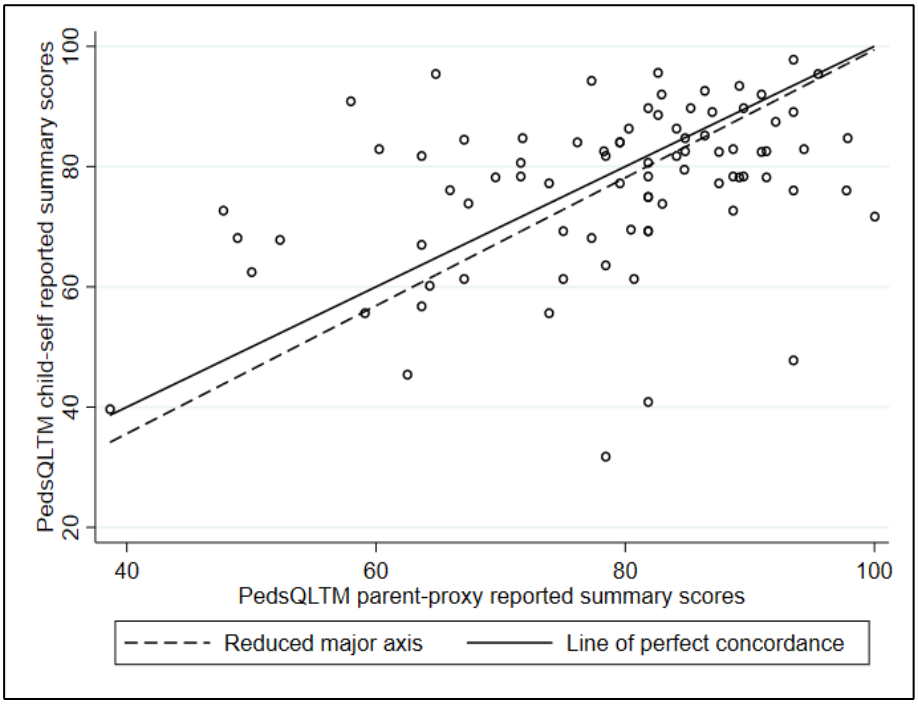


Figure 5.2b. Concordance between child and parent reported HRQoL scores using the PedsQL™.

5.4.3 *Comparison of agreement for CHU9D dimensions and PedsQL™ items:*

Table 5.2 presents the agreement coefficients (AC$_1$), for the CHU9D dimensions and the corresponding representative PedsQL™ items, for the overall sample and by age-group. Child-parent agreement ranged from 0.31 to 0.83 for the CHU9D dimensions and 0.15 to 0.52 for the relevant PedsQL™ items. The agreement was higher for CHU9D dimensions than for the corresponding PedsQL™ items. Among the dimensions related to subjective (internal) experiences, agreement was the highest for “sad” (CHU9D= 0.83) and “feeling sad” (PedsQL™= 0.37) within the respective measures. The agreement was high for “pain” (0.73) with the CHU9D, whereas its equivalent dimension in the PedsQL™ showed the lowest agreement (0.15) compared to all other dimensions within the measure. The weakest agreement across the CHU9D dimensions was observed for “tired” (0.31) followed by “worried” (0.45). In addition to the items related to the psychosocial health mentioned above, a poor agreement was also observed for the PedsQL™ item “having trouble sleeping” (0.16). For the physical functioning related dimensions, agreement ranging between moderate to good was observed with both, the CHU9D and the PedsQL™.

Across the age-groups, for the CHU9D dimensions, the only statistically non-significant agreement was observed between parents and children aged 6-7 years for “tired” (0.19). Moreover, for most dimensions, the agreement was lower for the 6–7-year-olds. In contrast, agreement across the majority of the relevant PedsQL™ items was higher for the youngest age-group (6-7 years) relative to the older age-groups (8-10 and 11-12 years). Furthermore, an insignificant agreement was observed for several items such as “having low energy level”, “feeling angry” and “having trouble sleeping” with both the older age-groups. They also demonstrated a poor agreement for the “getting aches and pain” item. Additionally, an insignificant agreement was also seen between parents and 11-12-year-olds for the “worrying what will happen to them” item.

As discussed in Chapter 4 (section 4.4.5), the dimension level agreement was higher for the sample and the age subgroup with linear weights (AC_2) relative to the estimates presented in Table 5.2 (see Supplementary information, Table 5.4 for more details).

Table 5-2 Comparison of child-parent agreement in CHU9D dimensions with relevant PedsQL™ items by age group.

CHU9D dimensions	Overall (N=85)	6 to 7 years (N=23)	8 to 10 years (N=30)	11 to 12 years (N=32)	Relevant PedsQL items	Overall (N=85)	6 to 7 years (N=23)	8 to 10 years (N=30)	11 to 12 years (N=32)
Worried	0.45	0.43	0.42	0.49	**2.1 Feeling afraid or scared**	0.32	0.45	0.36	0.23
					2.5 Worrying about what will happen to him or her	0.28	0.48	0.26	0.18$^{\#}$
Sad	0.83	0.86	0.89	0.76	**2.2 Feeling sad**	0.37	0.6	0.4	0.22
Pain	0.73	0.75	0.74	0.7	**1.7 Getting aches and pains**	0.15	0.27	0.1$^{\#}$	0.15$^{\#}$
Tired	0.31	0.19$^{\#}$	0.45	0.27	**1.8 Having a low energy level**	0.26	0.53	0.2$^{\#}$	0.13$^{\#}$
Annoyed	0.56	0.5	0.63	0.55	**2.3 Feeling angry**	0.2	0.53	0.1$^{\#}$	0.07$^{\#}$
School work/ homework	0.49	0.4	0.58	0.48	**4.3 Keeping up with schoolwork**	0.3	0.54	0.23	0.22
Sleep	0.54	0.6	0.46	0.58	**2.4 Having trouble sleeping**	0.16	0.37	0.13$^{\#}$	0.07$^{\#}$
Daily routine	0.52	0.44	0.55	0.55	**1.5 Taking a bath or**	0.52	0.52		

CHU9D dimensions	**Overall (N=85)**	**6 to 7 years (N=23)**	**8 to 10 years (N=30)**	**11 to 12 years (N=32)**	**Relevant PedsQL items**	**Overall (N=85)**	**6 to 7 years (N=23)**	**8 to 10 years (N=30)**	**11 to 12 years (N=32)**
(eating, having a bath/ Shower, getting dressed)					**shower by him or herself (N=22)**				
Able to join in activities **(playing out with friends, doing sports, joining things)**	0.63	0.66	0.58	0.66	**1.3 Participating in sports activity or exercise**	0.52	0.64	0.42	0.54
					3.1 Getting along with other children	0.43	0.65	0.44	0.27
					3.5 Keeping up when playing with other children	0.5	0.61	0.39	0.52

Agreement statistics estimated using AC_1.

Agreement coefficient not statistically significant at 95% confidence level

Altman’s scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good

(Results for the subgroups presence of long-term health condition and gender are discussed below briefly to keep the chapter focused on age groups.)

In general, agreement was higher for children without any health conditions and for boys across more dimensions/items for both measures. A similar pattern to the overall sample findings was observed across the subgroups of health condition and gender. Higher agreement within the dimensions was found with the CHU9D compared to the relevant PedsQL™ items. Very good agreement was observed for the “sad” dimension across both subgroups, and high agreement was also noted for the “pain” dimension (AC_1 estimate ranging from 0.71 to 0.74). The only non-significant agreement was for the “tired” dimension in the subgroup with a reported health condition (AC_1= 0.23). In contrast to the findings for the “pain” dimension of the CHU9D, the PedsQL™ item “getting aches and pains” showed poor and non-significant agreement across both subgroups. For more details, please refer to Supplementary information, Table 5.5.

5.5 Discussion

This study is the first, to date, to investigate child-parent agreement of child overall and dimension level HRQoL in a community-based sample of children using two generic HRQOL measures, the CHU9D and PedsQL™ 4.0. This study showed contrasting agreement for overall and dimension level HRQoL using the two measures. The agreement between parents and children for HRQoL scores was stronger for the PedsQL™, but weaker for the CHU9D. Conversely, agreement for the individual dimensions was stronger for the CHU9D compared to the PedsQL™ items.

The discrepancy in the consistency of agreement may be attributed, at least in part, to the statistical method used to measure the agreement. This study used two different methods to estimate agreement between the child and parent ratings: CCC for overall HRQoL and

Gwet's AC_1 for dimension level HRQoL. Utilities or summary scores combine responses from different dimensions to estimate the overall HRQoL of the child. However, when analysing inter-rater agreement, the dimension/item level responses can offer a more direct measurement of agreement as it provides the disaggregated evaluations of the two raters, i.e., the child and the parent. This may be more informative about the specific areas of agreement or disagreement between the child and the parent and, therefore, provide a better understanding of the concordance in evaluations of each aspect of HRQoL. Furthermore, the estimation of CCC in this study may have been affected by an increased level of variation in ratings resulting from the high number of rater pairs, which could have potentially led to an underestimation of the true magnitude of the CCC [161].

The inter-rater differences in HRQoL scores across age groups using both measures did not correspond with the trends in agreement observed at the individual dimension level. For instance, in comparison with the other age groups, the 11-12 years age group had the greatest inter-rater gap with the CHU9D utilities. However, the dimension level agreement was similar across age-groups. Additionally, while the same age-group had the smallest inter-rater difference with the PedsQL^{TM} summary scores, they demonstrated lower agreement levels across most of its items compared to the youngest age-group. Hence, it is important to acknowledge that the differences in the aggregated child and proxy reported HRQoL scores do not provide a measure of agreement [222].

Towards the opposite end of the age spectrum, a recent systematic review investigated the level of agreement between adult proxies and older adults with cognitive impairment [276]. Their findings indicated that there was some evidence suggesting higher levels of agreement in more observable HRQoL dimensions, such as physical health and mobility, compared to less observable dimensions like emotional well-being [276]. Typically, the available evidence indicates that parents also tend to be more concordant at reporting HRQoL dimensions related to the more easily observable attributes compared to those that are more subjective (internal) to the child [106, 222]. However, in this study, we found that with the CHU9D, a high level of agreement was obtained for the psychosocial health dimension “sad”. It is plausible that the responses for emotional states like ‘sad’ may be more consistent between self and proxy reports, possibly due to shared perceptions of distinct behavioural or emotional cues associated with sadness [277]. In contrast, dimensions like ‘sleep’ or ‘daily routine’ may reflect variability in interpretation or observation by proxies, leading to lower agreement. Nevertheless, recent study comparing the dimensionality of the EQ-5D-Y-5L, the HUI, the CHU9D and the PedsQL™ , also suggested that the CHU9D may be more suitable for assessing the emotional functioning aspect of HRQoL [96]. The agreement for physical health-related dimensions (“daily routine” and “able to join in activities”) was low but moderate. These findings contrasted with the agreement observed for similar PedsQL™ items. For example, agreement was higher for PedsQL™ physical health items, i.e., “participating in sports activity or exercise”and “taking a bath or shower by him or herself” as compared to the “feeling sad” item. Previous studies have reported a low agreement for pain using preference-based [180, 194, 195, 250, 278] and non-preference-based measures [100, 279]. In this study, a substantially higher agreement was observed for the “pain” dimension with the CHU9D as compared to the “getting aches and pains” item of the PedsQL™. Therefore, these findings suggest a possible interaction between the measure used and the dimension under consideration in determining the degree of agreement.

The findings in this study indicated a higher agreement for the CHU9D dimensions compared to the corresponding PedsQL™ items. Whilst both the measures were developed for use in children and adolescents in the development and validation of the instrument, the CHU9D followed a bottom-up approach that directly involved children in the development and validation of the instrument [91], whereas the PedsQL™ adopted a top-down approach and was developed based on a broader study of HRQoL in children with cancer [280]. The difference in agreement may also be attributed to the timeframe of assessment for each measure. In the CHU9D, respondents are asked about the (child's) health 'today' whereas the PedsQL^{TM} asks the respondent to report on their health over the 'past one month'. Thus, one possible explanation for the higher agreement found within the CHU9D dimensions may be its shorter time frame, which may reduce recall bias and result in less variability in perceived HRQoL [281]. Another contributing factor may be the difference in what the CHU9D and PedsQL^{TM} measures assess. The CHU9D measures the severity of impairment whereas the PedsQL^{TM} which measures frequency. For example, in the CHU9D dimension "sad", the response levels range from "don't feel sad" to "feel very sad", whilst the PedsQL^{TM} response levels for the corresponding item "feeling sad" range from "never" to "almost always"[281].

Studies reporting the level of child-parent agreement predominantly focus on samples including children aged 8 years and above [282-286]. The evidence for agreement in younger age-groups, e.g., 6-7 years old and capable of self-reporting their HRQoL using the PedsQL^{TM} or the CHU9D is limited [126, 218]. In this study, dyads comprising the youngest age-group (6-7 years) reported relatively lower agreement with the CHU9D. This may be owing to children in this age-group differing in their understanding of HRQoL as compared to their parents [287]. Younger children under 10 years of age have been reported to have difficulties with comprehension and recall of health-related events, as well as the associated frequency and severity [287]. However, except for the "tired" dimension, there was no clear association between age and agreement across any other CHU9D dimensions. In contrast to the CHU9D findings, the older age-groups, particularly the 11-12-year-olds, showed worse agreement for the PedsQL^{TM} items compared to the youngest one comprising 6-7-year-olds. The evidence in the literature examining the relationship between age of the child and agreement using both preference and non-preference-based measures is inconsistent [126, 216, 222, 280, 288, 289]. This study found conflicting results in the same population for the two measures. The reasons for these discrepancies are unclear. Further research including mixed methods studies, that combine quantitative investigations with in-depth qualitative research using cognitive interviewing techniques, for example think aloud may be helpful in providing a more detailed understanding of the reasons for these discrepancies in reporting child HRQoL [290].

The existing literature on the influence of health status of the child on agreement is inconsistent for both preference and non-preference-based measures [222, 223]. Some studies suggest that in chronic illnesses, greater severity of the disease [291] or a higher frequency of exacerbations [292] may be associated with higher levels of child-parent agreement. However, for chronic conditions like cancer, there is a lack of consensus regarding the degree of agreement [106, 222]. Conversely, acute illnesses have been associated with lower inter-rater agreement [176]. Considering that the literature is inconclusive, and the limited sample size of these studies, further research with a larger sample size is warranted to substantiate these findings.

5.6 Limitations

This study has limitations that are important to highlight. The study was conducted in a community-based sample of South Australian children who were relatively healthy. Hence, the findings may not be generalisable to more diverse samples including children with regular contact with health services and children with disabilities. Whilst the study sample was relatively small, good representation was achieved across age groups and approximately one-third of children were living with health conditions and/or living in areas of relative disadvantage. However, the main findings, particularly in relation to age-group analyses need to be interpreted with caution and further research needs to be conducted to substantiate these findings in larger community based and patient samples. The CHU9D preference weights employed in this study were established using adolescents aged 11-17 years and then applied to a sample that included a younger age group. It is recognised that the value sets derived from children/adolescents may differ from those derived from adults adopting a child's perspective [114]. Nevertheless, additional research is required to determine the youngest age at which children can provide valuations, taking into account ethical considerations, and to explore the potential impact of this on valuing child HRQoL across different age groups. Moreover, as the preference weights were used to estimate the CHU9D utilities, an additional preference weighted step not currently available for the PedsQL™ this makes score comparisons between the two measures difficult. Finally, the study investigated agreement between child-parent dyads using the CHU9D and PedsQL™ measures only and hence the findings may not necessarily be generalisable to other measures for measuring HRQoL in child populations.

5.7 Conclusions

This study a low found child-parent agreement for overall HRQoL across both measures, with CHU9D exhibiting a lower agreement relative to the PedsQL™. In contrast, at the individual dimension level, inter-rater agreement was higher for CHU9D than for PedsQL™. CHU9D showed the highest agreement with the dimensions of "sad" and "pain", whereas for the PedsQL™, agreement was the highest for the physical health items. There was no clear interaction between age and CHU9D dimensions. However, for the relevant PedsQL™ items, the dimension level agreement was stronger for the youngest children (6-7 years) in the sample and weaker for older children (8-10 and 11-12 years), particularly for the psychosocial health items. Further research in larger and more diverse study samples and across age groups is needed to substantiate these findings. The introduction of a preference-based scoring algorithm for the PedsQL™ will also facilitate empirical comparisons of child parental agreement at overall utility level and enable the impact of child and parent perspectives on HRQoL benefits for economic evaluations of interventions targeted at paediatric populations to be assessed.

5.8 Supplementary information

Table 5-3 Description of child and proxy reported HRQoL values and agreement using CHU9D and the PedsQL™ 4.0 generic core scales by presence/absence of health condition and child gender.

	Child			Parent			Agreement	
	N	**Mean (SD)**	**Median (IQR)**	**N**	**Mean (SD)**	**Median (IQR)**	**N**	**CCC (95% CI)**
CHU9D								
Long-term condition:								
No	56	0.83 (0.15)	0.87 (0.21)	58	0.8 (0.15)	0.85 (0.19)	55	0.21 (-0.05 ,0.45)
Yes	25	0.79 (0.17)	0.86 (0.22)	26	0.76 (0.18)	0.79 (0.22)	25	0.37 (-0.01 ,0.66)
Gender:								
Female	45	0.80 (0.17)	0.86 (.22)	47	0.78 (0.17)	0.82 (0.24)	45	0.34 (0.06, 0.57)
Male	35	0.83 (0.14)	0.88 (0.25)	36	0.81 (0.15)	0.87 (0.17)	35	0.16 (-0.17, 0.46)
PedsQL™								
Long-term condition:								
No	57	77.23 (12.72)	78.41 (13.04)	59	79.42 (11.16)	81.82 (15.91)	57	0.28 (0.03, 0.5)
Yes	26	75.98 (15.28)	81.82 (23.86)	26	77.26 (15.34)	81.82 (16.9)	26	0.55 (0.21, 0.77)

	Child			Parent			Agreement	
	N	**Mean (SD)**	**Median (IQR)**	**N**	**Mean (SD)**	**Median (IQR)**	**N**	**CCC (95% CI)**
Gender:								
Girl	47	74.75 (15.08)	77.27 (20.89)	47	77.35 (12.97)	78.41 (20.45)	47	0.21 (-0.07, 0.46)
Boy	36	79.56 (10.68)	82.21 (10.25)	37	80.56 (12.05)	81.82 (10.23)	36	0.72 (0.53, 0.85)

CCC= Concordance correlation coefficient

Altman's scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good**.**

Table 5-4 Comparison of child-parent agreement in CHU9D dimensions with relevant PedsQL™ items by age group using linear weights (AC2).

CHU9D dimensions	Overall (N= 85)	6 to 7 years (N=23)	8 to 10 years (N=30)	11 to 12 years (N=32)	Relevant PedsQL items	Overall (N= 85)	6 to 7 years (N= 23)	8 to 10 years (N= 30)	11 to 12 years (N= 32)
Worried	0.76	0.72	0.78	0.79	**2.1 Feeling afraid or scared**	0.56	0.58	0.69	0.55
					2.5 Worrying about what will happen to him or her	0.51	0.63	0.48	0.55
Sad	0.95	0.96	0.97	0.93	**2.2 Feeling sad**	0.61	0.76	0.68	0.56
Pain	0.89	0.88	0.88	0.91	**1.7 Getting aches and pains**	0.45	0.48	0.43	0.6

CHU9D dimensions	**Overall (N= 85)**	**6 to 7 years (N=23)**	**8 to 10 years (N=30)**	**11 to 12 years (N=32)**	**Relevant PedsQL items**	**Overall (N= 85)**	**6 to 7 years (N= 23)**	**8 to 10 years (N= 30)**	**11 to 12 years (N= 32)**
Tired	0.65	0.58	0.67	0.68	**1.8 Having a low energy level**	0.53	0.68	0.55	0.51
Annoyed	0.83	0.79	0.84	0.84	**2.3 Feeling angry**	0.51	0.67	0.55	0.47
School work/ homework	0.79	0.77	0.81	0.78	**4.3 Keeping up with schoolwork**	0.58	0.69	0.57	0.58
Sleep	0.83	0.88	0.78	0.85	**2.4 Having trouble sleeping**	0.46	0.52	0.47	0.51
Daily routine (eating, having a bath/ Shower, getting dressed)	0.80	0.76	0.84	0.79	**1.5 Taking a bath or shower by him or herself (N=22)**	0.65	0.65		
Able to join in activities (playing out with friends, doing sports, joining things)	0.85	0.85	0.82	0.88	**1.3 Participating in sports activity or exercise**	0.75	0.75	0.73	0.78
					3.1 Getting along with other children	0.67	0.79	0.71	0.62
					3.5 Keeping up when	0.7	0.71	0.68	0.73

CHU9D dimensions	Overall (N= 85)	6 to 7 years (N=23)	8 to 10 years (N=30)	11 to 12 years (N=32)	Relevant PedsQL items	Overall (N= 85)	6 to 7 years (N= 23)	8 to 10 years (N= 30)	11 to 12 years (N= 32)
					playing with other children				

Agreement statistics estimated using weighted (linear) AC2.

Agreement coefficient not statistically significant at 95% confidence level

Altman’s scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good

Table 5-5 Comparison of child-parent agreement in CHU9D dimensions with relevant PedsQL™ items by presence/absence of health condition and child gender.

CHU9D dimensions	Long term health condition		Gender		Relevant PedsQL items	Long term health condition		Gender	
	No (N=59)	Yes (N=26)	Girl (N=47)	Boy (N=37)		No (N=59)	Yes (N=26)	Girl (N=47)	Boy (N=37)
Worried	0.47	0.4	0.5	0.39	**2.1 Feeling afraid or scared**	0.27	0.44	0.26	0.4
					2.5 Worrying about what will happen to him or her	0.34	0.13#	0.19	0.39
Sad	0.89	0.71	0.81	0.86	**2.2 Feeling sad**	0.43	0.24#	0.33	0.44
Pain	0.72	0.74	0.74	0.71	**1.7 Getting aches and pains**	0.14#	0.17#	0.14#	0.16#

CHU9D dimensions	**Long term health condition**		**Gender**		**Relevant PedsQL items**	**Long term health condition**		**Gender**	
	No (N=59)	**Yes (N=26)**	**Girl (N=47)**	**Boy (N=37)**		**No (N=59)**	**Yes (N=26)**	**Girl (N=47)**	**Boy (N=37)**
Tired	0.35	0.23#	0.3	0.33	**1.8 Having a low energy level**	0.23	0.32	0.11#	0.45
Annoyed	0.58	0.53	0.52	0.62	**2.3 Feeling angry**	0.27	0.03#	0.15#	0.27
School work/ homework	0.46	0.57	0.47	0.53	**4.3 Keeping up with schoolwork**	0.31	0.3	0.31	0.3
Sleep	0.57	0.47	0.54	0.55	**2.4 Having trouble sleeping**	0.19	0.1#	0.18	0.14#
Daily routine (eating, having a bath/ Shower, getting dressed)	0.56	0.43	0.47	0.58	**1.5 Taking a bath or shower by him or herself***	0.5	0.64#	0.38#	0.68
Able to join in activities (playing out with friends, doing sports, joining things)	0.7	0.48	0.63	0.63	**1.3 Participating in sports activity or exercise**	0.54	0.47	0.46	0.6
					3.1 Getting along with other children	0.48	0.31	0.47	0.38

CHU9D dimensions		**Long term health condition**	**Gender**		**Relevant PedsQL items**		**Long term health condition**	**Gender**	
	No (N=59)	**Yes (N=26)**	**Girl (N=47)**	**Boy (N=37)**		**No (N=59)**	**Yes (N=26)**	**Girl (N=47)**	**Boy (N=37)**
					3.5 Keeping up when playing with other children	0.5	0.49	0.34	0.69

Agreement statistics estimated using AC1.

Agreement coefficient not statistically significant at 95% confidence level

Altman’s scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good

*Sample size: Long term condition: No=19, Yes=3; Gender: Girl=11, Boy=10

CHAPTER 6: THE CHILD'S PERCEPTION OF HEALTH-RELATED QUALITY OF LIFE (HRQOL): A MIXED METHOD APPROACH.

6.1 Overview

This chapter presents the following findings of the mixed method study involving Study sample 1. The study included 39 children who participated in the think aloud interview for their responses to the EQ-5D-Y-3L measure:

1) The self-report validity of responses to the EQ-5D-Y-3L in a community sample of school-aged children, and

2) The impact of response issues identified by application of the Tourangeau response model framework on child self-reported HRQoL and inter-rater agreement between the child and a parent proxy assessor.

6.2 Summary

Objective: This study examines the self-report validity of the EQ-5D-Y-3L in children of varying ages using a retrospective think aloud method.

Methods: A mixed methods study was conducted in a community-based sample of children aged 6-12 years (N=39). In a semi-structured interview, children self-completed the EQ-5D-Y-3L and then engaged in retrospective think aloud. Conversations were audio-recorded and transcribed for analysis in NVivo using the Tourangeau four-stage response model framework to assess comprehension, judgment, recall, and response mapping issues. Fisher's exact test was used to explore statistical differences between child-self reported HRQoL across subgroups. Of the overall sample, as discussed in Chapter 4, the inter-rater agreement for the subsample of children who participated in the think-aloud exercise for the

EQ-5D-Y-3L and their parent dyads, was evaluated using the CCC for overall HRQoL and Gwet's Agreement AC_1 at the dimension level. For detailed methodological information, please refer to Chapter 3, section 3.3 onwards.

Results: Response issues were detected in 46% of the participants (n=18). Comprehension issues, especially in understanding 'discomfort', were noted in the "having pain or discomfort" dimension. Recall-related issues involved responses influenced by usual tendencies (e.g., being *usually* worried) or past incidences (e.g., feeling pain *sometimes*). Judgement-related issues were the most common, particularly in the "doing usual activities" dimension, where responses reflected their *self-perceived ability* to engage in activities rather than health-related limitations. Lifestyle factors, like diet and exercise, were frequent considerations in EQ VAS ratings. Younger children had a higher proportion of response issues (6-7 years: 64%, 8-10 years: 62%), compared to older children (11-12 years: 20%). Children with response issues demonstrated significantly lower EQ-5D-Y-3L scores (mean=0.78, SD=0.04) compared to those without (mean=0.95, SD=0.02). The overall inter-rater agreement was higher for those without any response issues (CCC=0.33) than those with (CCC=0.14). Additionally, higher agreement was noted across all the five dimensions in the subgroup with no response issues relative to those with.

Conclusions: Children in the general community may have different perceptions of HRQoL when responding to the EQ-5D-Y-3L possibly due to their limited experience with health-related challenges. The retrospective think aloud approach adopted highlighted several response issues, particularly in younger children (ages <11 years) indicating the need for careful interpretation of self-reported HRQoL in this group.

6.3 Key Points

- The study used the EQ-5D-Y-3L measure in healthy children aged 6-12 years and found that 46% of the sample struggled to relate the HRQoL dimensions as intended.
- The study found that lower HRQoL scores were often reported due to response issues rather than actual health-related problems.
- Younger age groups, particularly those aged 6-7 and 8-10 years, exhibited more response issues.

6.4 Results

6.4.1 Sample characteristics for the overall qualitative sample

Please refer to Appendix, Section 10.3.1.

6.4.2 Sample characteristics for the EQ-5D-Y-3L sample

Table 6.2 presents the sociodemographic characteristics of the child-parent dyads. A total of 39 children participated, with a median age of 9 (IQR=4 years). Girls were slightly over-represented, accounting for 56% of the participants. Most children reported themselves to be in 'good' to 'excellent' health, with only 8% reporting 'fair' health on the single SRH item. According to parental reports, 28% of the children had one of the following conditions: asthma (45%), autism spectrum disorder (18%), dental caries (18%), and sleep problems (18%). The parents in the sample had a median age of 41 (IQR=9), and approximately one-fifth of the dyads consisted of father-child pairs. When considering SEIFA area-decile numbers, a lower proportion of respondents resided in postcodes associated with relatively disadvantaged quintiles (25%) compared to the Australian population [268].

Table 6-1 Sociodemographic characteristics of the child and parent participants.

Variable	N=39	% sample
Child age		
Mean (standard deviation)	9.1 (2)	
Median (IQR)	9 (4)	

Variable	N=39	% sample
Parent age*		
Mean (standard deviation)	41.1 (5)	
Median (IQR)	40 (8)	
Child gender		
Girl	22	56
Boy	17	44
Parent gender*		
Female	31	82
Male	7	18
Child self-rated general health**		
Excellent	6	16
Very good	16	43
Good	12	33
Fair	3	8
Child long term health condition		
Yes	28	72
No	11	28
Specific health condition		
Autism Spectrum Disorder	2	18
Asthma	5	45
dental caries	2	18
Sleep disorders	2	2
Socio-economic condition according to post-code**		
Relatively advantaged quintile	27	73
(SEIFA decile 7,8,9,10)		
Relatively disadvantaged quintile	10	27
(SEIFA decile 1,2,3,4,5,6)		

*n=38 reported, **n=37 reported. SEIFA= Socio-Economic Indexes for Australia

6.4.3 Response issues by age-group, presence of long-term health condition and gender

A total of 46% (n=18) children experienced at least one or more response issues. The highest proportion of response issues was observed among children in the younger age-groups (6-7 years: 64% and 8-10 years: 62%), whereas the older children (11-12 years) had the lowest proportion of response issues (20%). As illustrated in Table 6.2, the most common response issues were related to judgement (28%) and recall (23%), while comprehension issues were relatively less frequent (18%). Comprehension issues were mainly observed among the youngest age-group, while judgement and recall issues were predominant in children aged 8-10 years.

Overall, there was a potential association between age group and response issues albeit the significance was only marginal (p-value=0.08; Fisher's exact test). However, significant differences were observed between age-group and two specific types of response issues, namely comprehension (p-value=0.02) and judgment (p-value=0.03).

Children with reported health conditions and boys exhibited a marginally higher proportion of response issues compared to their respective comparison groups, as shown in Table 6.3. However, these differences were not statistically significant.

Table 6-2 Description of response issues by age-group, presence of long-term health condition and gender.

Demographic characteristic	**Comprehension n (%)**	**Judgement n (%)**	**Recall n (%)**	**Total (unique) n (%)**
Age group				
6-7 years (N=11)	5 (0.45)	3 (0.27)	3 (0.27)	7 (0.64)

Demographic characteristic	**Comprehension n (%)**	**Judgement n (%)**	**Recall n (%)**	**Total (unique) n (%)**
8-10 years (N=13)	2 (0.15)	7 (0.54)	4 (0.31)	8 (0.62)
11-12 years (N=15)		1 (0.07)	2 (0.13)	3 (0.2)
Presence of long-term health condition				
No (N=28)	4 (0.14)	7 (0.25)	5 (0.18)	12 (0.43)
Yes (N=11)	3 (0.27)	4 (0.36)	4 (0.36)	6 (0.55)
Gender				
Girl (N=22)	2 (0.09)	8 (0.36)	7 (0.32)	10 (0.44)
Boy (N=17)	5 (0.29)	3 (0.18)	2 (0.12)	8 (0.47)
Overall (N=39)	7 (0.18)	11 (0.28)	9 (0.23)	18 (0.46)

6.4.4 Relationship between response issues and 1) self-reported HRQoL scores and 2) inter-rater agreement

Children with response issues had significantly lower (p-value=0.0007) EQ-5D-Y-3L scores (mean=0.78, SD=0.04; median=0.81, IQR=0.1) compared to those with no response issues (mean=0.95, SD=0.02; median=1, IQR=0). Table 6.4 shows the inter-rater agreement for the dimension level and overall EQ-5D-Y-3L scores according to the presence of response issues. Poor child-parent agreement (CCC= 0.14, 95% CI= -0.31, 0.54) was observed among children with response issues, while a higher (fair) agreement (CCC=0.33, 95% CI= -0.06, 0.63) was observed in children with no response issues.

Agreement at the dimension level was lower for all dimensions in children with response issues, except for the “feeling worried, sad or unhappy” dimension. The lowest agreement was observed within the “doing usual activities” and “having pain/discomfort” dimensions. Correspondingly, the highest number of problems reported (Table 6.5) as well as response issues was observed within the “doing usual activities” dimension with judgement and “having pain/discomfort” dimension with comprehension (Fig 6.1).

Table 6-3 Inter-rater agreement for dimension level and overall EQ-5D-Y-3L scores by presence of response issues.

	Walking about	**Looking after myself**	**Doing usual activities**	**Having pain/discomfort**	**Feeling worried, sad or unhappy**	**Overall**
	AC_1 (95% CI)	**AC_1 (95% CI)**	**AC_1 (95% CI)**	**AC_1 (95% CI)**	**AC_1 (95% CI)**	**CCC (95% CI)**
Overall sample (N=38)	0.83 (0.69, 0.97)	0.74 (0.56, 0.91)	0.63 (0.43, 0.84)	0.61 (0.41, 0.82)	0.57 (0.35, 0.78)	0.28 (-0.03, 0.54)
No response issue (N=21)	0.9 (0.75, 1)	0.85 (0.66, 1)	0.85 (0.66, 1)	0.73 (0.49, 0.98)	0.53 (0.23, 0.84)	0.33 (-0.06, 0.63)
Some response issue (N=17)	0.74 (0.46, 1)	0.59 (0.25, 0.92)	0.33 (-0.07, 0.72)	0.45 (0.09, 0.82)	0.61 (0.29, 0.94)	0.14 (-0.31, 0.54)

AC_1= Gwet’s Agreement Coefficient; CCC=Concordance Correlation Coefficient; Altman’s scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good.

Table 6-4 Frequency of problems reported with the child self-report version of the EQ-5D-Y-3L.

Dimensions	**No problems** **n (%)**	**Some problems** **n (%)**	**A lot of problems** **n (%)**

Walking about	33 (0.87)	5 (0.13)	
Looking after myself	33 (0.87)	5 (0.13)	
Doing usual activities	29 (0.76)	9 (0.24)	
Having pain/discomfort	27 (0.71)	10 (0.26)	1 (0.03)
Feeling worried, sad or unhappy	32 (0.84)	6 (0.16)	

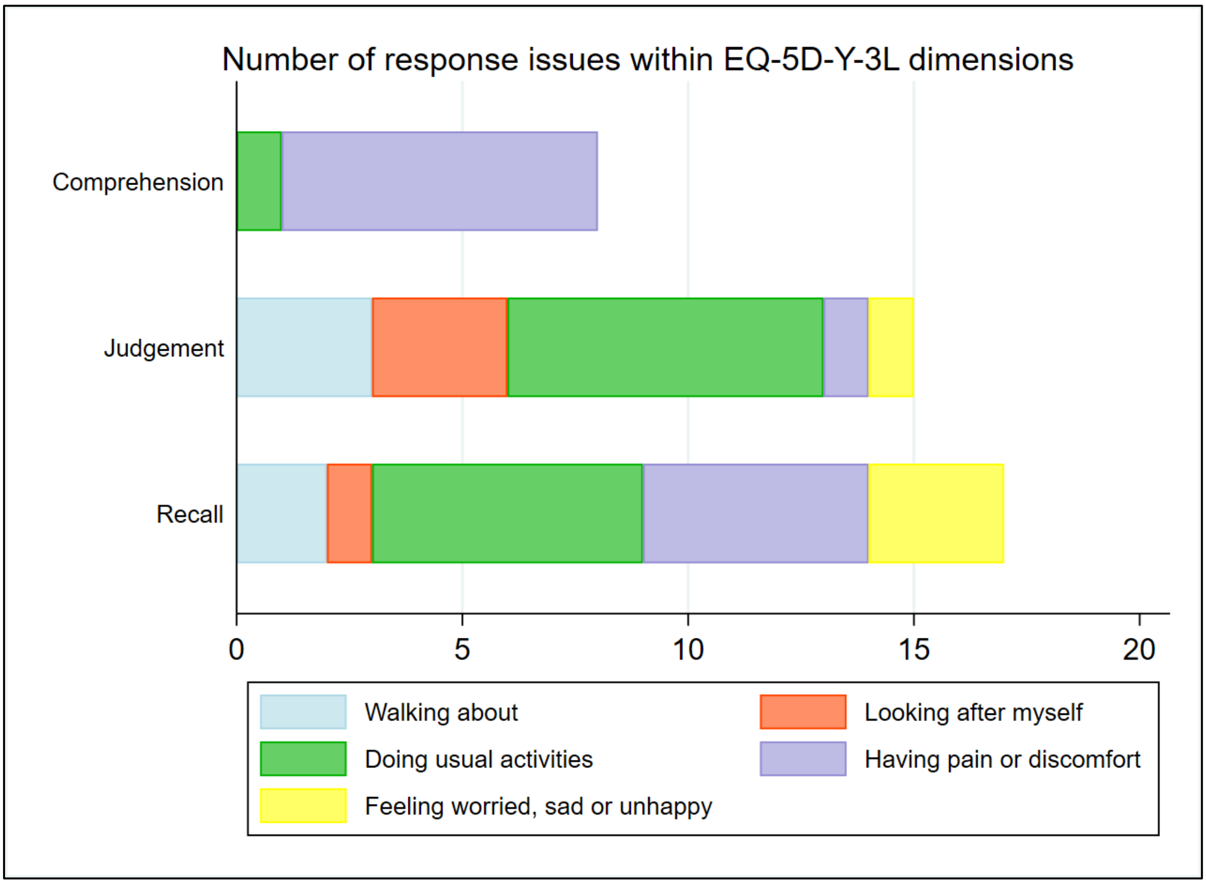


Figure 6.1 Response issues within EQ-5D-Y-3L dimensions.

6.4.5 Response issues by EQ-5D-Y-3L dimensions and VAS- qualitative findings

Table 6.6 highlights the key qualitative findings related to response issues encountered by children in understanding and responding to the EQ-5D-Y-3L and the EQ VAS.

In the “walking about” dimension, four children in total had judgement (n=3) and/or recall (n=2) related response issues. Children conflated health related ability to walk with ability to walk without tripping (non-health related). One child inferred ability to walk to mean walking independently without being accompanied by adults. Whereas another child stated that they have problems with walking but recalled non-active episodes that happens ‘sometimes’.

Similarly, in the “looking after myself” dimension, one child mentioned age-related ability as the reason for having problems with washing or dressing themselves. Two children reported problems due to non-health related issues such as frustration with clothes or dislike of washing themselves. Both responses were inferred based on their overall patterns of behaviour rather than issues specifically encountered on the day of assessment.

In the “doing usual activities” dimension, all judgement related response issues were general non-health related inabilities rather than health related limitations. Similar to other dimensions, recall related problems were identified when children responded based on their general behavioural tendencies or specific events outside of the recall period. Only one comprehension issue was identified with a child that didn’t know what ‘usual activities’ meant.

Comprehension issues were predominant within the “having pain or discomfort” dimension (n=7) followed by recall issues (n=5). Two children did not know what the word ‘discomfort’ meant. Three children inferred ‘pain and discomfort’ to mean ‘emotional pain’. The judgement related response issue for the “having pain or discomfort” dimension was related to the concept of health. One child said they don’t have ‘pain or discomfort’ as they ‘work on their health’. Five children that had recall related issues mentioned episode/s of pain from outside of the recall period.

Only a few response issues were identified within the “feeling worried, sad or unhappy” dimension, mostly related to recall. In one case, a child recalled a specific event of feeling worried (from outside the recall period), while another child recalled their inclination or general tendency to worry.

A common theme that emerged with the rating of the EQ VAS was the association of the concept of health with emotional wellbeing and lifestyle. Five children took into consideration their emotional health when rating themselves on the EQ VAS. Nine children rated themselves based on healthy habits such as eating fruits and vegetables and exercising.

None of the response issues were categorised as response mapping issues.

Table 6-5 Selected quotes from the qualitative analysis of response issues by EQ-5D-Y-3L Dimensions, Participant Age, and Selected Responses.

Quote	Dimension	Age of Participant	Selected Response
"because, sometimes when I walk, it hurts, over here sometimes"	Walking About	8 yrs.	Some problems
"Sometimes I make like mistakes in like – sometimes in sports games I make a mistake about tripping or accidentally hurting by bumping because in – we play basketball sometimes and once we were playing and I accidentally bumped into someone and made them fall and I helped them up though"	Doing Usual Activities	9 yrs.	Some problems
"Because I sometimes when – like it’s really hard to figure stuff out… I meant like when I’m doing maths and stuff"	Doing Usual Activities	8 yrs.	Some problems

Quote	Dimension	Age of Participant	Selected Response
"I don't have pain or discomfort, because I usually always fit in, and no-one forces me to not do that stuff, like be rude to me…"	Having Pain or Discomfort	7 yrs.	No problems
"Discomfort, I thought it means you're not comfortable with all your friends talking behind your back. That's what I thought it meant"	Having Pain or Discomfort	8 yrs.	No problems
"Well, I'm not like – no-one's really happy all the time. But I wouldn't say that I'm sad all the time, I'm just at a normal, maybe some things might have made me upset, like small petty things"	Feeling Worried, Sad, or Unhappy	11 yrs.	Some problems
"Yeah. I'm not like the happiest, but I'm not upset – just like pretty good, I'm feeling"	VAS	11 yrs.	Rating: 71
"Like I could be healthier with my choices of eating and stuff, and I could do – eat more healthier things and do more healthier things. But I'm not really unhealthy. I don't eat heaps of chocolate and not much vegetables and fruit. I still do quite a few – I do basketball for a few hours a week and stuff. I still do exercising and stuff"	VAS	11 yrs.	Rating: 75

6.5 Discussion

Despite its ubiquitous use, research exploring how children understand and interpret the items in the EQ-5D-Y is limited [21,61-66]. This study aimed to address this gap in the literature by investigating the response process in children aged 6-12 years who self-completed the EQ-5D-Y-3L, using the Tourangeau four-stage response model. The findings indicated that the response process varied depending on the age group, with children ≤ 10 years of age demonstrating a higher proportion of comprehension, judgement, and recall related issues as compared to older children.

Wille et al. conducted a multinational study to adapt the EQ-5D-Y from the adult EQ-5D-3L [144]. They found that the items were well accepted and generally comprehensible in a predominantly healthy sample of children and adolescents aged 8-18 years. Nevertheless, the major challenge identified by the study was children’s difficulty distinguishing between health-related impairment and age-related inability [36]. The findings by Amien and colleagues also indicated younger children (ages 5-7 years) reported a significantly higher frequency of problems with “looking after myself” as compared to older children (ages 8-10 years), primarily due to their need for assistance, which was unrelated to their medical condition [236]. In this study, only two children, ages 8 and 11 years, respectively, reported problems with “walking about” and “looking after myself” due to age-related independence/limitation. However, in a similar vein, a predominant issue identified in this study was judgement related particularly for the “doing usual activities” dimension. Children responded thinking about their non-health related ability to perform the task, rather than their health-related limitations. More specifically, they reported problems if they considered their general limitations in relation to athletic or academic performance in comparison to other children. This is consistent with the findings of Cremeens et al., who reported that children frequently used social comparisons to judge their own ability to perform certain tasks [140].

Children also took into account their emotional well-being when responding to the “having pain and discomfort” dimension and their overall health using the EQ VAS. Children tended to associate the word ‘discomfort’ with emotional discomfort. Amein et al. reported that children may not be familiar with the word ‘discomfort’; however, in their study, this lack of familiarity did not affect their understanding of the question being asked [236]. In a previous study, it was found that school-age children (aged 5-11 years) associated the term “healthy” with behaviours such as eating fruits and vegetables [293]. In this study, children also considered a “healthy” lifestyle, which included diet and exercise, when evaluating their overall health using the EQ VAS. As noted by Brazier and McCabe, EQ VAS tends to capture different aspects of health, including concepts like fitness [294]. They also considered their level of happiness, an indicator of quality of life [295], when evaluating their overall health.

The EQ-5D-Y requires the child to report their health status ‘today’. In their review, Arbuckle et. al state that children under the age of 6 years may have limited introspective abilities and struggle with distinguishing between past, present, and future [296]. In this sample, while many children did understand that they had to report their health status ‘today’, recall issues were identified mainly within the “having pain/discomfort” dimension. Specifically, it was observed that some children based their responses on past experiences of pain or discomfort, even if they were not experiencing any pain or discomfort the day of the interview. Future investigations comparing child specific HRQoL measures with different timeframes (e.g., EQ-5D-Y and the PedsQL™) taking into account children’s developmental capabilities and potential recall biases may be needed to provide deeper insights into children’s responses to HRQoL assessments.

Most importantly, this study found that the presence of response issues was associated with lower self-reported HRQoL scores. Whilst it was found that child self-reported HRQoL did not vary by the presence or absence of long-term health-condition/s, it was found to vary by response issues; children with identified response issues reported a lower HRQoL overall compared to the sub-group with no response issues (mean difference= 0.17, standard error=0.04). It is possible that this lower reported HRQoL may result from incorrect interpretation and response to the HRQoL dimension/s themselves, rather than actual health-related limitations. Similarly, dimension level child-parent agreement was lower in the sub-group with identified response issues (relative to the children with no identified response issues) for all dimensions, except the “feeling worried, sad or unhappy” dimension. The higher level of agreement within the “feeling worried, sad or unhappy” dimension may be attributed to the reason for the response issues with this dimension. Meaning, when reporting on “feeling worried, sad or unhappy”, children reported on their general tendency to worry which likely would have also been the same reasoning guiding the parent’s proxy response resulting in a higher level of agreement.

6.6 Limitations

This study has limitations that warrant consideration. While associations between understanding the HRQoL measure and lower reporting of HRQoL was observed, the cross-sectional design limits the ability to draw causal inferences. The lower levels of agreement among participants with response issues remain challenging to interpret, partly due to the nature of the retrospective think-aloud methodology. There may be a possibility that children were unable to fully articulate their initial reasoning when completing the questionnaire. Additionally, loss of data due to technical errors reduced the sample size, potentially affecting the representativeness and robustness of the findings. Finally, while the study’s sample size was adequate for overall analyses, it may have been insufficient to support subgroup comparisons, such as differences in agreement based on age, which limits the exploration of nuanced patterns. Future research could incorporate think-aloud protocols from both children and proxies to explore whether they make similar errors when responding to specific questions or items, particularly those involving recall. Addressing such issues in future studies would significantly enhance the reliability and generalisability of findings in child-proxy HRQoL assessments.

Prior studies have investigated the reliability and validity of self-reports in children as young as 5 years of age using only quantitative analysis [146, 270, 297, 298]. However, this study provides valuable insights into the reliability of self-reported EQ-5D-Y-3L responses and inter-rater agreement in Australian children using a mixed methods study design.

6.7 Conclusions

This study aimed to examine the validity of children's responses when self-reporting their own HRQoL using the EQ-5D-Y-3L and its impact on the HRQoL scores and inter-rater agreement. The findings highlighted that younger children (particularly children aged 6-7 years and 8-10 years) may face several challenges in their understanding and comprehension of the HRQoL dimensions relative to older children (aged 11-12 years). These challenges may result in younger children self-reporting a lower level of health-related quality of life and inconsistencies with parental proxy reported HRQoL, which are not solely based on the actual quantification of health-related impairments intended to be captured by the measure. These findings underscore the importance of enhancing child-specific measures and ensuring that children fully understand the questions posed to them.

CHAPTER 7: UNDERSTANDING OF THE CHU9D DIMENSIONS IN CHILDREN: A RETROSPECTIVE THINK ALOUD STUDY.

7.1 Overview

This chapter presents the following findings of the mixed method study involving Study sample 1. The study included 36 children who participated in the think aloud interview for their responses to the CHU9D measure:

1) The self-report validity of responses to the CHU9D in a community sample of school-aged children, and

2) The impact of response issues identified by application of the Tourangeau response model framework on child self-reported HRQoL and inter-rater agreement between the child and a parent proxy assessor.

7.2 Summary

Objective: This study investigated the understanding and the interpretation of the CHU9D dimensions by children through a retrospective think aloud method and its impact on the inter-rater agreement.

Method: A sample of community-based children aged 6-12 years (N=36) and their parents independently completed the self and the proxy versions of the CHU9D, respectively. After the completion of the measure, children were asked to verbalise their thoughts, which were recorded and transcribed. Data were analysed in NVivo, with two raters applying the Tourangeau four-stage response model framework to identify response-related issues in comprehension, judgment, recall, and response mapping. Fisher's exact test was used to explore statistical differences between child-self reported HRQoL across subgroups. The

overall inter-rater agreement was assessed with CCC and dimension level using Gwet's AC_1 (please refer to Chapter 3, section 3.3 onwards for specific details).

Results: Response issues were detected in n=18 (49%) children. Children most frequently encountered problems with the "activities" and "daily routine" dimensions. A common issue was the confusion between the responses "joining in *any* activities" versus "joining in *most* activities", with children interpreting these choices as a constraint on the number of activities they could participate in or the availability of opportunities/activities. In addition, children reported problems within these dimensions due to preference (e.g., reluctance to participate in "activities" or "daily routine") rather than considering limitations due to health issues. Additionally, there was a lack of clarity about what constituted "activities" among some children. A notable observation was that sometimes children referenced past incidents rather than their current or 'today' HRQoL status. No significant subgroup differences were identified by age, gender, or health status, highlighting a sample-wide issue in the understanding and interpretation of the CHU9D dimensions mentioned above. Children with response issues demonstrated significantly lower CHU9D utilities (mean= 0.78, SD= 0.03) than those without (mean= 0.88, SD= 0.04). Overall inter-rater agreement was poor regardless of the presence/absence of response issues. Dimension level inter-rater agreement varied, with generally higher levels observed in the subgroup of children without any response issues, with the exception of the "tired" dimension, which exhibited poor agreement (AC_1= 0.23). Interestingly, in this dimension, children with response issues showed better agreement (AC_1=0.62, good). The strongest agreement was observed in the "sad" dimension for the subgroup without response issues (AC_1=0.95, very good).

Conclusions: While the CHU9D is a valuable tool for assessing child HRQoL, this study highlights the importance of ensuring that children understand the dimensions and response levels. Enhancing the clarity of the CHU9D may help mitigate interpretational difficulties and improve the reliability of HRQoL measurements. The findings underscore the need for further adaptations to the measure to better suit the cognitive and linguistic capabilities of children.

7.3 Key Points

- The study administered the CHU9D measure in healthy children aged 6-12 years and found response issues in 49% of the sample.
- Response issues were primarily detected in the understanding and interpretation of the “activities” and “daily routine” dimensions.
- Similar to the EQ-5D-Y-3L, children with response issues on the CHU9D reported significantly lower HRQoL scores, indicating that these lower scores may be due to challenges in understanding and interpretation rather than actual health-related problems.

7.4 Results

7.4.1 Sample characteristics for the overall qualitative sample

Presented in section 6.3.1.

7.4.2 Sample characteristics for the CHU9D sample

The sample characteristics were similar to those of the EQ-5D-Y-3L sample. As presented in Table 7.1, 36 children participated in the think aloud interviews for their CHU9D responses, with a median age of 9 years (IQR = 4 years). The parents in the sample had a median age of 40 years (IQR = 9 years). Among the children, 42% (n = 15) were boys and 58% (n = 21) were girls. Of the children, 67% (n = 24) did not have a long-term condition, while 33% (n=12) did. Specific health conditions reported included autism spectrum disorder (17%, n = 2), congenital heart disorder (8%, n=1), asthma (50%, n=6), dental caries (17%, n=2), and sleep problems (8%, n=1). The children’s SRH item indicated that 31% (n=11) considered their health to be excellent, 61% (n=22) rated it as very good, and 8% (n=3) rated it as good. The parents' gender distribution was 22% male (n = 8) and 78% female (n=28). A larger proportion of the child-parent dyads (64%, n=23) lived in relatively advantaged areas, while 36% (n=13) resided in relatively disadvantaged areas.

Table 7-1 Sociodemographic characteristics of the child and parent participants.

Variable	N	%
Child age		
Mean (standard deviation)	36	9 (2)
Median (IQR)		9 (4)
Parent age		
Mean (standard deviation)	36	41 (6)
Median (IQR)		40 (9)
Child gender		
Boy	15	42
Girl	21	58
Long term condition		
No	24	67
Yes	12	33
Child Health condition		
Autism spectrum disorder	2	17
Congenital heart disorder	1	8
Asthma	6	50
Dental caries	2	17
Sleep problems	1	8
Child self-rated general health		
Excellent	11	31
Very good	22	61
Good	3	8

Variable	**N**	**%**
Parent gender	3	4
Male	8	22
Female	28	78
Socio-economic condition according to post-code		
Relatively advantaged quintile	23	64
(SEIFA decile 7,8,9,10)		
Relatively disadvantaged quintile	13	36
(SEIFA decile 1,2,3,4,5,6)		

SEIFA= Socio-Economic Indexes for Australia

7.4.3 Response issues by age-group, presence of long-term health condition and gender

Table 7.2 shows the frequency and the type of response issues by subgroups, namely, age-group, presence of long-term health condition and gender and for the overall sample. Of the 36 children, similar to the EQ-5D-Y-3L sample, half (n=18) were identified to have response issues. Only one comprehension issue was noted, involving a 6–7-year-old girl child with no health condition. Judgement issues were frequent (n=17), identified in all but one of the participants exhibiting response issues, while recall issues were observed in seven participants. The proportion of response issues were marginally higher among children aged 8–10 years, those with health condition, and boys in this sample.

There were no significant differences in the presence of response issues by age-group (Fisher's exact test, p-value=0.912), gender (p-value=0.999) or health condition (p-value=0.725).

Table 7-2 Description of response issues by age-group, presence of long-term health condition and gender.

Demographic characteristic	Comprehension n (%)	Judgement n (%)	Recall n (%)	Total (unique) n (%)
Age group				
6-7 years (N=10)	1 (0.1)	4 (0.4)	4 (0.4)	5 (0.5)
8-10 years (N=14)		8 (0.57)	3 (0.21)	8 (0.57)
11-12 years (N=12)		5 (0.42)	1 (0.08)	5 (0.42)
Presence of long-term health condition				
No (N=21)	1 (0.05)	9 (0.43)	4 (0.19)	10 (0.48)
Yes (N=15)		8 (0.53)	4 (0.27)	8 (0.53)
Gender				
Girl (N=24)	1 (0.04)	10 (0.42)	5 (0.21)	11 (0.46)
Boy (N=12)		7 (0.58)	3 (0.25)	7 (0.58)
Overall (N=36)	1 (0.03)	17 (0.47)	8 (0.22)	18 (0.5)

7.4.4 Relationship between response issues and 1) self-reported HRQoL scores and 2) inter-rater agreement

The mean score for those with response issues (mean=0.78, SD=0.03; median=0.78, IQR=0.2) was significantly lower than that for those without (mean= 0.88, SD= 0.04; median=0.90, IQR=0.09), as indicated by the Wilcoxon rank-sum test (p-value= 0.0019).

The overall inter-rater agreement was poor across the sample (CCC=0.07). For the sub-sample with response issues, the agreement was slightly higher (CCC=0.03) compared to those without (CCC=0.01), indicating virtually no agreement within the sample (Table 7.3).

The inter-rater agreement across the dimensions ranged from moderate to very good in the sample as shown in Table 7.3. The classification of the sample into *no issues* and *with issues* based on the overall presence or absence of any response issue did not consistently correspond to specific dimension-level issues and their corresponding agreement. For instance, lower agreement (Gwet's AC_1) was observed within the dimensions "sad" and "pain", for those with issues ("sad" =0.77, "pain" =0.75) than those without issues ("sad" =0.94, "pain" =0.88). Notably, no response issues were identified within these dimensions specifically as shown in Fig 7.1. Similarly, a notable difference between the two groups was observed in the "tired" dimension. Children with no issues demonstrated poor agreement (0.22), while those with issues (in other dimensions) showed good agreement (0.62). Since no response issues were noted within the "tired" dimension, this suggests that the disagreement may not stem from a difference in understanding of the dimension but rather a genuine difference in perspective as the greatest number of problems (vs no problems) were reported within this dimension (Table 7.4). However, this may also indicate the role of chance, or some other variable not captured in this study.

In the "activities" dimension, very good agreement was observed among children without issues (0.83), while only a moderate agreement was evident among those with issues (0.57). Correspondingly, the highest number of response issues were identified in this dimension. The second highest number of response issues was noted in the "daily routine" dimension. Agreement in this dimension was only marginally higher for children without response issues (0.58) than those with (0.43). Similarly, good agreement was observed in the "sleep" dimension among those with no issues (0.64) while those with issues showed only a moderate agreement (0.43). Similar levels of agreement were observed in the remaining dimensions.

Table 7-3 Inter-rater agreement for dimension level and overall CHU9D scores (and their 95% confidence intervals) by presence of response issues.

Overall/Dimension	Overall sample	No issues	With issues
Overall HRQoL*	**0.07 (-0.26, 0.38)**	**0.01 (-0.42, 0.44)**	**0.03 (-0.42, 0.47)**
Worried	0.52 (0.32, 0.73)	0.57 (0.27, 0.87)	0.48 (0.18, 0.79)
Sad	0.85 (0.73, 0.98)	0.94 (0.82, 1)	0.77 (0.53, 1)
Pain	0.82 (0.68, 0.96)	0.88 (0.71, 1)	0.75 (0.51, 1)
Tired	0.42 (0.21, 0.63)	0.22 (-0.08, 0.52)	0.62 (0.33, 0.91)
Annoyed	0.54 (0.34, 0.74)	0.58 (0.29, 0.87)	0.51 (0.2, 0.81)
Schoolwork	0.51 (0.31, 0.71)	0.45 (0.14, 0.75)	0.57 (0.27, 0.86)
Sleep	0.53 (0.34, 0.73)	0.64 (0.36, 0.91)	0.43 (0.13, 0.73)
Daily routine	0.5 (0.3, 0.71)	0.58 (0.29, 0.87)	0.43 (0.12, 0.74)
Able to join in activities	0.7 (0.53, 0.87)	0.83 (0.62, 1)	0.57 (0.28, 0.86)

*Overall HRQoL agreement was estimated using Concordance Correlation Coefficient (CCC), while the dimension level agreement was estimated using Gwet's Agreement Coefficient (AC_1); Altman's scale interpretation: Less than or equal to 0.2=Poor, between 0.21 & 0.4=Fair, between 0.41 & 0.6=Moderate, between 0.61 & 0.8=Good, between 0.81 & 1=Very Good.

Table 7-4 Frequency of problems reported with the child self-report version of the CHU9D.

	Response level				
Dimensions	**No problems**	**Little problems**	**A bit of problems**	**Quite a lot of problems**	**A lot of problems**

	n (%)	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Worried	25 (69)	8 (22)	1 (3)	1 (3)	1 (22)
Sad	34 (94)	2 (6)			
Pain	27 (75)	6 (17)	2 (6)	1 (3)	
Tired	4 (11)	25 (69)	5 (14)	1 (3)	1 (3)
Annoyed	33 (92)	1 (3)	1 (3)		1 (3)
Schoolwork	27 (75)	7 (19)	1 (3)		1 (3)
Sleep	24 (67)	12 (33)			
Daily routine	28 (78)	7 (19)	1 (3)		
Able to join in activities	27 (75)	7 (19)	2 (6)		

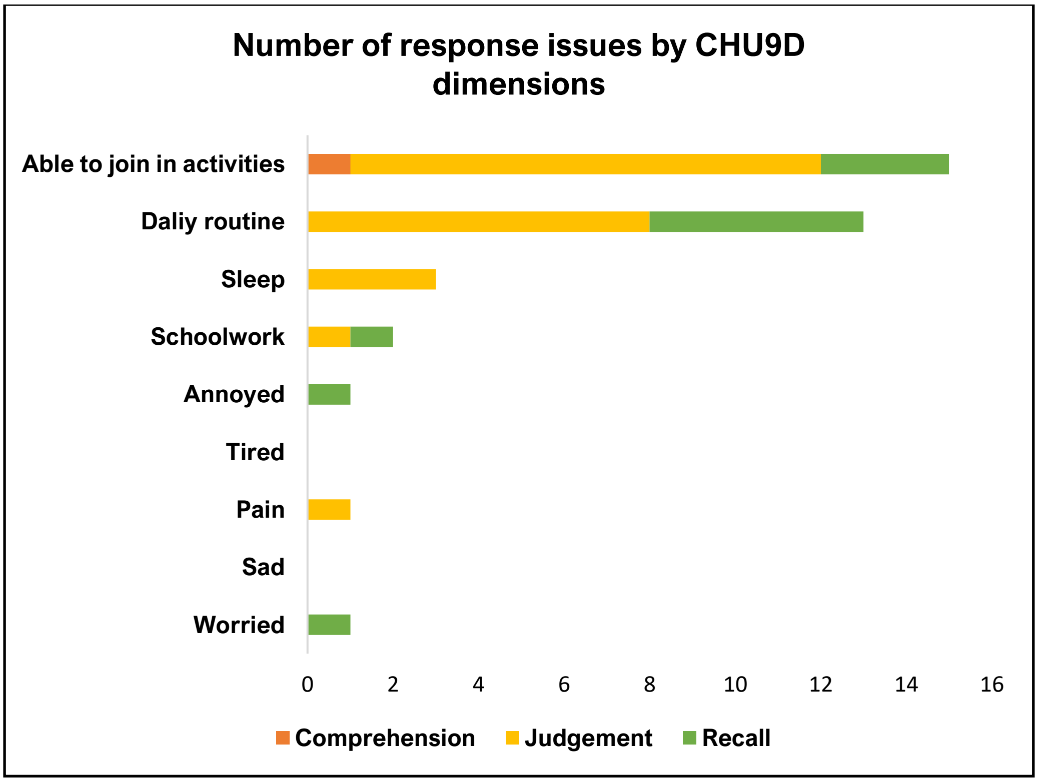


Figure 7.1 Response issues within CHU9D dimensions.

7.4.5 Response issues by CHU9D dimensions - qualitative findings

Table 7.5 presents the key qualitative findings related to response issues encountered by children in understanding and responding to the CHU9D.

As mentioned above, one of dimensions with the most response issues was "activities" (n=15). The only comprehension issue noted was due to the child misunderstanding the word "activities". Within the activity dimension, judgement issues were noted predominantly when some children confused their ability to perform activities with the availability of opportunities to engage in them (n=4). For example, one child indicated they could join in 'most' activities (rather than 'any') but clarified that this was because they were doing activities with family members and, therefore, could not be involved in other activities. Another child associated their inability to join activities with the school holidays, implying they had more opportunities to engage in activities during school days. Additionally, there was evidence of confusion between ability and capability or preference. Four children mentioned they could join in more activities but only if it was a sport they knew how to play or if they were not tired.

Another dimension for which a high number of response issues was found was “daily routine” (n=12). Most of them were judgement issues (n=7). When asked about their daily routine, children confused health-related ability to perform daily activities with self-perceived capability (n=5), availability of opportunity (n=1), preference (n=1) or answered based on a specific incident (n=1). A common theme that emerged when asked about their daily routine was that children mentioned having "a few problems" because they were engaged in other activities (e.g., reading a book, being distracted) instead of following their typical “routine” due to external influences (from parents or siblings) or personal preferences. In addition to these identified response issues, overlapping recall issues were also noted with five children within this dimension.

Within the “sleep” dimension, judgement issues (n=3) were observed as children provided answers based on a specific incident, for instance, one child linked sleep problems to a late-night activity with their family. In the “schoolwork” dimension, one child reported problems with schoolwork due to the day of the interview being a school holiday. In the “pain” dimension, another child misjudged the severity or nature of their pain and attributed their pain to mosquito bites, misinterpreting the response scale. This led to an indication of a higher level of discomfort than warranted by such irritations, which may not align with the intent of the question.

Table 7-5 Selected quotes from the qualitative analysis of response issues by CHU9D Dimensions, Participant Age, and Selected Responses.

Quote	Dimension	Age of Participant	Selected Response
" Yeah. A few problems. Just because I like – I get distracted because my little sister's always watching her favourite TV show called [Cocomelon]…"	Daily routine	7 yrs.	Few problems
"Well, usually I just stay up late because that's me, I'm a night owl. So is my dad. So, I just stay up late. Then me and my brother because we were playing, I was used to being awake for that long. So, a few problems but I went to sleep, I got sleep. I'm good"	Sleep	11 yrs.	Few problems
"Because we're going to [relatives' house] and we're here so that's three activities. So – but we probably can't fit anymore now"	Activities	8 yrs.	Join in with some activities
"The last bit of the day I can. But the – for the morning I can't because I need to do my room"	Activities	6 yrs.	Join in with most activities
"Well, maybe some activities, because maybe I don't think too positive about one of the activities and I just think "Oh I just can't do this" and I won't join into it. But this year, because I'm in Year 6, I'm trying to join in as most activities as I can...maybe some games that I like, maybe like sport, like netball. I've never been really interested in some sports, and I just don't want to join in with it or something"	Activities	11 yrs.	Join in with most activities

Quote	Dimension	Age of Participant	Selected Response
"Because sometimes I don't want to do my daily routine because sometimes I don't want to have a shower or like stuff like that"	Daily routine	10 yrs.	Few problems
" My daily routine. Well, usually I – well, Mum begs me to get ready for school, or et cetera, et cetera. Or when we're going out in the morning, and I don't listen. I just keep reading my book or watching my – or watching the news. Drives my mum nuts"	Daily routine	12 yrs.	Few problems
" Because… activities [we] open we make stuff or we colour in stuff"	Activities	6 yrs.	Join in with any activities

*[ ] quotes edited for clarity

7.5 Discussion

This study found that the response issues identified led to significant discrepancies in the reported median HRQoL utilities between those with and without these issues. The variation in reported HRQoL scores was higher for those with response issues than those without. Given that the children were generally in good health and there were no differences based on age, gender, or health status, this variation likely arises from misunderstandings of the HRQoL dimensions rather than genuine health-related limitations. This implies that the challenges children face in understanding HRQoL dimensions may lead to biased HRQoL values. Consequently, in CUAs, these values may not accurately represent the QALY gains, potentially affecting health economic evaluations and decision-making processes in this population.

As mentioned, the frequency of response issues by subgroup (age group, long-term health condition, gender) did not show significant differences. This may indicate that younger children (under eight years of age), children with health conditions, or children of a certain gender were not particularly prone to response issues. This finding suggests that children aged 6-7 years may be able to meaningfully respond to HRQoL measures such as the CHU9D. This may stem from the CHU9D's development, which involved qualitative interviews with children in its 'bottom-up' approach, unlike the EQ-5D-Y which was designed using a more traditional 'top-down' approach [299]. The CHU9D's development process incorporated feedback from children to ensure that the measure was understandable and relevant to their experiences [90].

It is notable that overall agreement using CCC was poor regardless of the presence or absence of response issues. This may be a statistical artifact due to the low power of the sample and this has been addressed previously (see section 4.5, paragraph 3). Notably, a poor overall agreement (agreement estimate < 0.2) has been reported in other studies with large sample sizes as well. For instance, the study by Rogers et. al included 486 child-proxy dyads [226], conducted in cohort of children with dental caries (a common condition). However, as identified in the systematic review in Chapter 2, studies assessing inter-rater agreement with the CHU9D are scarce [222]. This highlights the need for assessment of inter-rater agreement in diverse samples in future studies to draw more robust conclusions.

Except for the dimensions “sleep” and “tired”, the inter-rater agreement was consistently lower for the group with any response issues. Interestingly, no response issues were noted in the “tired” dimension, and similarly, no issues were observed within the “sad” dimension. This may indicate that the disagreement in these dimensions noted in this study sample were due to genuine differences in perspective. Both are psychosocial dimensions; however, the agreement was the highest with “sad” and lowest with “tired” across the nine CHU9D dimensions. This may suggest that parents in this sample were better at concordantly reporting emotional distress using the CHU9D than physical fatigue.

This study revealed that half of the sample exhibited response issues, primarily in the judgment stage, as classified using the Tourangeau framework. Issues in the judgment stage arise when respondents struggle to retrieve relevant information pertinent to the question, leading to answers that do not accurately reflect the intended measurement. Children frequently retrieved irrelevant, non-health-related information to respond to certain dimensions of the CHU9D measure. While this does not imply that their responses were erroneous, it does indicate that their answers did not align with the intention of the questions. For instance, many children interpreted “routine” in the context of their typical behaviours or habits, or what is ideally expected of them, rather than their ability to perform necessary daily tasks like getting dressed. This interpretation suggests that children were focusing on personal preferences or activities they chose to engage in, rather than addressing the intended assessment of their health-related capability to perform essential daily functions.

Another example is the dimension “able to join in activities.” Some children selected response options based on the availability of activities they could join, rather than their health-related ability to join those activities. For instance, a child may report being able to join in “some” vs “any” activities simply because there were not many activities available and not because their health limited their participation in any of the activities. These misinterpretations suggest that children were influenced by their personal context and the opportunities available to them, rather than focusing on their health-related limitations.

Such response issues reveal that while children can provide valuable insights into their HRQoL, their interpretations can diverge significantly from the intended constructs of the measures. These examples underscore the challenge in designing HRQoL measures that are both comprehensible and relevant to children. The complexity of interpreting terms like "routine" or "joining in activities" highlights the need to ensure that the language and examples used in these measures are appropriate for children to elicit accurate responses, necessitating a careful reconsideration of how questions are phrased, and the contexts provided within the child-specific HRQoL measures.

A study by Guerriero et al., investigated whether using animation on a touch screen device was a more effective way self-reporting HRQoL in children aged 4-14 years compared to traditional paper questionnaires. In a hospital setting with 438 children, five versions of the CHU9D were administered. They found that children in all age groups favoured animation on a tablet, highlighting its potential for self-assessment of HRQoL in children as young as 4 years [257]. Future research should focus on refining HRQoL measures to minimise such potential misunderstandings. For instance, enhancing the existing measures to incorporate visual illustrations [300], animations, or interactive components to improve children's comprehension and engagement, may result in more precise and meaningful data in the assessment of child HRQoL. Providing context and examples may also help to obtain more accurate and representative responses from children.

7.6 Limitations

Please refer to section 6.6 of Chapter 6.

7.7 Reflections

The process of conducting think-aloud interviews with 6-7-year-olds was both challenging and insightful and provided valuable reflections for future research in this area. Overall, the approach was reasonably successful, as many children in this age group were able to articulate their thought processes while responding to the questionnaire, offering insights into their interpretation of the HRQoL questions and response selection. Success was determined based on the children's ability to engage with the tasks, articulate their reasoning, and provide consistent responses with prompts. However, variability in cognitive and verbal abilities was evident, which influenced the depth and clarity of the data collected.

For researchers considering similar think-aloud studies with younger children, several practical insights emerged. First, the use of age-appropriate language and familiar tools, such as laptops or tablets instead of traditional pen-and-paper methods, was crucial in maintaining the children's engagement and ensuring their comprehension. Second, shorter interview durations helped manage attention spans. Third, establishing a comfortable and supportive environment allowed children to express themselves more freely. Nonetheless, challenges such as limited vocabulary, abstract reasoning abilities, and occasional reliance on parental input were observed.

Despite these challenges, this study highlights the feasibility of involving children aged 6-7 years in think-aloud research, though it requires careful methodological considerations. For example, additional probes and scaffolding techniques [301] can enhance understanding without leading the child's responses. While this thesis concludes that self-reported HRQoL is generally reliable from age 11 onwards, the think-aloud findings contribute to the broader discussion about younger children's ability to self-report. In conclusion, with appropriate adaptations, researchers can gather meaningful data from this age group, though findings should be interpreted with caution given the developmental constraints.

By reflecting on these experiences, this work contributes a practical framework for researchers aiming to conduct think-aloud studies with younger children and encourages a reconsideration of the scepticism surrounding self-reports in children under 8 years of age. Further research is needed to refine these methods and explore their applicability in diverse contexts.

7.8 Conclusions

This study highlights several challenges in obtaining meaningful HRQoL measurements from children. Misunderstandings of terms, context-specific answers, and confusion between different aspects of their abilities and preferences, rather than health-related limitations, contributed to a lower reported HRQoL compared to children without these issues. These findings underscore the importance of enhancing child-specific measures and ensuring that children fully understand the questions posed to them. This will enhance the validity of child-self reported HRQoL measures, leading to more accurate assessments in both clinical and research settings.

Overall, it is recommended that current child HRQoL measures, such as the EQ-5D-Y-3L and CHU9D, be used with caution for children aged 10 years and younger, as this group demonstrated greater difficulty in meaningfully self-reporting their HRQoL. In contrast, children aged 11 years and older exhibited fewer response issues and higher inter-rater agreement with parental proxy reports, indicating a stronger ability to understand and accurately reflect their HRQoL using these measures. Notably, some younger children aged 6–7 years demonstrated the potential to reliably self-report their HRQoL using the CHU9D. As such, rather than imposing a strict cut-off age of 11 years for self-reporting, it is important to consider providing additional support or adaptations to the measures for younger children to enhance the validity of their self-reported HRQoL data.

CHAPTER 8: CHALLENGES IN EVALUATING HEALTH-RELATED QUALITY OF LIFE (HRQOL) IN CHILDREN: SELECTING THE VALUE-SET FOR THE EQ-5D-Y-3L.

8.1 Overview

This chapter presents the methods and findings for the second empirical study using a secondary dataset, i.e., Study 2, which analysed the impact of using different value sets on the inter-rater agreement when utilising the EQ-5D-Y-3L. It explored the effect of applying international child-specific preference weights versus Australian adult weights on the consistency of child-proxy HRQoL reports.

8.2 Summary

Objective: A child-specific value set for the EQ-5D-Y is more appropriate for evaluating the health-related quality of life (HRQoL) in children. Given that the Australian child-specific value-set for the EQ-5D-Y-3L is still under development, this study explores the impact of applying EQ-5D-Y-3L ('Y') value sets (child perspective) from nine other countries and the Australian EQ-5D-3L value set (adult perspective) on the assessment of the inter-rater gap between child-self and proxy reports of HRQoL.

Methods: An online cohort of 845 dyads, comprising children aged 6-10 years and their parents from across Australia, participated. This study analysed child and proxy-reported HRQoL data using the EQ-5D-Y-3L self and proxy (EQ-5D-Y-3L Proxy Version 1) reports respectively. HRQoL values were derived using the EQ-5D-3L Australian (adult) value set and published EQ-5D-Y-3L value sets from nine other countries (collectively referred to as 'Y' value sets). Analyses were stratified by age (6-7 and 8-10 years), gender, and health condition presence (Yes/No). Group differences were identified using paired t-tests. The Intraclass Correlation Coefficient (ICC) assessed concordance between self and proxy HRQoL values, while Gwet's Agreement Coefficient (AC_1) was used for dimension-level agreement.

Results: The highest HRQoL values were reported using the Indonesian value set (child=0.95, proxy=0.96), while Slovenia reported the lowest (child=0.87, proxy=0.89), following the Australian adult value set (child=0.85, proxy=0.86). Compared to child-self reports, proxy ratings were consistently higher with statistically significant difference across

all ‘Y’ value sets (mean difference range= 0.007 to 0.016; p-value<0.005) but not when the Australian adult value set was applied. Girls (mean difference range=0.008 to 0.022; p-value<0.005) as well as children without health conditions (mean difference range=0.010 to 0.025; p-value<0.005) reported significantly lower HRQoL than proxy estimates across all value sets. Proxies also reported significantly higher HRQoL for older children (8-10 years) across ‘Y’ weights (mean difference range=0.008, 0.020; p-value<0.005), but not with the Australian adult weights. Remarkably, children with health conditions reported higher HRQoL than their proxies, with significant differences noted only with the Australian adult set. Inter-rater agreement was mostly consistent across all value sets (adult and ‘Y’) (ICC range=0.62 to 0.71). The dimension level agreement was very good (0.82-0.96) for all dimensions except “feeling worried, sad or unhappy” which showed a lower agreement (0.69, good).

Conclusions: This research highlights that the choice of value set can critically affect the assessment of child HRQoL, with child-specific weights offering more uniform outcomes across different contexts than adult-based weights. Nonetheless, agreement coefficients for HRQoL values derived from country-specific child value sets and the Australian adult weights were comparable. Therefore, it can be recommended to use any available ‘Y’ value set for assessing child HRQoL until the Australian-specific version is unavailable, rather than relying solely on the adult EQ-5D-3L value set. These findings underscore the importance of carefully selecting value sets for evaluating child HRQoL and suggest further research into developing and applying child-specific preference weights.

8.3 Key points

- There is a lack of guidance on whether to use international child-specific value sets or the Australian adult value set for EQ-5D-Y-3L child HRQoL assessments.

- Child HRQoL values using international child-specific value sets were more consistent than those from the Australian adult value set, with proxies overestimating HRQoL across all child-specific value sets.
- The main findings illustrate the need for country-specific child value sets in the evaluation of health outcomes in child populations.

8.4 Introduction

Child-specific preference-based measures are designed to capture the HRQoL of children from a perspective that incorporates the society's valuation of different health states [2, 75]. Among the measures designed for this purpose, the EQ-5D-Y-3L measure is widely recognised for its reliability and validity in assessing child HRQoL [77, 131]. The EQ-5D-Y-3L was developed by the EuroQol group and represents an adaptation of the original adult version of the EQ-5D-3L with modifications to the wording to better suit its use in child populations[7] [77]. Contrary to the well-established methods for valuing adult HRQoL, there are several concerns regarding fundamental aspects of the methods for valuing child HRQoL including whether adult weights from the EQ-5D can be used to obtain child utilities [302] (see Chapter 1, section 1.4).

Employing the adult-weights accompanying the adult HRQoL measure to derive child utilities facilitates the comparison of utility values across adults and children, however, this approach is met with significant criticism regarding the suitability of adult-weights to accurately reflect the experiences associated with child health states [70, 303, 304]. Adults' valuation of health states often diverges when considering their own conditions (own perspective) versus those of children (child perspective). Notably, adults may attribute lower values to their own health

---

[7] However, the findings in Chapter 6 suggest that further adaptations may be necessary to make the measure more appropriate for self-reporting in children aged 10 years and younger.

state compared to similar health states in children [69, 88, 305]. This discrepancy stems from the adult perspective, which has been explored in several notable studies, including those by Reckers-Droog et al. [80], Powell et al. [306], and Lipman et al. [81, 307]. Adults may perceive the impact of certain conditions on children as less severe, or there may be a reluctance to make life-year trade-offs for children, who are often viewed as having inherently more valuable years ahead of them [69]. Thus, using inappropriate value sets to calculate utilities in child populations can lead to significant misestimations of QALY gains and consequently result in potentially misguided healthcare decisions.

The EuroQol group recently published a protocol in an effort to standardise the valuation methods used to derive the value sets for the EQ-5D-Y-3L across different countries [89]. The health states derived from the EQ-5D-Y-3L are valued from the perspective of the society (or taxpayers), thus circumventing the ethical concern associated with involving children in direct valuation tasks that would require them to compare different health states with ‘being dead’. However, the preferences are elicited from adults valuing the hypothetical health states of a child, i.e., *child perspective*. The protocol for obtaining value sets recommends a two-step technique: an initially DCE complemented by a c-TTO task (see Chapter 1, section 1.6.1 for more details) [89].

To date, value sets for the EQ-5D-Y-3L have been published and are available for nine countries. Value sets specific to Australia are currently under development but are not yet available [78, 308]. The absence of country-specific child value sets currently presents challenges for clinicians and researchers in selecting the most appropriate existing value set for evaluating child HRQoL using the EQ-5D-Y-3L in Australia. Consequently, there may be a reliance on adult measures and their accompanying utility values, as seen in the health technology assessments/appraisals submitted to regulatory bodies such as the PBAC in Australia [309] and the NICE in the UK [310]. Additionally, although not recommended due to the reasons mentioned previously, some studies employing the child-specific EQ-5D-Y-3L measure have applied the respective country-specific adult value set for the EQ-5D as an alternative to calculate the child utilities [146, 148, 311-314].

Presently, there is little evidence available to guide the decision about whether to apply child-specific weights from a different country or whether to apply adult weights from the same country, thereby maintaining the cultural representation of the sample. The extent to which cultural differences versus the perspective used to value health states (i.e., adult versus child perspectives) influence HRQoL assessments is yet to be explored. To address this, the study will examine the impact of employing different cross-cultural value sets on the consistency of child-proxy reports of child preference-weighted HRQoL using the EQ-5D-Y-3L within an Australian setting.

8.5 Methods

8.5.1 Participants

This study is part of a focused secondary analysis on Sample 2 derived from the larger Paediatric Multi-Instrument Comparison (P-MIC) study [315]. The larger sample size of the P-MIC study was necessary to evaluate the impact of various value sets effectively. The P-MIC study was designed to assess the comparative effectiveness of paediatric HRQoL measures within the Australian context. This subset was specifically utilised due to its inclusion of child-proxy dyads for reporting child HRQoL, allowing for a detailed examination of inter-rater agreement and health state utilities. Detailed documentation on the recruitment strategies and data collection methodologies employed by the P-MIC study is available elsewhere [316]. Ethical approval for the overarching P-MIC study was obtained from the Royal Children’s Hospital Human Research Ethics Committee on March 21, 2021 (HREC/71872/RCHM-2021) [315].

The sample utilised here is a subset of Sample 2 (N=1,774), recruited from across the Australian States and Territories through the online panel Pureprofile Pty Ltd Australia. The recruitment period spanned from March 2022 to September 2022. This sample included children aged 6-10 years and their parents/caregivers from the general population, with quotas in place to ensure even representation across each age group (6, 7, 8, 9, 10). Each participant dyad was invited to self-complete the survey through an online link provided via REDCap. Written informed consent on behalf of the dyad was obtained from each parent or caregiver participant. Participants received a token of appreciation valued between AUD$3 and AUD$4 for their participation [315].

Child HRQoL data collected using child and proxy reports of the EQ-5D-Y-3L and the single SRH item in the initial survey is analysed here. The proxy perspective reported in this study is proxy-proxy (EQ-5D-Y-3L Proxy Version 1), i.e., from the proxy’s own perspective. Additionally, caregivers were asked to provide their own HRQoL data using the EQ-5D-5L and the SRH item as well as the demographic information for the dyad. This included details such as the ages and gender of both the child and the parent/caregiver, the presence of any medical condition in the child, family income, caregiver’s level of education, their state or territory of residence and the caregiver’s relationship to the child.

8.5.2 EQ-5D measures

8.5.2.1 EQ-5D-3L

The EQ-5D-3L measure is widely used for assessing HRQoL within adult populations. This measure is used for self-reporting purposes in individuals aged 16 and older. It comprises a descriptive system with five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The 3L version categorises responses into three levels, ranging from “no problems“ through “some/moderate problems” to “extreme problems/unable to”. The 5L version expands this scale by introducing two additional response categories, “slight problems” and “severe problems”, thereby offering a nuanced gradation beyond the 3L response categories [317, 318] .

8.5.2.2 EQ-5D-Y-3L

In response to the need for a version tailored to younger populations (ages 8-16), the EQ-5D was adapted into the child-friendly EQ-5D-Y, available in both 3L and 5L formats. This adaptation rephrases the dimensions to resonate more closely with children’s experiences, specifying mobility as “walking about”, self-care as “looking after myself”, and usual activities contextualised in terms of child-centric activities such as engaging in school and play. The response levels for the EQ-5D-Y-3L version are simplified to “no problems”, “some problems”, and “a lot of problems” with slightly different labels of “not”, “a bit”, and “very” for the dimension assessing anxiety/depression specified as “feeling worried, sad, or unhappy” [2, 75].

8.5.2.3 Valuation of the EQ-5D and the EQ-5D-Y-3L

The Australian value set available for the EQ-5D-3L was derived using the TTO method only but includes a simulation approach [243]. In this study, this value set was applied to the each of the 243 possible health states generated by the EQ-5D-Y-3L. The proxies reported their own HRQoL using the EQ-5D-5L measure, which was valued using the corresponding Australian value set derived through the DCE method with a duration approach [319].

The International Valuation Protocol for the EQ-5D-Y-3L was developed in response to the need for appropriate value sets for the estimation of QALYs and to address methodological inconsistencies and limitations identified in previous studies that published value sets for this measure [89]. The protocol addresses three primary concerns. Firstly, it advocates for adopting the taxpayer’s perspective to elicit preferences from the adult population. Secondly, it recommends that valuation tasks should be conducted from the viewpoint of a child, specifically targeting the perspective of a 10-year-old. Lastly, it proposes a two-step methodology for calculating value sets: an initial online DCE complemented by face-to-face interviews involving a c-TTO task [89].

Since the publication of the valuation protocol, value sets for nine countries are available: Japan, Slovenia, Spain, Germany, Belgium, China (Mainland), Hungary, Indonesia and Netherlands [78]. The methodological component of all the published value sets aligned closely with the recommended protocol. DCEs were conducted with 1,000 adults (with some studies also including an adolescent sample aged 11-17 years [320]) to assess the importance of various health dimensions. Each participant was presented with one of 10 blocks, each containing 15 questions. These questions compared two health states using the EQ-5D-Y-3L system, asking respondents to choose the preferred state for a hypothetical 10-year-old child. At this stage, the DCE values were unanchored and reflected the relative significance of the different levels of the dimensions. To establish the final value set by anchoring the DCE values onto a QALY scale ranging from 0 (representing death) to 1 (indicating perfect health), face-to-face interviews with ~200 adult participants engaging in c-TTO tasks were conducted. The protocol recommended a minimum of 10 health states to be included in the c-TTO tasks [89]. However, some studies differed in the number of health states included in the c-TTO tasks, ranging from 10-28 health states [321-324]. Additionally, contrary to the recommended method in the protocol, some studies also conducted the interviews for the c-TTO tasks online [244, 322, 325, 326], perhaps influenced by the heightened reliance on virtual interviews following the COVID-19 pandemic [78].

The protocol currently lacks specific guidance on the optimal method for anchoring latent scale DCE values. To date, all but one study approached the estimation of the value set by either mapping the DCE data onto the c-TTO scale or by rescaling it based on the worst health state (33333). In contrast, the value set for China employed a hybrid model that utilised both DCE and c-TTO data jointly [321]. Variations in methodology were also evident in the mapping functions used to link the two data, such as the Indonesian value set, which incorporated a power term for non-linear mapping [324]. This diversity in approaches highlights the evolving nature of methodologies in the field, as researchers seek to refine and adapt strategies that best captures the complexities of valuation of the preferences for child health states.

8.5.3 Statistical analysis

Sociodemographic data were analysed using descriptive statistics. The subgroups were analysed by age groups (6-7 years and 8-10 years), gender (girl, boy, non-binary), and the presence of a health condition (Yes/No). The categorisation into two distinct age groups, 6-7 years, and 8-10 years, was chosen to facilitate a comparison of responses between younger children under eight years of age, the age above which self-completion of the EQ-5D-Y-3L questionnaire is recommended, and their older counterparts.

The child and proxy reported HRQoL in this sample was converted into preference-weighted HRQoL (HRQoL values) using the EQ-5D-3L Australian (adult) value set [243] and the published EQ-5D-Y-3L value sets from nine other countries (collectively referred to 'Y' weights/ 'Y' value sets) [244, 321-328], resulting in ten paired HRQoL value reports. The HRQoL values were described using mean (and SD) to compare self and proxy-reported assessments. Median (and IQR) values were also reported for completeness. To identify group differences, paired t-tests were administered. The concordance between the self and proxy HRQoL values was assessed using the ICC [157], while Gwet's AC_1 was utilised to evaluate HRQoL agreement at the dimension level [151, 161, 264]. Given the predominance of healthy children in the sample and the EQ-5D-Y-3L's three-level response design, the unweighted version of AC_1 was deemed appropriate. This was based on the reduced likelihood of marked disagreements within the sample, rendering the weighted Gwet's AC_2 less appropriate due to its potential to overstate agreement for minor discrepancies between categories [314]. The ICC and Gwet's AC_1 values span a range from -1 to 1, and their levels of significance were interpreted following Altman's scale, which classifies agreement as poor (≤0.2), fair (≤0.4), moderate (≤0.6), good (≤0.8), and very good (1) (please see Chapter 2, section 2.4.5 for more details) [161]. The statistical procedures were executed in Stata 16.1 (Stata Corp LLC, College Station, TX, USA) [265] and the significance level was set at 0.05.

8.6 Results:

8.6.1 Participant characteristics

Table 8.1 provides an overview of the demographic characteristics of children and their proxies included in the study. Child-parent reports were collected for 845 dyads. The mean age of children in the overall sample was 7.9 years (SD= 1.5 years) and 36.2 years (SD=6.6 years) for the parents. Both boys and girls were equally represented in the child sub-sample. Just under a third of the children in the sample (29%) were reported to have some medical condition. Of the whole sample, autism spectrum disorder was the most prevalent at 18% followed by attention deficit hyperactivity disorder at 5%. Less prevalent conditions included hearing impairment, anxiety disorder, intellectual disability, and learning disability, each at 1%, with other disorders like 16p11.2 deletion syndrome, asthma, Duchenne muscular dystrophy, dyslexic, and Stage 4 Radial Aplasia each represented by less than 1% of the sample. In general, children self-reported their health mostly as very good (49%) or excellent (29%) on the SRH item.

The subset of proxies in the study was primarily female, constituting 87% of the sample. A significant majority, 97%, were parents of the child participants. Other proxies included carers (1.3%), grandparents (0.95%), siblings (0.24%), and other relatives (0.36%). Proxies' health ratings for children on the SRH item were more evenly distributed, with good (30%), very good (39%), and excellent (26%) being the most common responses. Proxies reported a mean EQ-5D-5L utility value for their own health at 0.78 (0.19). Educational attainment among proxies varied; 36% had attained at least a bachelor’s degree, while 40% held a Certificate or Diploma. In terms of weekly household income, the most common income brackets were $1,000-$1,999 and $2,000 or more, each accounting for 40% and 36% of the sample, respectively. Most participants resided in Queensland (28%), New South Wales (26%), and Victoria (24%). Participants from Western Australia (10%) and South Australia (8%) also contributed to the sample, whereas Tasmania, the Australian Capital Territory (ACT), and the Northern Territory (NT) were less represented, comprising 2%, 1%, and 0.4% of the sample, respectively.

Table 8-1 Socio-demographic characteristics of child-proxy dyad.

Demographic characteristic	**Variable**	**Child N (%)**	**Proxy N (%)**
Age (years)	Mean (SD)	7.9 (1.47)	36.21 (6.56)
	Median (IQR)	8 (2)	36 (9)
Gender	Male	419 (0.50)	104 (0.12)
	Female	424 (0.50)	736 (0.87)
	Transgender Female	-	2 (0.002)
	Not Described	1 (0.001)	2 (0.002)
	Prefer Not to Answer	1 (0.001)	1 (0.001)
Child Medical Conditions Child	No	600 (0.71)	-
	Yes	245 (0.29)	-
Child general health item	Poor	4 (0.005)	9 (0.01)

Demographic characteristic	Variable	Child N (%)	Proxy N (%)
	Fair	27 (0.03)	39 (0.05)
	Good	157 (0.19)	250 (0.30)
	Very Good	416 (0.49)	326 (0.39)
	Excellent	241 (0.29)	221 (0.26)
Relation to the child	Parent	-	821 (0.97)
	Grandparent	-	8 (0.01)
	Carer unrelated to the child	-	11 (0.01)
	Sibling	-	2 (0.002)
	Other relative	-	3 (0.003)
EQ-5D-5L utility (Proxy)	Mean (SD)	-	0.78 (0.19)
	Median (IQR)	-	0.78 (0.25)
Proxy Highest Level of Education	Bachelor's degree or above	-	307 (0.36)
	Certificate/Diploma	-	335 (0.40)
	Year 12	-	102 (0.12)
	Year 9-11	-	78 (0.09)
	Certificate I/II	-	20 (0.02)
	Year 8 or below	-	3 (0.004)

Demographic characteristic	Variable	Child N (%)	Proxy N (%)
Income (per week)	Less than $500	-	35 (0.04)
	$500-$999	-	164 (0.20)
	$1,000-$1,999	-	333 (0.40)
	$2,000 or more	-	301 (0.36)
Residence State	Australian Capital Territory		9 (0.01)
	New South Wales		221 (0.26)
	Victoria		203 (0.24)
	Queensland		235 (0.28)
	South Australia		67 (0.08)
	Western Australia		87 (0.1)
	Tasmania		19 (0.02)
	Northern Territory		3 (0.004)

Note: Percentages may not total 100% due to rounding.

8.6.2 Differences in child-proxy reported preference-weighted HRQoL

Table 8.2 presents a comparative analysis of child and proxy reported preference-weighted HRQoL derived from all the available value sets for the EQ-5D-Y-3L in addition to the EQ-5D-3L value set for Australian adults. The dyadic differences are examined for the overall sample and by gender, age-group, and health condition.

Since the sample consisted of children from the general population, the HRQoL values reported were generally high across all value sets, with the highest mean HRQoL values observed using the Indonesian value set (child=0.95, proxy=0.96) and Australian adult the lowest (child=0.85, proxy=0.86). Proxy ratings were consistently higher than child ratings. This difference was statistically significant for HRQoL values across all ‘Y’ value sets (p-value<0.05). In contrast, the child-self and proxy HRQoL values, when reported using the Australian adult value set, were not significantly different.

For both genders, the overall findings were similar across all value sets. Proxy ratings were higher than child ratings except with boys using Australian (adult) weights. However, the differences were not statistically significant for boys. In contrast, girl child and proxy ratings were significantly different across all value sets. Similarly, the HRQoL values derived from all ‘Y’ weights were significantly higher from the proxy perspective for older children aged 8-10 years. However, this difference was not observed for values calculated using Australian (adult) weight. Children without any reported health condition reported significantly lower HRQoL values across all value sets. Children with a reported medical condition were the only subgroup where the child-self reported HRQoL values exceeded those reported by proxies across all value sets. Notably, only the Australian (adult) value set showed a significant difference between child and proxy ratings within this subgroup.

Table 8-2 Comparative analysis of child and proxy reported HRQoL mean (standard deviation) values using the Australian Adult EQ-5D-3L and the EQ-5D-Y-3L value sets.

	Australia Adult	Belgium	Slovenia	Germany	Hungary	Japan	Netherlands	Spain	Indonesia	China
Overall										
Child	0.85 (0.006)	0.88 (0.006)	0.87 (0.006)	0.89 (0.005)	0.9 (0.005)	0.92 (0.003)	0.91 (0.005)	0.87 (0.006)	0.95 (0.003)	0.92 (0.004)
Proxy	0.86 (0.006)	0.9 (0.005)	0.89 (0.005)	0.9 (0.005)	0.91 (0.004)	0.93 (0.002)	0.92 (0.004)	0.89 (0.005)	0.96 (0.003)	0.93 (0.003)
Difference	-0.006	-0.015	-0.016	-0.011	-0.014	-0.008	-0.010	-0.016	-0.007	-0.010
p-value	0.20	<0.001	<0.001	0.01	<0.001	<0.001	<0.001	<0.001	0.004	0.003
Subgroups:										
Gender										
Boy										
Child	0.85 (0.009)	0.88 (0.008)	0.87 (0.009)	0.89 (0.007)	0.9 (0.007)	0.92 (0.004)	0.91 (0.006)	0.87 (0.009)	0.95 (0.005)	0.92 (0.006)
Proxy	0.85 (0.008)	0.89 (0.007)	0.88 (0.008)	0.89 (0.007)	0.91 (0.006)	0.93 (0.003)	0.91 (0.006)	0.88 (0.008)	0.96 (0.004)	0.92 (0.005)
Difference	0.003	-0.009	-0.010	-0.005	-0.009	-0.005	-0.006	-0.011	-0.005	-0.007

	Australia Adult	Belgium	Slovenia	Germany	Hungary	Japan	Netherlands	Spain	Indonesia	China
p-value	0.68	0.15	0.16	0.37	0.15	0.10	0.26	0.12	0.14	0.20
Girl										
Child	0.86 (0.009)	0.88 (0.008)	0.87 (0.009)	0.89 (0.008)	0.9 (0.007)	0.92 (0.004)	0.91 (0.006)	0.87 (0.009)	0.95 (0.005)	0.92 (0.006)
Proxy	0.87 (0.008)	0.9 (0.007)	0.9 (0.007)	0.9 (0.006)	0.92 (0.006)	0.93 (0.003)	0.92 (0.005)	0.9 (0.007)	0.96 (0.004)	0.93 (0.005)
Difference	-0.015	-0.020	-0.022	-0.017	-0.019	-0.010	-0.015	-0.022	-0.008	-0.013
p-value	0.02	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.006	0.001
Age-group										
6-7 years										
Child	0.86 (0.009)	0.89 (0.009)	0.88 (0.01)	0.89 (0.008)	0.91 (0.008)	0.92 (0.004)	0.91 (0.007)	0.88 (0.009)	0.95 (0.005)	0.92 (0.006)
Proxy	0.86 (0.009)	0.9 (0.008)	0.89 (0.008)	0.9 (0.007)	0.92 (0.007)	0.93 (0.003)	0.92 (0.006)	0.89 (0.008)	0.95 (0.005)	0.93 (0.005)
Difference	-0.004	-0.012	-0.012	-0.009	-0.011	-0.006	-0.008	-0.012	-0.005	-0.009
p-value	0.59	0.08	0.1	0.13	0.08	0.06	0.11	0.10	0.19	0.08

	Australia Adult	Belgium	Slovenia	Germany	Hungary	Japan	Netherlands	Spain	Indonesia	China
8-10 years										
Child	0.85 (0.008)	0.88 (0.008)	0.87 (0.008)	0.88 (0.007)	0.9 (0.007)	0.92 (0.004)	0.9 (0.006)	0.87 (0.008)	0.95 (0.005)	0.91 (0.006)
Proxy	0.86 (0.007)	0.9 (0.006)	0.89 (0.007)	0.9 (0.006)	0.91 (0.006)	0.93 (0.003)	0.92 (0.005)	0.89 (0.007)	0.96 (0.003)	0.92 (0.005)
Difference	-0.008	-0.018	-0.019	-0.013	-0.016	-0.009	-0.012	-0.020	-0.008	-0.011
p-value	0.21	<0.001	<0.001	0.02	<0.001	<0.001	0.01	<0.001	0.007	0.01
Health condition										
No										
Child	0.88 (0.006)	0.91 (0.006)	0.9 (0.006)	0.91 (0.005)	0.92 (0.005)	0.93 (0.003)	0.93 (0.004)	0.9 (0.006)	0.97 (0.003)	0.93 (0.004)
Proxy	0.9 (0.005)	0.93 (0.004)	0.93 (0.004)	0.93 (0.004)	0.95 (0.003)	0.94 (0.002)	0.95 (0.003)	0.93 (0.004)	0.98 (0.002)	0.95 (0.003)
Difference	-0.018	-0.023	-0.025	-0.019	-0.021	-0.011	-0.017	-0.025	-0.010	-0.016
p-value	0.02	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Yes										

	Australia Adult	Belgium	Slovenia	Germany	Hungary	Japan	Netherlands	Spain	Indonesia	China
Child	0.79 (0.013)	0.82 (0.013)	0.80 (0.015)	0.83 (0.012)	0.85 (0.012)	0.89 (0.006)	0.86 (0.01)	0.81 (0.014)	0.91 (0.009)	0.88 (0.010)
Proxy	0.76 (0.013)	0.81 (0.012)	0.80 (0.013)	0.82 (0.011)	0.84 (0.011)	0.89 (0.005)	0.86 (0.009)	0.80 (0.012)	0.91 (0.007)	0.87 (0.009)
Difference	0.022	0.004	0.005	0.007	0.003	0.001	0.005	0.004	0.001	0.004
p-value	0.22	0.62	0.64	0.39	0.69	0.78	0.50	0.66	0.89	0.56

Significance level<0.05, represented as grey highlights.

8.6.3 Inter-rater agreement for the preference-weighted HRQoL (Australian (adult) value set) and the dimension level HRQoL

Table 8.3 presents the dyadic agreement [ICC, (95% CI)] for the HRQoL values calculated using the Australian (adult) value set and the dimension level HRQoL. According to Altman's interpretation, the child-proxy agreement for HRQoL values was good (0.67) for the overall sample. While the level of agreement was similar across the subgroups categorised by age and gender, a significant difference in agreement was observed between children with and without any reported health condition. Specifically, a lower agreement (0.55) was reported between proxies and relatively healthy children as compared to children with health conditions (0.72).

The dimension level agreement was very good (0.81-1) for all dimensions except "feeling worried, sad or unhappy" which showed a lower but good agreement (0.69). This observation was generally consistent across all subgroups based on age, gender, and health condition. Notably, children a with health condition had a substantially lower agreement (0.58) within the "feeling worried, sad or unhappy" dimension as compared to children without any health condition (0.74). Additionally, they demonstrated a significantly lower agreement within the "looking after myself" (0.8) and the "doing usual activities" dimensions (0.73) in comparison to their healthy counterparts.

Table 8-3 Agreement between child self and proxy reports using the EQ-5D-Y-3L for the overall sample and by subgroups.

		Overall HRQoL	**Walking about**	**Looking after myself**	**Doing usual activities**	**Having pain/ discomfort**	**Feeling worried, sad or unhappy**
	N (%)	ICC (95% CI)	AC_1 (95% CI)	AC_1 (95% CI)	AC_1 (95% CI)	AC_1 (95% CI)	AC_1 (95% CI)
Overall	845	0.67 (0.63, 0.71)	0.96 (0.94, 0.97)	0.89 (0.87, 0.91)	0.86 (0.84, 0.89)	0.82 (0.79, 0.85)	0.69 (0.65, 0.73)
Subgroup:							
Gender							
Boy	419 (0.5)	0.63 (0.57, 0.69)	0.95 (0.93, 0.97)	0.87 (0.83, 0.9)	0.83 (0.79, 0.87)	0.82 (0.77, 0.86)	0.67 (0.61, 0.73)
Girl	424 (0.5)	0.71 (0.66, 0.76)	0.96 (0.94, 0.98)	0.91 (0.88, 0.94)	0.9 (0.87, 0.93)	0.83 (0.79, 0.87)	0.71 (0.66, 0.77)
Age-group							
6-7 yrs.	374 (0.44)	0.65 (0.59, 0.71)	0.96 (0.94, 0.98)	0.84 (0.8, 0.89)	0.83 (0.79, 0.88)	0.85 (0.81, 0.89)	0.7 (0.64, 0.76)
8-10 yrs.	471 (0.56)	0.69 (0.64, 0.73)	0.95 (0.93, 0.97)	0.93 (0.9, 0.95)	0.88 (0.85, 0.92)	0.8 (0.76, 0.85)	0.68 (0.63, 0.74)

		Overall HRQoL	**Walking about**	**Looking after myself**	**Doing usual activities**	**Having pain/ discomfort**	**Feeling worried, sad or unhappy**
Health condition							
No	592 (0.7)	0.55 (0.49, 0.61)	0.96 (0.95, 0.98)	0.92 (0.9, 0.95)	0.91 (0.89, 0.94)	0.83 (0.79, 0.86)	0.74 (0.7, 0.78)
Yes	253 (0.3)	0.72 (0.66, 0.78)	0.94 (0.91, 0.97)	0.8 (0.74, 0.86)	0.73 (0.66, 0.8)	0.82 (0.76, 0.87)	0.58 (0.5, 0.66)

Altman's scale interpretation: Less than or equal to 0.2=Poor (red), between 0.21 & 0.4=Fair (orange), between 0.41 & 0.6=Moderate (yellow), between 0.61 & 0.8=Good (green), between 0.81 & 1=Very Good (blue).

8.5.4 Inter-rater agreement for the preference-weighted HRQoL derived using ‘Y’ value sets:

Table 8.4 provides a detailed comparison of child-proxy inter-rater agreement using the EQ-5D-Y value sets for different countries. Agreement was described as ICC (95% CI) for the overall sample and the subgroups. The ICCs for overall inter-rater agreement was consistent across countries, ranging from 0.62 to 0.71, categorised as good according to Altman’s scale. The Indonesian value set reported the highest overall agreement (0.71), while the Chinese reported the lowest (0.62).

The agreement was observed to be good across the majority of value sets and subgroups with the exceptions of the Chinese value set and the subgroup comprising children with no health condition. Girl child subgroups demonstrated higher agreement than boys, with ICCs ranging from 0.70 to 0.76 for girls and from 0.55 to 0.68 for boys. For the age group 6-7 years, good agreement was observed with values ranging from 0.65 to 0.69, except with the Chinese value set where the agreement was only moderate (0.55). Children aged 8-10 years had slightly higher agreement than the younger age group, with ICCs from 0.65 to 0.70. Notably, children without health conditions showed a moderate level of agreement, the lowest across all subgroups, with ICCs ranging from 0.42 to 0.56. In contrast, children with health conditions exhibited good agreement, with ICCs between 0.70 and 0.83.

Table 8-4 Comparison of inter-rater agreement for HRQoL values using EQ-5D-Y value sets for different countries.

	Belgium ICC (95% CI)	**Slovenia ICC (95% CI)**	**Germany ICC (95% CI)**	**Hungary ICC (95% CI)**	**Japan ICC (95% CI)**	**Netherlands ICC (95% CI)**	**Spain ICC (95% CI)**	**Indonesia ICC (95% CI)**	**China ICC (95% CI)**
Overall	0.69 (0.66, 0.73)	0.69 (0.65, 0.72)	0.69 (0.65, 0.72)	0.67 (0.63, 0.71)	0.67 (0.64, 0.71)	0.68 (0.64, 0.71)	0.69 (0.66, 0.73)	0.71 (0.68, 0.75)	0.62 (0.58, 0.66)
Subgroups:									
Gender									
Boy	0.65 (0.6, 0.71)	0.65 (0.59, 0.7)	0.66 (0.6, 0.71)	0.62 (0.56, 0.68)	0.64 (0.58, 0.69)	0.64 (0.58, 0.69)	0.66 (0.6, 0.71)	0.68 (0.62, 0.72)	0.55 (0.48, 0.62)
Girl	0.73 (0.69, 0.77)	0.73 (0.69, 0.78)	0.72 (0.67, 0.76)	0.72 (0.67, 0.76)	0.71 (0.66, 0.76)	0.72 (0.67, 0.76)	0.73 (0.68, 0.77)	0.76 (0.71, 0.79)	0.70 (0.65, 0.74)
Age-group									
6-7 years	0.68 (0.62, 0.73)	0.68 (0.62, 0.73)	0.68 (0.62, 0.73)	0.66 (0.6, 0.71)	0.67 (0.61, 0.72)	0.67 (0.61, 0.72)	0.68 (0.62, 0.73)	0.74 (0.69, 0.78)	0.58 (0.51, 0.65)
8-10 years	0.7 (0.65, 0.75)	0.7 (0.65, 0.75)	0.69 (0.64, 0.74)	0.68 (0.63, 0.72)	0.68 (0.63, 0.73)	0.69 (0.64, 0.73)	0.7 (0.65, 0.75)	0.69 (0.64, 0.74)	0.65 (0.60, 0.70)

	Belgium ICC (95% CI)	Slovenia ICC (95% CI)	Germany ICC (95% CI)	Hungary ICC (95% CI)	Japan ICC (95% CI)	Netherlands ICC (95% CI)	Spain ICC (95% CI)	Indonesia ICC (95% CI)	China ICC (95% CI)
Health condition									
No	0.54 (0.48, 0.59)	0.54 (0.48, 0.59)	0.56 (0.5, 0.62)	0.5 (0.44, 0.56)	0.52 (0.46, 0.58)	0.54 (0.48, 0.59)	0.54 (0.48, 0.59)	0.42 (0.35, 0.48)	0.43 (0.37, 0.50)
Yes	0.77 (0.71, 0.81)	0.76 (0.7, 0.81)	0.74 (0.68, 0.79)	0.75 (0.69, 0.8)	0.76 (0.7, 0.81)	0.74 (0.68, 0.79)	0.77 (0.71, 0.81)	0.83 (0.78, 0.86)	0.70 (0.63, 0.76)

Altman's scale interpretation: Less than or equal to 0.2=Poor (red), between 0.21 & 0.4=Fair (yellow), between 0.41 & 0.6=Moderate (orange), between 0.61 & 0.8=Good (green), between 0.81 & 1=Very Good (blue).

8.7 Discussion

This study was the first to explore the estimation of children’s HRQoL (aged 6-10 years) through proxy reports compared to self-reports within an Australian dyad sample from the general population, employing both the ‘Y’ value sets —comprising all the published EQ-5D-Y-3L value sets— and the EQ-5D-3L Australian adult weights. This study examined how the choice of value set (adult vs child weights) impacts the evaluation of discrepancies and the level of agreement in the HRQoL values as reported by children and proxies. Notably, in contrast to the HRQoL values calculated using the ‘Y’ value sets, the differences in child and proxy reported HRQoL values obtained with adult weights were not significantly different. Across all the value sets, the inter-rater agreement was consistently observed to be good.

In general, the application of country specific adult-weights, in this study, resulted in lower HRQoL values relative to the ‘Y’ value sets. This is unsurprising, as noted by Devlin et al. [245], the EQ-5D-Y-3L child values exceed those of adults derived using the EQ-5D-5L for similar health states. Consequently, assuming the same length of survival, the higher HRQoL values associated with ‘Y’ value sets may render the improvement in QALYs from interventions in the child populations seem relatively less significant than when adult weights are applied, potentially decreasing the cost-effectiveness of the intervention [245, 303]. In addition, as mentioned in Chapter 3 (section 3.3.1), the Australian values for the EQ-5D-Y-3L were assumed to resemble European rather than Asian preference patterns [245]. This study found that when comparing the Australian adult value set with the ‘Y’ value sets from the nine countries, the value sets from Spain, Slovenia, Belgium, and Germany produced HRQoL values that were closer to those generated using the Australian adult value set, i.e., at the lower end of the range of values.

A previous study published used the Australia-specific adult value set for the EQ-5D-3L to derive utilities for child HRQoL assessments and examine child-parent agreement in a South

Australian sample [314]. The study demonstrated a low inter-rater agreement between child-parent dyads for the preference-weighted HRQoL [314]. This outcome was consistent with the results derived from applying the CHU9D with its associated child-weights in the same sample [269]. This study found that the choice of value sets was important in highlighting the systematic differences in child and proxy assessments of HRQoL, with statistically significant discrepancies consistently observed across all 'Y' value sets. However, such differences were not evident when employing the Australian adult weights, where the child-proxy differences in reported HRQoL values did not reach statistical significance. Nevertheless, in relation to inter-rater agreement, consistent agreement was noted for the overall sample and across subgroups and across all value sets (adult and 'Y' value sets) with the exception of the Chinese value set for the subgroups comprising boys and children aged 6-7. Moreover, the findings did not indicate any closer alignment of the agreement levels estimated using the Australian value set to either the European (Western) or the Asian (Eastern) value sets.

To date, systematic reviews by Kwon et al. [104], Khadka et al. [115] and Jiang et al. [116], investigating the divergences in self and proxy reports of HRQoL values have reported the direction of this divergence to be inconsistent, varying according to the measure used, the health condition or the type of proxy. Nonetheless, a notable observation was that parental proxies tended to report significantly higher HRQoL relative to child self-report in the general population sample [104, 115, 116]. In this study, the findings indicated that proxies tended to overestimate the HRQoL of children. This finding was also observed for the subgroup of children without health conditions where the proxies significantly overestimated their child's HRQoL across all value sets. Moreover, the lowest agreement across all value sets was also noted among children with no health condition/s. This indicates that there may be a systematic difference in understanding and interpretation of the HRQoL dimensions which results in the overreporting of problems by children themselves relative to their caregiver. The mixed methods study presented in Chapter 6, involving the Study 1 sample, provides insights into the challenges that arise when generally healthy children self-report using the EQ-5D-Y-3L questionnaire. Employing a retrospective think aloud approach, the study found that children encounter difficulties in the comprehension of the dimension, applying relevant information to the response categories as well as retrieving the relevant information from the specified time frame. Children experiencing these difficulties reported a significantly lower HRQoL values as compared to those without such issues.

Consistent with the findings from other research, HRQoL is influenced by both gender [137, 329-334] and age [137, 330-334]. In a study of Australian adolescents (aged 11-17 years), girls and older adolescents were reported to have significantly lower HRQoL than their respective counterparts, assessed using the EQ-5D-Y-3L [313]. Literature suggests that the challenges encountered during transitioning into puberty potentially results in the impaired HRQoL [331]. However, in this study, even younger girls (aged 6-10 years) reported lower HRQoL compared to boys. Moreover, the significant gap between girls and older children aged 8-10 years, and proxy reported HRQoL, resulting in overestimation by proxies, was observed across all value sets. The source of the discrepancy remains uncertain. It may stem from psychosocial-related factors, with girls or older children reporting issues within specific dimensions that are not fully recognised by their parents/caregivers. For instance, in a Swedish study, adolescent girls (aged 13-18) and older adolescents (aged 15-16) reported more problems with the “doing usual activities”, “having pain or discomfort” and “feeling worried, sad or unhappy” dimensions of the EQ-5D-Y-3L compared to boys and younger counterparts [137]. This implies that that proxy-reports may not be reliable for girls and older children, highlighting potential gaps in understanding this group’s specific HRQoL issues.

The systematic review and meta-analysis in Chapter 2, investigating child-parent agreement using preference-based HRQoL measures, reported low overall agreement (ICC=0.49) [222]. The HRQoL values derived using both adult and ‘Y’ child-weights exhibited relatively higher (good) levels of agreement across the overall sample (ICC= 0.62 to 0.71). In the study presented in Chapter 4, in the South Australian dyadic sample (Study 1), the inter-rater agreement with the EQ-5D-Y was found to be only fair (CCC= 0.28). This study was conducted through in-person interviews with dyads, during which children and their parents separately reported the child's HRQoL. However, a limitation of the present study is its reliance on an online sample of dyads wherein it cannot be determined whether proxies influenced the child-reports and vice-versa.

Across the dimensions, good to very good agreement was observed for the overall sample and for all subgroups except for "feeling sad, worried, or unhappy". Although this dimension reported good agreement, the agreement coefficient was the lowest compared to all other dimensions. Notably, only a moderate agreement (AC_1=0.58) was observed in children with reported health conditions, underscoring a theme of varying perceptions of HRQoL in this dimension based on health status [222]. Previous studies have shown that agreement between child-proxy dyads in the psycho-social dimension of "feeling sad, worried, or unhappy" tends to be low, often attributed to its subjective nature [172, 194, 195, 250, 335]. In contrast, as presented in Chapter 5, within the CHU9D's overlapping "sad" dimension, higher agreement between children and proxies has been observed [269]. This discrepancy may stem from the CHU9D treating "sad" and "worried" as separate dimensions, unlike the EQ-5D-Y, which combines these psycho-social aspects into a single "feeling sad, worried, or unhappy" dimension [313].

A limitation of this study is that the sample does not reflect the diversity of the Australian population due to the online nature of the survey administration, restricting participation to child-proxy dyads with internet access. This limitation potentially affects the generalisability of the findings to the wider population. Despite this, the online methodology remains a valuable tool for engaging diverse participant groups. Furthermore, the study consisted of predominantly healthy children, and a broader representation of children experiencing health conditions would enable a more comprehensive analysis. Future research should aim to employ more inclusive sampling strategies that encompass a broader spectrum of health statuses and socio-demographic characteristics to overcome these limitations.

8.8 Conclusions

The findings of this study indicate that the choice of value set may depend on the specific outcome of interest. Employing child-specific weights for the EQ-5D-Y led to more uniform outcomes across the ‘Y’ value sets when examining discrepancies between child and proxy reports of child HRQoL, than did utilising an Australian-specific adult value set. Thus, utilising any available ‘Y’ value set for assessing child HRQoL until the Australian-specific version is unavailable, rather than relying solely on the adult EQ-5D-3L value set is recommended. This finding also highlights a cultural consensus on the valuation of child health. On the other hand, the agreement coefficients for the HRQoL values derived using country-specific child value sets and the Australian adult weights were found to be comparable. These results also suggest that while there is a range of parent-child agreement across different cultural contexts as reflected by the ICCs, the variances are relatively narrow. The consistency of the ICCs across diverse international value sets implies that cultural differences in the valuation of HRQoL states may have a limited impact on the level of agreement between parent and child reports.

8.9 Supplementary information

Table 8-5 Comparison of child-proxy EQ-5D-Y-3L HRQoL values and agreement from Studies 1 and 2 using the German EQ-5D-Y-3L value set

Variable	**N (dyad)**	**Study 1**	**N (dyad)**	**Study 2**
Overall	85		845	
Child		0.93 (0.11)		0.89 (0.005)
Proxy		0.92 (0.10)		0.9 (0.005)
Difference		0.009		-0.011
p-value		0.50		0.01
Agreement		0.29 (0.08, 0.47)		0.69 (0.65, 0.72)
Subgroups:				
Gender				
Boy	37		419	
Child		0.95 (0.07)		0.89 (0.007)
Proxy		0.93 (0.09)		0.89 (0.007)
Difference		0.03		-0.005
p-value		0.12		0.37
Agreement		0.25 (-0.05, 0.50)		0.66 (0.6, 0.71)

Variable	**N (dyad)**	**Study 1**	**N (dyad)**	**Study 2**
Girl	47		424	
Child		0.91 (0.13)		0.89 (0.008)
Proxy		0.91 (0.12)		0.9 (0.006)
Difference		-0.001		-0.017
p-value		0.85		<0.001
Agreement		0.29 (0.01, 0.53)		0.72 (0.67, 0.76)
Age-group				
6-7 years	23		374	
Child		0.93 (0.09)		0.89 (0.008)
Proxy		0.92 (0.12)		0.9 (0.007)
Difference		0.009		-0.009
p-value		0.86		0.13
Agreement		0.23 (-0.17, 0.56)		0.68 (0.62, 0.73)
8-10 years	30		471	
Child		0.91 (0.14)		0.88 (0.007)
Proxy		0.91 (0.12)		0.9 (0.006)

Variable	**N (dyad)**	**Study 1**	**N (dyad)**	**Study 2**
Difference		-0.002		-0.013
p-value		0.77		0.02
Agreement		0.42 (0.08, 0.67)		0.69 (0.64, 0.74)
Health condition				
No	59		592	
Child		0.94 (0.09)		0.91 (0.005)
Proxy		0.94 (0.07)		0.93 (0.004)
Difference		-0.004		-0.019
p-value		0.70		<0.001
Agreement		0.17 (-0.08, 0.40)		0.56 (0.5, 0.62)
Yes	26		253	
Child		0.91 (0.15)		0.83 (0.012)
Proxy		0.87 (0.14)		0.82 (0.011)
Difference		0.04		0.007
p-value		0.1		0.39
Agreement		0.34 (-0.04, 0.62)		0.74 (0.68, 0.79)

Table 8.5 presents the results of child-proxy HRQoL values and agreement for the EQ-5D-Y-3L (self and proxy version 1) from Study 1 (face-to-face sample) and Study 2 (online sample). The HRQoL values were derived using the German ‘Y’ value set, and the corresponding agreement was assessed using the ICC. Study 1 showed higher mean HRQoL values for both children (0.93, SD=0.11) and proxies (0.92, SD=0.10), with minimal differences (0.009, p-value= 0.50) and moderate agreement (ICC = 0.29). In contrast, Study 2 revealed lower mean HRQoL values for children (0.89, SD=0.005) and proxies (0.90, SD=0.005), with a significant difference (-0.011, p-value= 0.01) but substantially higher agreement (ICC= 0.69).

When comparing the subgroups, agreement was consistently higher in Study 2 across gender, age group, and health condition subgroups, whereas Study 1 exhibited greater variability, with notably lower agreement in boys (ICC = 0.25) and children without health conditions (ICC = 0.17). Differences between child and proxy-reported values were consistent across most subgroups with varying levels of statistical significance, except for boys and 6–7-year-olds. Specifically, girls, children aged 8–10 years, and those without a health condition rated themselves lower than proxies, while children with a health condition rated themselves higher than proxies.

These findings suggest that the larger sample size and standardised online data collection methods in Study 2 reduced variability and enhanced agreement, though they also highlighted statistically significant differences that were not evident in Study 1. Importantly, these differences remain small in magnitude, often below the minimally clinically important difference (MCID) for the EQ-5D-Y-3L (0.03), underscoring the need to separate statistical significance from policy relevance. The Chapter 8 results, in conjunction with those from Chapter 4, indicate that most of the observed differences by gender, age and health condition were consistent across datasets and reinforce the importance of considering these factors in interpreting HRQoL data.

CHAPTER 9: DISCUSSION

9.1 Overview

This chapter discusses the implications of the findings from both studies presented in the thesis. It addresses the research questions posed in Chapter 1, provides recommendations for future research, and highlights the significance of enhancing the currently used age-appropriate HRQoL measures. The chapter concludes with a reflection on the contributions of the thesis to the field of child HRQoL assessment.

9.2 Economic evaluation and HRQoL measurement in children

Advancements in health technologies such as pharmaceuticals, medical devices, and public health interventions have the potential to enhance health and quality of life outcomes [336]. Whilst they are often found to be beneficial, these innovations in healthcare technologies are also key drivers of the rising healthcare expenditures in Australia [337]. The Australian healthcare system is predominantly funded by government expenditures (State and Federal) [338]. There was a 6% increase in healthcare expenditures in the financial year 2021-2022 compared to the previous one, constituting approximately 10% of Australia’s GDP [339]. Given that the healthcare system operates within a constrained budget, it is imperative for regulatory bodies, such as the PBAC, to determine the ‘value for money’ of one intervention over another, especially for subsidy considerations [5]. Economic evaluations, particularly CUAs, are instrumental in guiding decision-makers in this process. As mentioned in Chapter 1, CUAs compare alternative interventions by examining the ratios of cost, measured in monetary terms ($AUD) to benefit, measured in terms of gains in health outcomes (commonly, QALYs) [4]. The validity of allocation decisions aimed at maximising value for healthcare budgets is, thus, contingent upon the accuracy of the benefit metric used (i.e., QALYs), which in turn rely on the precise measurement and valuation of HRQoL.

The measurement and valuation of HRQoL in the adult population, while complex, is relatively less challenging compared to children due to several factors. Typically, adults are able to articulate the state of their own health and the impact on their quality of life more clearly than children, who may lack the cognitive and linguistic capabilities to fully express their experiences and perceptions of health. Additionally, children's developmental stages may cause the same health issue to affect them differently depending on their age [340]. Consequently, the establishment of well-defined guidelines for the measurement and valuation of HRQoL in child populations remains an ongoing challenge [95].

In recent years, the valuation of HRQoL in child populations has garnered substantial attention [70, 81]. However, in addition to advancing the methods used to value HRQoL in children, it is also equally essential to evaluate the accuracy of the measurements provided by the HRQoL measures themselves. Given the challenges inherent in obtaining self-reports from children, proxy reports are often utilised to complement or a substitute for self-reports. Several pertinent questions arise in this context. Firstly, it is imperative to consider whose perspective is being utilised (child or proxy) in the assessments of child HRQoL. When proxy reports are used, it is essential to assess the extent to which they represent the child's self-reports. Secondly, for child-self reports, the validity of the responses to the HRQoL measures must be scrutinised. That is, to investigate if the child-self responses are genuinely reflective of the actual heath status of the child as intended by the measure. This thesis sheds light on these critical questions by evaluating the inter-rater agreement between child self-reports and proxy-reports as well as examining the validity of responses provided by children to preference-based HRQoL measures. It is important to note that the inter-rater gap examined may also depend on various factors, including the choice of value set, especially in the absence of a child-specific value set such as for the Australian EQ-5D-Y-3L. As a secondary aim, this thesis also investigated the impact of applying child-specific versus adult value sets to the EQ-5D-Y-3L measure and examined the resulting differences in reported HRQoL values and the corresponding inter-rater gap.

9.3 Meeting the challenges in measuring child HRQoL: Main findings

To address the research questions outlined in Chapter 1, two studies involving child-parent dyads in Australia were conducted. The extent to which these questions have been successfully addressed based on the findings from these studies is summarised below.

9.3.1 What is the level of inter-rater agreement found in existing literature between self-reports by children and proxy-reports of child HRQoL?

To answer this, a systematic review and meta-analysis was conducted to include all published literature on this topic until 7th March 2024. The use of correlation coefficient to measure agreement has been a statistical issue criticised in the literature [341]. Correlation coefficients, while useful in determining the relationship between the scores from the raters, do not adequately measure agreement or the degree to which the scores align. For instance, two raters may have a high correlation in their scores even if one consistently rates higher or lower than the other, which would not indicate true agreement. The systematic review in this thesis addresses this issue by considering this important distinction and analysing the agreement and correlation separately.

The review revealed a moderate inter-rater agreement (ICC= 0.50, 95% CI= 0.36 to 0.62) between children and proxies using child-specific preference-based measures. This level of agreement was higher than that reported in Study 1 discussed in Chapters 4 and 5 but lower than the agreement observed in Study 2, Chapter 8. HRQoL dimensions related to physical health and functioning demonstrated higher agreement compared to the psychosocial related dimensions (e.g., ‘‘feeling worried, sad or unhappy” and “having pain or discomfort” dimensions of the EQ-5D-Y measures). Among the types of proxies, parents demonstrated a higher level of agreement with children relative to health professionals, likely due to their closer perceptual access to their child’s health and behaviours. Additionally, agreement decreased with the increased severity of health conditions. For instance, studies involving children with paediatric cancer exhibited lower levels of inter-rater agreement relative to non-cancer related studies. This may be attributed to the complexity and variability of such severe health conditions, coupled with the subjective nature of pain and emotional distress [342], leading to larger discrepancies between the child’s self and proxy-reports.

9.3.2 How does the proxy perspective influence inter-rater agreement in the measurement of child HRQoL?

In Chapter 4, a community-sample of 85 child (aged 6-12 years) and parent dyads provided responses to the EQ-5D-Y-3L self, proxy-proxy, and proxy-child reports, respectively. The findings indicate that the perspective adopted by parents when responding as proxies to the EQ-5D-Y-3L did not impact the overall agreement. The overall agreement between self-reported and proxy-reported HRQoL was low (fair) across both proxy-proxy (CCC=0.28) and proxy-child (CCC=0.26) perspectives. However, when the agreement was examined at the dimension level, a notable difference was observed with the “feeling worried, sad or unhappy” dimension. As mentioned above, a low inter-rater agreement is often seen with psychosocial dimensions due to their subjective nature. However, by asking the parents to switch their perspective from their own to that of their child, a substantially higher agreement was observed, comparable to those with the physical health-related dimensions. The EQ VAS results further illustrated this discrepancy between proxy perspectives. When parents adopted their child’s perspective, they significantly overestimated the child’s health status on the EQ VAS. This suggests that parents, when trying to adopt their child’s viewpoint, tend to believe their children perceive their health more positively than they actually do. In reality, children may be more critical of their own health, considering factors that may not be fully apparent to their parents. This was also evident in a study by Theunissen et al. among relatively healthy children who reported lower a HRQoL compared to their parent-reports [343].

9.3.3 How does the age of the child influence the inter-rater agreement in the measurement of child HRQoL?

The age of the child may play a role in the inter-rater agreement between child self-reports and proxy reports in the measurement of child HRQoL, depending on the measure used. In Chapter 5, the inter-rater agreement was explored using the CHU9D and the PedsQL^{TM} for the overall HRQoL and the nine dimensions of the CHU9D and the overlapping PedsQL^{TM} across different age groups (6-7, 8-10, and 11-12 years). Varying levels of overall and dimension-level agreement was observed across the age-groups.

The overall agreement suggested poor agreement for 6–7-year-olds in the sample, with both the CHU9D and the PedsQL^{TM}. The overall agreement was higher for the PedsQL^{TM} than for the CHU9D. While both measures are designed for child populations, the CHU9D is preference-based [56, 344], whereas the PedsQL^{TM} is not [246]. The different weights assigned to various dimensions in the CHU9D may have caused these dimensions to contribute differently to the overall score, impacting the corresponding agreement.

For the CHU9D dimensions, all the three age-groups generally exhibited comparable levels of agreement with their parents across most dimensions. The only exception was the “tired” dimension, where the agreement was statistically non-significant (AC_1= 0.19). This suggests that younger children may have more difficulty accurately conveying their health status or that parents may have more difficulty interpreting this dimension in relation to younger children. In contrast, for the PedsQL^{TM} items, the youngest age group (6-7 years) showed higher levels of agreement with their parents relative to the older age groups (8-10 and 11-12 years). Notably, insignificant agreement was observed for psychosocial health-related items across the two older age groups. These findings suggest that as age increases, the divergence between child self-reports and their parents' increases, particularly for the more subjective dimensions of the PedsQL™. This finding contradicts an earlier study conducted by Cremeens et al., which showed the largest differences in reported HRQoL in older age groups (6.5 years and above) using the PedsQL™. However, they found no significant differences between child-parent reports when the child was between 5.5-6.5 years [126]. There is some overlap between the age groups in this study and Study 1, which may account for the differing results.

The findings for the overall and the dimension-level agreement across the age-groups presented differing narratives. This discrepancy may arise during the aggregation of HRQoL scores; high agreement in some dimensions may offset low agreement in others, leading to an overall agreement that does not fully capture the variability observed at the dimension level. Moreover, in the case of the CHU9D, if certain dimensions have more weights in the overall score calculation, the agreement in those dimensions will have a larger impact on the overall agreement. Consequently, disparities in less weighted dimensions may not significantly affect the overall agreement. Additionally, the differences in dimension-level inter-rater agreement for the two measures highlight the nuanced distinctions between their descriptive systems. For example, the inter-rater agreement for HUI 2 was higher than that for HUI 3, despite both measures evolving from the same HUI classification system and being designed to be complementary in nature. Although they have attributes with the same names, HUI 2 and HUI 3 reflect different underlying constructs [202] as discussed in Chapter 2. These differences suggest that the wording or format of child-specific HRQoL measures may also impact inter-rater agreement.

In addition, the findings from chapters 4 and 8 (Table 8.5) suggest that the age of the child may influence agreement between child and proxy-reported HRQoL values. For younger children aged 6-7 years, agreement was lower in both studies, with an ICC (equivalent to

CCC) of 0.23 in Study 1 and 0.68 in Study 2. In comparison, older children aged 8-10 years demonstrated higher agreement, with an ICC of 0.42 in Study 1 and 0.69 (slightly higher) in Study 2. The differences between child and proxy-reported values were small and non-significant for the younger age group in both studies (Study 1, p-value= 0.86; Study 2: p-value= 0.13). For older children, the differences were small but statistically significant in Study 2 (p-value = 0.02), though not in Study 1 (p-value= 0.77).

While the data suggests that agreement increases with age, this interpretation should be approached with caution. Although 6-7-year-olds exhibited lower agreement levels, older children aged 8-10 years were also identified to have response issues, as observed in Chapters 5 and 6. Additionally, the differences between child and proxy reports in the larger sample from Study 2 were small in magnitude, further complicating the interpretation of age-related trends in agreement. Future studies could address these uncertainties by employing longitudinal designs to explore whether agreement improves as children mature, while larger, more diverse samples could provide more robust insights into the role of age and developmental factors in child-proxy agreement.

9.3.4 How well do children understand the dimensions of HRQoL measures?

The mixed methods study presented in Chapters 6 and 7 provided valuable insights into this question. The qualitative aspect of Study 1 illustrated how children understood and interpreted the HRQoL dimensions presented in the EQ-5D-Y-3L and CHU9D measures. Response issues, defined as deviations from responding in a manner aligned with the construct's intent, were classified using Tourangeau's framework into comprehension, judgment, recall, and response mapping issues. The sample was divided into two groups: those participating with the EQ-5D-Y-3L and those with the CHU9D measure in the qualitative think aloud interviews.

Interestingly, across both the samples, nearly half the participants were categorised as having a response issue predominantly related to judgment. These response issues translated into significantly lower reported HRQoL for children with response issues compared to those without. For the EQ-5D-Y-3L, the median HRQoL score was 0.81 (IQR= 0.1) for children with response issues versus 1 (IQR= 0) for those without. Similarly, for the CHU9D, the median HRQoL score was 0.78 (IQR= 0.2) for children with response issues versus 0.90 (IQR = 0.09) for those without. This discrepancy was also reflected in the inter-rater agreement. For the EQ-5D-Y-3L, the inter-rater agreement in the subgroup with response issues was lower within the dimensions where the most response issues were noted, specifically the subjective dimensions such as “doing usual activities” and “having pain or discomfort”. This was consistent with the findings in the available literature presented in Chapter 2 [222], potentially suggesting that these dimensions are more prone to misinterpretation or misunderstanding by children. In contrast, the CHU9D showed a slightly different pattern. While dimensions with the most response issues (“joining in activities” and “daily routine”) also exhibited lower inter-rater agreement in the subgroup with response issues, other dimensions like “sad” showed lower agreement within this group even though no response issues were observed in that dimension in the overall sample. This could potentially reflect genuine disagreement rather than response issues.

The discrepancy in self-reported HRQoL between the two groups may suggest that the overall HRQoL is capturing aspects of their well-being that are not solely health-related, as intended by the measures. It is important to note that terms like “daily activities” or “daily routine” may be understood differently by younger children, because unlike adults, who self-direct their daily activities, children often have these activities imposed upon them, and these can vary depending on what their parents/guardians have outlined. Consequently, for children, adhering to a routine or performing their daily activities may also be interpreted as their ability to meet the expectations set by their parents/guardians. This can lead to misunderstandings in interpreting HRQoL dimensions, as children may conflate their health-related ability to perform necessary tasks with their willingness or motivation to engage in those activities.

9.3.5 What is the minimum age at which children can reliably self-report their HRQoL?

The findings presented in Chapter 4 demonstrate that the lowest inter-rater agreement across age groups was observed for three of the five dimensions of the EQ-5D-Y-3L, particularly the “doing usual activities” dimension among 6–7-year-olds. However, Chapter 5 noted that with the CHU9D, 6–7-year-olds showed a comparable, and even marginally higher, inter-rater agreement relative to older age groups. This discrepancy could be attributed to the differences in the design of these measures. The EQ-5D-Y was adapted from the adult measure EQ-5D to enable the tracking of changes in HRQoL from childhood through adulthood [86]. Although the language was modified to be more child-friendly, the descriptive system may still retain adult-centric concepts that younger children find challenging to interpret. In contrast, the CHU9D was developed using a ‘bottom-up’ approach, specifically developed with the involvement of children, perhaps making it inherently more comprehensible and relatable for younger age groups [90]. Consequently, children may find the CHU9D easier to understand and respond to accurately, resulting in higher inter-rater agreement for this measure in this age-group.

Further analysis in the mixed methods study suggested a correlation between the understanding of the “doing usual activities” dimension and the corresponding low inter-rater agreement. Younger children may have misunderstood the intended meaning of the “doing usual activities” dimension, frequently recalling non-health-related factors in their responses. This misunderstanding likely contributed to the low inter-rater agreement observed, rather than reflecting genuine differences in perspectives on child HRQoL. Furthermore, the proportion of response issues was higher among 8–10-year-olds, 8 years being the recommended age threshold for the EQ-5D-Y-3L, compared to 11–12-year-olds. For the CHU9D, the lowest inter-rater agreement was observed for the “tired” dimension among 6–7-year-olds. However, findings in Chapter 7 indicated no response issues within this dimension, suggesting genuine disagreement between the dyads. Nonetheless, a marginally higher proportion of response issues was observed among the younger age groups (10 years and below) in this sample compared to the older age group. These issues were predominantly noted in the “joining in activities” and “daily routine” dimensions. In contrast, children aged 11 years and older demonstrated a better understanding of the CHU9D dimensions.

Overall, the evidence for the minimum age of self-report in children was mixed. The findings in this thesis indicates that children aged 11 years and older can reliably self-report their HRQoL. They showed fewer response issues and higher inter-rater agreement with parental proxy reports, indicating a better understanding and ability to accurately reflect their HRQoL using the EQ-5D-Y-3L and the CHU9D. However, it is important to acknowledge that some younger children (e.g., 6-7 years old) demonstrated the potential to reliably self-report HRQoL using measures like the CHU9D. Children aged 10 years and younger have more difficulty in meaningfully self-reporting their HRQoL. The presence of significant response issues and lower inter-rater agreement suggested that these children may struggle with understanding and meaningfully responding to HRQoL measures. Additionally, the discrepancies in inter-rater agreement may not solely reflect limitations in children’s ability to self-report but could also arise from parents’ difficulties in accurately reporting their younger children’s HRQoL using these descriptive systems. Consequently, it cannot be conclusively stated that the current format of the EQ-5D-Y-3L and the CHU9D is adequately adapted for children between 6-10 years of age to reliably report their HRQoL.

9.3.6 How does the use of different value sets impact the inter-rater gap when utilising the EQ-5D-Y-3L?

Study 2, presented in Chapter 8, addressed a significant limitation of Study 1, which utilised an Australian adult scoring algorithm to evaluate the self and proxy-reported HRQoL values generated using the EQ-5D-Y-3L. This approach was suboptimal as a child-specific value set is not currently available in Australia, raising questions about the most appropriate value set to use to minimise bias in this context. To explore this, Study 2 employed a large online sample of 845 child-parent dyads (children aged 6-10 years) from across Australia to investigate the impact of applying different EQ-5D-Y-3L (‘Y’) value sets from nine countries as well as the Australian EQ-5D-3L adult value set on the evaluation of self and proxy discrepancies in HRQoL values.

The use of different value sets resulted in varied HRQoL values due to differences in the weighting of dimensions, as well as aggregation methods. However, child-specific value sets produced more consistent outcomes in evaluating the inter-rater gap in HRQoL values compared to the Australian adult value set. Proxy ratings were consistently higher than child self-reports across all ‘Y’ value sets, with statistically significant differences (mean difference range= 0.007 to 0.016; p-value<0.05). This statistical difference in child-proxy HRQoL values was not observed with the Australian adult value set. Despite these variations, the inter-rater agreement was mostly consistent across all value sets (both adult and ‘Y’), with ICC values ranging from 0.62 to 0.71, which was higher than what was observed in Study 1. This difference may be attributed to the lower power of the first study. Additionally, in Study 1, interviewers ensured that child and parent participants completed the HRQoL measures independently. In contrast, for Study 2, while participants were instructed to complete the measures separately, the online nature of the study made it difficult to enforce this, potentially allowing for collusion between the child and the parent.

Importantly, as noted in Section 8.9, the results from Chapter 8, which included a larger sample size, demonstrated lower variability in HRQoL ratings, contributing to the higher agreement observed in the Study 2 sample. In contrast, the higher variability in Study 1, as reflected in the larger standard deviations, likely influenced the lower agreement between child and proxy ratings. While agreement was lower in Study 1, the higher variability may suggest that it captured a broader range of child-proxy perspectives, as observed in Chapter 4 from the child's perspective. As noted in 9.3.4, this was particularly evident in the predominantly healthy population, where children provided more diverse responses when completing the HRQoL measure independently with an interviewer that often incorporated non-health-related perspectives.

Despite this, statistically significant differences between child and proxy-reported HRQoL values were observed in Study 2, primarily due to the increased power of the larger sample size. However, it is important to note that these differences were smaller than the MCID of 0.03, indicating they might have weaker clinical relevance in real-world contexts, such as healthcare policy or resource allocation. The findings from this thesis must therefore be contextualised within the objectives of the research and the implications for practice.

9.4 Implications and future research directions

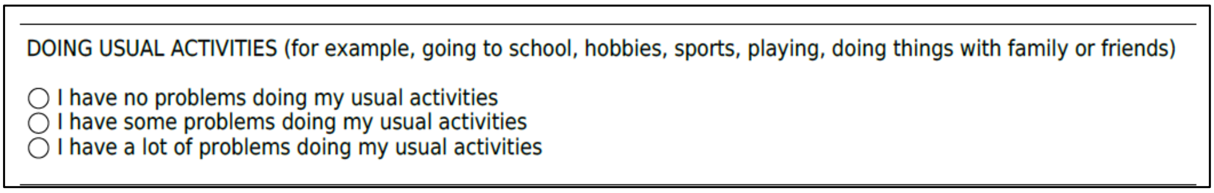
In Australia, the PBAC does not mandate a specific approach for the measurement and valuation of HRQoL, but is advised that the utility weights used should be representative of the general Australian population [5]. However, the lack of clear and consistent guidelines poses challenges in interpreting evidence derived from child-specific PBMs using both child-self and proxy reports to inform decision-making [309]. Among the available child-specific utility measures, a recent systematic review identified the EQ-5D-Y as the only preference-based HRQoL measure offering the comprehensive guidance for both self and proxy reporting on their website [345]. The evidence presented in this thesis and the findings emanating from it, aim to contribute to the development of standardised guidelines for child and proxy reporting of child HRQoL across most, if not all, child-specific generic measures currently in use.

It is important to note that the position adopted in this thesis aligns most closely with the perspective that there is no single ‘true’ latent HRQoL value, but rather multiple valid perspectives—those of the child and the proxy—each shaped by the individual’s experiences, observations, and interpretations. HRQoL is a multidimensional construct with ‘true’ values determined by agent and perspective. Consequently, the findings of this thesis are evaluated through this lens, recognising that discrepancies between child and proxy reports do not necessarily indicate error but reflect differences in lived experience and perspective. In line with this, it should also be noted that increasing inter-rater agreement should not always be the ultimate goal. Instead, self- and proxy-reports should be seen as complementary sources of information, each contributing unique insights into the child’s HRQoL.

Perspective is a crucial consideration in measuring child HRQoL. To enhance the consistency and reliability of HRQoL assessments, it is essential to provide detailed guidance on when and how to adopt different proxy perspectives. Adopting a proxy-child perspective can be particularly beneficial in increasing concordance, especially when assessing key dimensions such as mental and emotional well-being. This approach involves proxies, such as parents or caregivers, actively considering the child’s point of view and experiences rather than relying solely on their own perspective.

The results of this thesis suggest that there may be an interaction between the child’s age and the HRQoL measure used. Although studies have explored this interaction [216-219], the results have been inconclusive. Future research should explore this interaction in larger, more diverse samples and through longitudinal studies to track changes in child-proxy discrepancies in HRQoL reports as children age, using child-specific generic preference-based measures. Understanding how agreement evolves over time may provide valuable insights into developing more effective age-appropriate HRQoL measures.

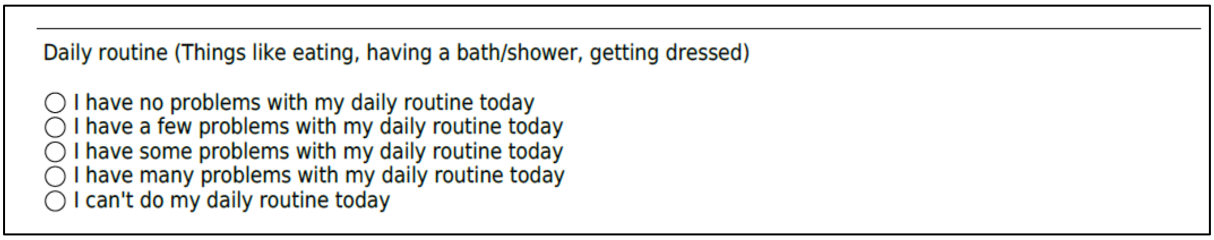
A key recommendation is to enhance the age-appropriateness of child-specific HRQoL measures. The findings do not imply that children below 11 years of age should not self-report; instead, they highlight the need to improve current measures to make them more suitable for younger children. For instance, the mixed methods study found that children often misunderstood the meaning of “routine” in both the EQ-5D-Y and the CHU9D. The current phrasing may not provide enough context for children to understand the intent of the question. Children often conflated their ability to perform necessary daily tasks with their typical variations in daily routines influenced by their interests and motivations. To address this, the question could be rephrased to include clarifying statements. The EQ-5D-Y-3L “doing activities” and the CHU9D “daily routine” dimensions are currently phrased as given in the figure below:



DOING USUAL ACTIVITIES (for example, going to school, hobbies, sports, playing, doing things with family or friends)

○ I have no problems doing my usual activities
○ I have some problems doing my usual activities
○ I have a lot of problems doing my usual activities

Figure 9.1a. EQ-5D-Y-3L “doing activities” dimension.



Daily routine (Things like eating, having a bath/shower, getting dressed)

○ I have no problems with my daily routine today
○ I have a few problems with my daily routine today
○ I have some problems with my daily routine today
○ I have many problems with my daily routine today
○ I can't do my daily routine today

Figure 9.1b. CHU9D “daily routine” dimension.

For these dimensions adding a phrase such as, “I have no problems…*due to my health*” in the response options may add more context and offer the respondent a clearer framework to guide their thought process. This phrasing may also be effective for the “having pain or discomfort” dimension wherein children often considered circumstances that resulted in “emotional pain” to respond to this dimension. For the “daily routine” and “joining in activities” dimension of the CHU9D, phrases such as “…if I want to” could be added in the examples could further clarify the intent. For instance, as shown in Fig. 9.1b, the example could be amended to “Things like eating, taking a bath/shower, getting dressed *if I want to*”. Additionally, incorporating illustrations or audiovisual components may be tested to further enhance understanding.

While the recommendation to adapt the EQ-5D-Y-3L and CHU9D to better suit younger children may help addresses the identified challenges of self-reporting in this age group, it is not without its limitations. One significant critique of such adaptations is that they may undermine the comparability of the measures across different versions. These validated standardised HRQoL measures derive much of their value from facilitation of consistent comparisons across diverse populations, settings, and age groups. Introducing age-specific adaptations risks creating disparities in how HRQoL is measured and interpreted, potentially limiting the applicability of results in broader health economic evaluations and cross-group analyses. It is important to note that any adapted versions would require further validation and psychometric testing to ensure they are both reliable and valid and to establish equivalence between the adapted and original version.

Another reason children did not report their health status as intended is that they used a different benchmark for comparison, other than assessing their health condition in terms of illness or wellness. Specifically, children often compared themselves to an ‘ideal’ version of themselves, considering optimal lifestyle habits, such as eating healthily and exercising. This was evident from their responses to the EQ VAS in the think aloud stage, where some children rated their health against this idealised standard rather than their actual health status. This lack of a clear benchmark is not only an issue for children but may also be for proxies. For instance, in a study by Blackmore et al., caregivers of children with intellectual disabilities reported the HRQoL of these children by proxy using the EQ-5D-5L. During the think aloud interviews, caregivers expressed uncertainty about the basis of comparison they should use, whether to compare the child to their peers or the health status of child themselves [346]. To address this issue, including contextual guidelines in the HRQoL measures could potentially improve consistency among respondents.

Another recommendation is to use any available ‘Y’ value set for assessing child HRQoL until Australian-specific version is unavailable, rather than relying solely on the adult EQ-5D-3L value set. The findings in this thesis underscore the importance of carefully selecting value sets for evaluating child HRQoL and suggest further research into developing and applying child-specific preference weights.

The inter-rater agreement between child self-reports and parent proxy reports was observed to be low. The findings of this thesis suggest that this low agreement may be influenced by several factors, including the specific HRQoL measure used, the perspective adopted by the proxy respondent, and the age of the child. However, it was also noted that inter-rater disagreement may not solely arise from genuine differences in perspective but also from children misinterpreting the intent of the HRQoL dimensions. An important consideration is the significant exposure of children today to various information sources through the internet, where they often search for health-related information [347], which in turn shapes their ideas about health and well-being. This suggests that the self-report issue in children may not be purely cognitive but also related to whether the measures are well-suited in their concepts and phrasing for use with younger populations (especially in ages 6-7 years). For example, terms and constructs in the HRQoL measures may not resonate with children’s experiences or may be misunderstood, leading to discrepancies in reporting. This issue is particularly significant from a longitudinal perspective where the child's HRQoL is being tracked over time. Changes in interpretation of the HRQoL dimensions as children grow older can lead to inaccuracies in their self-reports, further complicating the assessment of their HRQoL.

It is also crucial to question whether increasing inter-rater agreement should be the goal. Low inter-rater agreement indicates that proxy reports may not serve as direct substitutes for child self-reports but could potentially complement them. To understand this better, it is necessary to analyse proxy responses in the context of the child's actual experiences. This could involve further qualitative studies to assess how well parent reports align with the child's self-reported experiences when both reports are considered together. Such an approach would provide a more comprehensive understanding of the child's HRQoL, acknowledging both the child's perspective and the insights offered by the proxy. In addition, it would be valuable to explore inter-rater agreement based on parental gender in future research. Differences in agreement may reflect variations in time spent with the child, as mothers often assume primary caregiving roles. Investigating this aspect could provide important insights into how caregiving dynamics influence proxy reporting and the interpretation of child HRQoL.

9.5 Limitations

The main limitation of the two empirical studies in this thesis, Studies 1 and 2, was that the children in the sample were from the general population and may not have had experience or understanding of the impact of health conditions on the dimensions included in the measures used. However, this cohort is often used in research-related (e.g., studies relating to psychometric validation of measures) and clinical studies (e.g., studies with general population children as controls), making these findings particularly relevant. Another limitation was the sample size in Study 1. While sufficient for preliminary analyses, the low power of the study may limit the generalisability of the findings. The first empirical study (Study 1) was conducted with a relatively homogeneous group of participants from local areas with an insufficient number of participants from lower socioeconomic status groups, which may not be representative of the broader population. Due to the low power of the sample, it could not be conclusively determined whether the presence of illness was driving the observed lower agreement. However, in the larger sample analysed in Chapters 4 and 5, agreement was slightly lower in the subgroup of children with a health condition. This observation suggests that health status may indeed play a role in influencing self-proxy agreement, although this finding requires further validation in a larger and more representative sample of children with health conditions. This thesis also predominantly included mothers as the parental proxy in assessing child HRQoL, which is common in much of the existing child-caregiver research. The potential influence of parental gender (e.g., mothers versus fathers) as a source of variation in proxy-reported HRQoL was not explored, due to sample size constraints. Future research should aim to include larger, more diverse samples in terms of health conditions, socioeconomic status and parental gender to enhance the external validity of the results.

The thesis focused on specific measures, namely the EQ-5D-Y, CHU9D, and PedsQL™, which are the most frequently used in economic evaluations. Although these measures are highly relevant, the exclusion of other potentially useful measures may limit the comprehensiveness of the HRQoL assessment in children. Additionally, the qualitative findings were analysed using a framework analysis, which may have constrained the scope of the results. An inductive analysis approach could have potentially revealed other findings not captured in this study, providing a broader understanding of the data.

In Study 1, while parents completed the measures independently, an interviewer was present with the child to provide assistance if needed. Despite efforts to minimise potential bias introduced by the interviewer’s presence, this remains a limitation. The presence of an interviewer could inadvertently influence the child’s responses due to social desirability bias [348], where the child may alter their answers to align with perceived expectations of the interviewer. Additionally, the interviewer's presence could affect the child's level of comfort and engagement with the questionnaire, potentially confounding the accuracy and authenticity of the responses. This suggests that the mode of completion by the children and the parents may not have been entirely equivalent. Specifically, in Chapter 4, despite the distraction task of completing the EQ-5D-3L for themselves in between each proxy task, given that proxy-child report was completed subsequent to 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 [275] or consider randomising the order in which the two proxy reports are administered.

In Study 2, the sample consisted of an online group and demonstrated higher agreement. Although participants were instructed to complete the measures independently, this could not be enforced, and it is possible that children may have received some help from their parents or guardians. This potential lack of independence may have affected the degree of agreement estimated for the dyads in this sample. However, the sample was used to examine the impact of using the available EQ-5D-Y-3L value sets on inter-rater gap and agreement. Since the potential systematic error due to collusion between the child and the parent would be consistent across all value sets being compared, it would not impact the relative differences observed between the value sets. As such, the findings are independent of the nature of completion by children and their parent/guardians.

Despite these limitations, the studies in this thesis offer valuable insights into the measurement of HRQoL in children, with the innovative use of both quantitative and qualitative methodologies significantly strengthening the overall findings and contributing to the field of child HRQoL research.

9.6 Conclusions

This thesis presents the first comprehensive research conducted in Australia to examine child-proxy agreement in the assessment of child HRQoL and to evaluate children's understanding of HRQoL measures using an in-depth mixed methods approach. Key findings from this study include several important insights into the methodological challenges in the measurement of HRQoL in children. The different proxy perspectives (proxy-proxy or proxy-child) with the EQ-5D-Y-3L did not significantly affect overall inter-rater agreement, though the agreement for psychosocial dimensions improved when parents adopted the child's perspective. Age influenced inter-rater agreement, with child-specific measures specifically designed for children, namely the CHU9D and the PedsQL^{TM} showing comparable agreement in younger children (6-7 years) relative to older children (8-10, 11-12 years). In contrast, measures adapted from adult versions, i.e., the EQ-5D-Y-3L response issues were exhibited particularly among younger children (6-10 years), who often misunderstood specific HRQoL dimensions, leading to lower HRQoL values that captured non-health-related problems. Furthermore, employing child-specific value sets resulted in more consistent HRQoL outcomes than adult value sets, highlighting the importance of using child-specific values to appropriately assess health outcomes in child populations.

The evidence gathered from this research is invaluable for informing the development of comprehensive guidelines to guide researchers, policymakers, and practitioners in the field of child health. These guidelines will enhance the consistency and reliability of HRQoL assessments in economic evaluations and other studies involving child populations, cross sectionally or longitudinally, such as the Longitudinal Study of Australian Children (LSAC).

By addressing the nuances of child and proxy reporting, these guidelines can ensure more accurate and meaningful evaluations of children's HRQoL.

CHAPTER 10: APPENDIX

10.1 Supplementary information for Chapter 2

Table 10-1 Search Strategy (Adapted from Khadka et al., 2019)

Database: PubMed, Embase, Web of Science, PsycINFO, EconLit, CINAHL, Cochrane Library

Last run date for Khadka et al., 2019: 1.07.2017

First run, include articles with publication date from 30.6.2017 to 19.05.2021.

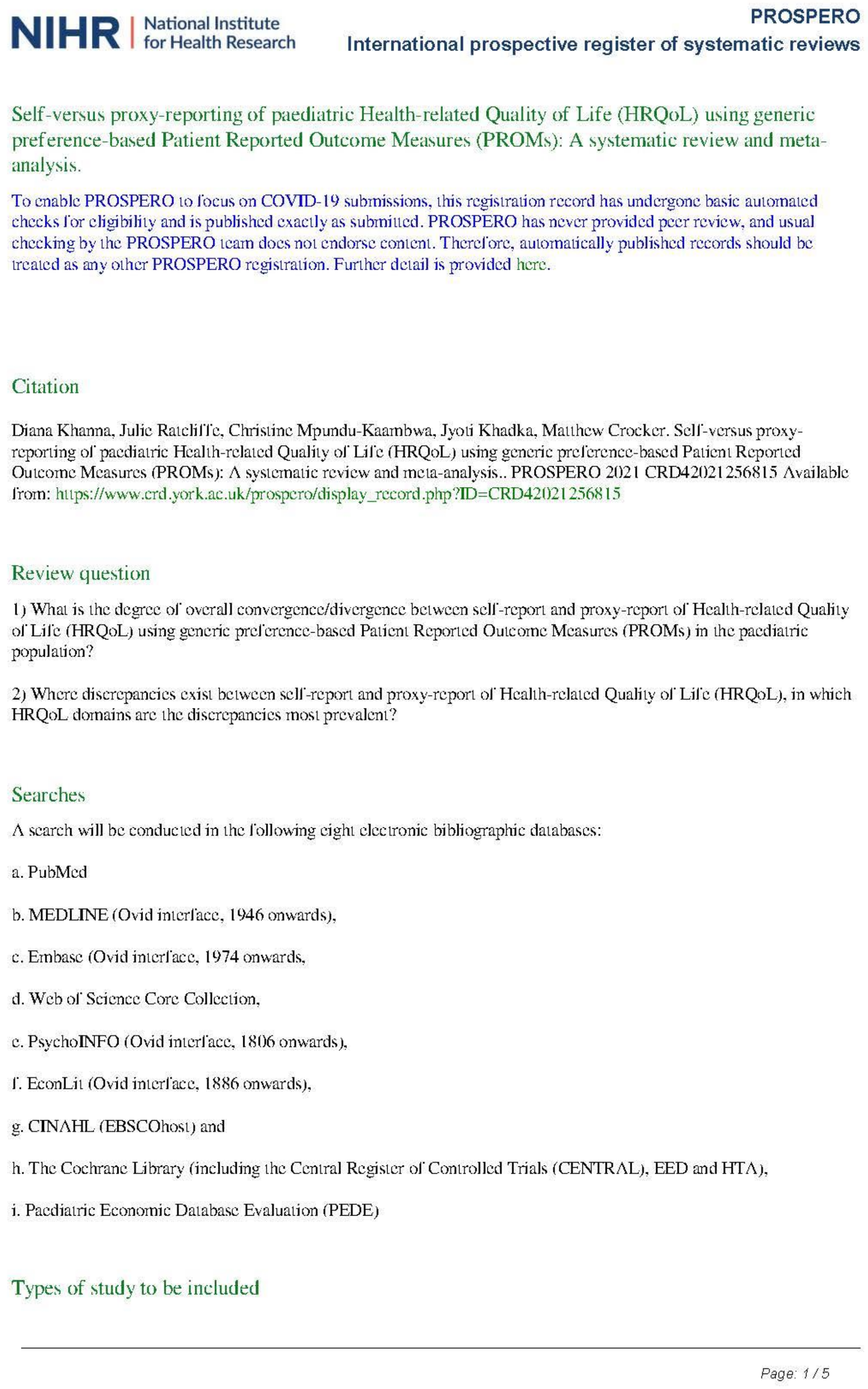
Secon run, include articles with publication date from 19.05.2021 to 7.03.2024.

Limit search to title and abstract

Search Category		**Search Terms**
Utility Terms	1	Utilit* or disutilit* or HSUV
	2	“quality adjusted life year*” or QALY or “quality-adjusted life year*” or “quality-adjusted life-year*”
	3	**OR (1 to 2)**
Indirect Valuation Method Terms	4	EQ-5D or “EQ 5D” or EQ5D or Euroqol or “Euro qol” or EQ-5D-Y or "EQ 5D Y"
	5	Short-form survey-6D or short form 6D or SF-6D or “SF 6D” or SF6D
	6	“health utilities index”
	7	“quality of well being” or “quality of well-being” or QWB
	8	16D Health-Related Quality of Life or 16D HRQoL or 17D Health-Related Quality of Life or 17D HRQoL
	9	AQoL-6D or Assessment of Quality of Life-6D
	10	“Child Health Utility 9 Dimension” or CHU9D or CHU-9D or “CHU 9D”
	11	Adolescent Health Utility Measure or AHUM
	12	15-dimensional instrument or 15 dimensional instrument
	13	preference-based measure of HRQoL or preference based measure of HRQoL
	14	multi-attribute utility instrument or multiattribute utility instrument
	15	**OR (4 to 14)**
	16	Standard Gamble or standard-gamble

Direct Valuation Method Terms	17	Time trade off or time trade-off
	19	best worst scaling or best-worst scaling
	19	Discrete choice experiment or discrete-choice experiment
	20	person trade off or person trade-off
	21	scoring algorithm or scoring-algorithm
	22	utility elicitation or direct elicitation
	23	**OR (16 to 22)**
	24	**3 OR 15 OR 23**
Childhood Terms	25	Child* or adolesc* or kid or kids or youngster* or teen* or youth* or infant* or newborn* or perinat* or neonat* or “parent proxy”
	26	Pediatri* or paediatri*
	27	**OR (25 to 26)**
Main Search	**28**	**24 AND 27**
	29	**Remove non-English Title and/or Abstract**
	30	**Remove Duplicates Across Databases**
	31	**Only include publications from year “2017 to Current”**

Figure 10.1 PROSPERO Registration document.



NIHR | National Institute for Health Research

PROSPERO
International prospective register of systematic reviews

Self-versus proxy-reporting of paediatric Health-related Quality of Life (HRQoL) using generic preference-based Patient Reported Outcome Measures (PROMs): A systematic review and meta-analysis.

To enable PROSPERO to focus on COVID-19 submissions, this registration record has undergone basic automated checks for eligibility and is published exactly as submitted. PROSPERO has never provided peer review, and usual checking by the PROSPERO team does not endorse content. Therefore, automatically published records should be treated as any other PROSPERO registration. Further detail is provided here.

Citation

Diana Khanna, Julie Ratcliffe, Christine Mpundu-Kaambwa, Jyoti Khadka, Matthew Crocker. Self-versus proxy-reporting of paediatric Health-related Quality of Life (HRQoL) using generic preference-based Patient Reported Outcome Measures (PROMs): A systematic review and meta-analysis.. PROSPERO 2021 CRD42021256815 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021256815

Review question

1) What is the degree of overall convergence/divergence between self-report and proxy-report of Health-related Quality of Life (HRQoL) using generic preference-based Patient Reported Outcome Measures (PROMs) in the paediatric population?

2) Where discrepancies exist between self-report and proxy-report of Health-related Quality of Life (HRQoL), in which HRQoL domains are the discrepancies most prevalent?

Searches

A search will be conducted in the following eight electronic bibliographic databases:

a. PubMed

b. MEDLINE (Ovid interface, 1946 onwards),

c. Embase (Ovid interface, 1974 onwards,

d. Web of Science Core Collection,

e. PsychoINFO (Ovid interface, 1806 onwards),

f. EconLit (Ovid interface, 1886 onwards),

g. CINAHL (EBSCOhost) and

h. The Cochrane Library (including the Central Register of Controlled Trials (CENTRAL), EED and HTA),

i. Paediatric Economic Database Evaluation (PEDE)

Types of study to be included

Primary and secondary research studies will be included. There are no restrictions on the type of studies to be included in the search because it is expected that the number of studies in this field is relatively small.

Condition or domain being studied

Paediatric health-related quality of life/ paediatric health and well-being.

Participants/population

Inclusion: Paediatric population including early childhood (1 to <5 years), child (5 to <13 years) and young person (13 to <18 years) age group.

Exclusion: Infants (0 to <1 year) and young person > 18 years, adults (over 22 and <70) and older people (over 70)

Intervention(s), exposure(s)

Preference-based PROMs to measure health-related quality of life (HRQoL) in the paediatric population.

Comparator(s)/control

Not Applicable

Context

No restrictions will be applied in terms of the study setting.

Main outcome(s)

A descriptive assessment of the degree of convergence/divergence between self-report and proxy-report of health-related quality of life (HRQoL) using generic preference-based Patient Reported Outcome Measures (PROMs) in the paediatric population.

Measures of effect

Intraclass correlation coefficients (ICC) (and other measures of agreement e.g. Cohen's Kappa statistic).

Additional outcome(s)

Investigate the extent of agreement/disagreement between self-report and proxy-report of HRQoL in the paediatric population in each domain of the respective PROM by comparing the intraclass correlation coefficients (ICC) (and other measures of agreement e.g. Kappa statistic) for self-reported and proxy reported scores for each domain of the PROM.

Data extraction (selection and coding)

Titles and/or abstracts of articles resulting from the search will be screened independently by two review authors to identify the studies comparing self-report and proxy-report of health-related quality of life (HRQoL) using preference-based Patient Reported Outcome Measures (PROMs) in the paediatric population that meet the inclusion criteria. Screening of the titles and/or abstracts of potentially eligible studies will be conducted against predefined eligibility criteria. Reference lists of included articles will also be reviewed for supplementary literature not identified using the search strategy. Backward and forward citation chasing will be carried out to help confirm the saturation of the initial searches. Where necessary, the study authors will be contacted for clarification and additional information to inform study selection. Full text articles will be retrieved and reviewed in full by two reviewers. Each step of the selection process will be outlined in a PRISMA-style flow chart. Disparities will be resolved by discussion and consultation with the review team. Summary data of each included article will be extracted into a data extraction form specifically designed for this review. The information to be extracted will include:

1. Descriptive data participants: study setting; study population and participant characteristics;

2. Descriptive information about study: date of publication; country of origin; sample size; study type;

3. Information about the measure: name of PROM; domains/dimensions; number of items; description of the items; response method; method of administration; psychometric properties (content validity, construct validity, reliability, responsiveness, floor or ceiling effects, acceptability and feasibility); interpretation and summary scoring;

4. Information about valuation of measure: health states valued; preference elicitation method; population preference weights. and management. An example of such a software tool is the Systematic Review Data Repository-Plus.

Risk of bias (quality) assessment

-Two independent reviewers will appraise the quality and suitability of the studies that compare self-report and proxy-report of health-related quality of life (HRQoL) using preference-based Patient Reported Outcome Measures (PROMs) in the paediatric population.

-The overall reporting quality score will be calculated as given by in the paper "A Systematic Review and Meta-analysis of Childhood Health Utilities" by Kwon et al., 2018 will be used to assess bias (Kwon, J., Kim, S.W., Ungar, W.J., Tsiplova, K., Madan, J. & Petrou, S. 2018, "A Systematic Review and Meta-analysis of Childhood Health Utilities", Medical Decision Making, vol. 38, no. 3, pp. 277-305). 15 variables representing a quality measure will be extracted from each of the selected articles that have met the inclusion criteria. A point will be assigned to each of these variables if they are reported in the article. The sum of all the points will indicate the overall reporting quality score of the article.

Strategy for data synthesis

-A descriptive summary of the included studies and the level of convergence/divergence between self-report and proxy-report of HRQoL using preference-based Patient Reported Outcome Measures (PROMs) in the paediatric population will be provided.

-The ICC between self-report and proxy-report of HRQoL using preference-based Patient Reported Outcome Measures (PROMs) in the paediatric population for each domain will be reported.

-A meta-analysis will be used to pool correlation coefficients for convergent validity using the Schmidt–Hunter method. Heterogeneity will be assessed using the I^2 statistic.

Analysis of subgroups or subsets

None planned

Contact details for further information

Diana Khanna

khan0420@flinders.edu.au

Organisational affiliation of the review

Flinders University, South Australia

www.flinders.edu.au

Review team members and their organisational affiliations

Ms Diana Khanna. College of Nursing and Health Sciences, Flinders University, South Australia

Professor Julie Ratcliffe. College of Nursing and Health Sciences, Flinders University, Australia

Dr Christine Mpundu-Kaambwa. College of Nursing and Health Sciences, Flinders University, Australia

Dr Jyoti Khadka. College of Nursing and Health Sciences, Flinders University, Australia

Mr Matthew Crocker. College of Nursing and Health Sciences, Flinders University, Australia

Type and method of review

Systematic review

Anticipated or actual start date

24 May 2021

Anticipated completion date

25 October 2021

Funding sources/sponsors

MRFF - PPHR Initiative - 2019 Targeted Health System and Community Organisation Research Grant Opportunity. Measuring and valuing changes in child health to facilitate robust decision making. (Reference ID: MRF1200816)

Grant number(s)

State the funder, grant or award number and the date of award

Funder: Australian Government, National Medical and Research Council

Grant Number MRF1200816, approved on the 14th of May, 2020

Conflicts of interest

Language

English

Country

Australia

Stage of review

Review Ongoing

Subject index terms status

Subject indexing assigned by CRD

Subject index terms

Child; Health Status; Humans; Patient Reported Outcome Measures; Quality of Life

Date of registration in PROSPERO

30 June 2021

Date of first submission

31 May 2021

Details of any existing review of the same topic by the same authors

This systematic review is an update of an earlier review and will include findings from recent publications (up to 31st May 2021), approach the primary question from a novel aspect and provide a with meta-analysis for pooled correlation co-efficients for convergent validity, absent in the earlier review. Article link: https://doi.org/10.1016/j.socscimed.2019.112543

Stage of review at time of this submission

Stage	Started	Completed
Preliminary searches	Yes	No
Piloting of the study selection process	No	No
Formal screening of search results against eligibility criteria	No	No
Data extraction	No	No
Risk of bias (quality) assessment	No	No
Data analysis	No	No

The record owner confirms that the information they have supplied for this submission is accurate and complete and they understand that deliberate provision of inaccurate information or omission of data may be construed as scientific misconduct.

The record owner confirms that they will update the status of the review when it is completed and will add publication details in due course.

Versions

30 June 2021

30 June 2021

Table 10-2 PRISMA (2020) Checklist.

Section and Topic	Item #	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review.	Cover page
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	1
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	2-4
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	4
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	5
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	4
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Appendix 1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	8-9

Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	5,6,9
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome dimension in each study were sought (e.g., for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	6
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	5, 7
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	8-9
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	7-8
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	N/A
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	7-8
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	8
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	7-8
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	8
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A

Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Appendix 3 and 4
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	7-8
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A
Study characteristics	17	Cite each included study and present its characteristics.	Table 2 and 3
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Table 2 and 3, Appendix 3 and 4
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Table 4 and 5
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	9-10
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	15-16
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	10-15
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A

Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	15-16
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	17-20
	23b	Discuss any limitations of the evidence included in the review.	19-20
	23c	Discuss any limitations of the review processes used.	20
	23d	Discuss implications of the results for practice, policy, and future research.	20
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	4
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	N/A
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	20
Competing interests	26	Declare any competing interests of review authors.	20
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	N/A

Table 10-3 Quality assessment criteria for quantitative studies adapted from Kmet et al., 2004.

No.	Criteria	YES	Partial	No
1	Question / objective sufficiently described?	2	1	0
2	Study design evident and appropriate?	2	1	0
3	Method of subject/comparison group selection or source of information/input variables described and appropriate?	2	1	0
4	Subject (and comparison group, if applicable) characteristics sufficiently described?	2	1	0
5	If interventional and random allocation was possible, was it described?	N/A	N/A	N/A
6	If interventional and blinding of investigators was possible, was it reported?	N/A	N/A	N/A
7	If interventional and blinding of subjects was possible, was it reported?	N/A	N/A	N/A
8	"Outcome and (if applicable) exposure measure(s) well defined and robust to measurement / misclassification bias? Means of assessment reported?"	2	1	0
9	Sample size appropriate?	2	1	0
10	Analytic methods described/justified and appropriate?	2	1	0
11	Some estimate of variance is reported for the main results?	2	1	0
12	Controlled for confounding?	N/A	N/A	N/A
13	Results reported in sufficient detail?	2	1	0
14	Conclusions supported by the results?	2	1	0

Table 10-4 A comparison of quality assessment scores using the two criteria: 1) Kmet et al. 2) Papaioannou et al.

Study number	Study name	Quality score Criteria 1 (Max score: 20)	Score	Quality score Criteria 2 (Max score: 7)	Score
1	Czyzewski et al., 1994	16	0.8	5	0.71

2	Barr et al., 1999	14	0.7	5	0.71
3	Glaser et al., 1999 (a & b)	17	0.85	5	0.71
4	Verrips et al., 2001	17	0.85	5	0.83
5	Brunner et al., 2003	16	0.8	6	0.86
6	Sung et al., 2004	18	0.9	6	0.86
7	Fu et al., 2006	16	0.8	7	1
8	Banks et al., 2008	17	0.85	6	0.86
9	Fluchel et al., 2008	19	0.95	7	1
10	Jelsma and Ramma, 2010	17	0.85	6	0.86
11	Penn et al., 2011	19	0.95	6	0.86
12	Belfort et al., 2016	17	0.85	5	1
13	Lee et al., 2011	17	0.85	7	1
14	Morrow et al., 2012	17	0.85	7	1
15	Rhodes et al., 2012	20	1	7	1
16	Ungar et al., 2012	17	0.85	7	1
17	Kulpeng et al., 2013	17	0.85	6	0.86
18	Wolke et al., 2013	18	0.9	7	1
19	Gusi et al., 2014	19	0.95	6	0.86
20	Sims-Williams et al., 2017	18	0.9	7	1
21	Bharij et al., 2017	18	0.9	7	1
22	Bray et al., 2017	19	0.95	6	0.86
23	Perez Sousa et al., 2017	19	0.95	6	0.86
24	Perez Sousa et al., 2018 (b)	19	0.95	4	0.8
25	van Summeren et al., 2018	19	0.95	5	1
26	Rogers et al., 2019	20	1	7	1
27	Shiroiwa et al., 2019	18	0.9	7	1

28	Sinlapamongkolkul et al., 2020	19	0.95	7	1
29	Lin et al., 2020	19	0.95	6	0.86
30	Zhou et al., 2021	19	0.95	7	1
31	Ralph et al., 2022	18	0.9	6	0.86
32	Abraham et al., 2022	19	0.95	7	1
33	Hetherington et al., 2022	19	0.95	7	1
34	Fitriana et al., 2022	19	0.95	7	1

10.2 Supplementary information for Chapter 3

10.2.1 Interview protocol and materials

10.2.1.1 An interview protocol for the Child

(1) Introduction

Interviewer prompts:

- ‘Hello, my name is ______. How are you? How do you feel?

This is what we would like you to tell us.

(2) The child will be provided with the information on paper and an opportunity to express their views in writing about their participation.

Interviewer prompts:

- I would like you to answer the questions that come up on the screen about your health and how you are feeling today. Some of the questions might be harder for you to answer than others. If you want to stop at any time, just tell me and we will stop.
- Please read every question carefully.
- What answer comes to your mind first?
- Choose the box that fits your answer best and click on it.
- Please remember this is not a test so there are no wrong or right answers. It is important that you try and answer all the questions. If you are having any difficulties and need some help, I am here to help you - please just let me know.

- Keep in mind that your responses will not be shown to anyone, including your parent, without asking you if this is OK first.

(3) Eye tracking mode is enabled. The child is guided through a simple calibration procedure and is asked to focus on specific points on the screen. During this procedure the eye tracker uses an eye-tracking software installed in the laptop to estimate the geometric characteristics of the child's eyes to perform gaze point calculation.

- So, let's get started.
- First you will have to look at the screen and make sure the two dots (*point towards the dots*) are in the box. And the way you can get them in the box, is by adjusting the position of your head. Check it out. *Proceed to help the child to be in the optimal position for eye-tracking.*
- Now that we've got the dots in the box, we can go to the next step.
- In this, a dot will appear on the screen. You will have to follow the dots with your eyes and watch it become bigger.
- *Before accepting the calibration:* Excellent job! Now a set of questions will appear on the screen that you have to answer, like mentioned before.
- Again, remember that there are no right or wrong answers.
- Mum/dad will be sitting there (point towards the parent) and answering a few questions for us.
- So, if something is difficult to understand, please ask me. I am here to help.

- It will be best if you can try to stay still and maintain your positions to answer the questions.
- When you have answered the last question, we will have a chat. Let me know once you are done.
- Before you start answering, let me show you an example. *Proceed to show an example with CHU9D.*
- So, if you are ready, let's so this.

(4) The child self-completes CHU9D, PedsQL and EQ-5D-Y (randomised order for participants).

(5) Interviewer submits the survey and proceeds to process the recording of the tracked activity on Tobii pro lab application. For the following sections, the child will be shown a replay of the recording and will be assessed simultaneously using Retrospective Think Aloud (RTA) technique.

(6) Interviewer commences recording and prompts child to think aloud using the following prompts.

Interviewer prompts:

- Thank you very much! You have done very well.
- Now for the next part of this exercise, I will be asking you a few questions.
- What you're going to say is really important. So, I will also be recording the conversation we will be having to make sure that we don't forget anything. And don't worry. You can

say what you feel. If you want me to stop, you can tell me, and I will stop. OK? (*Start recording*)

- So here we go.

(7) For a single instrument (either CHU9D or EQ-5D-Y) picked using a stratified randomisation method based on the child's chronological age, the interviewer prompts the child for an explanation of why they responded in the way that they did to each individual question (retrospective think aloud).

- You will now see on the screen the survey you just completed.
- All you have to do here is look at your answers from Section A. It looks something like this (*show the child the paper version of the questionnaire*).
- While you are looking at your answers you have to tell me what answer you chose and what the first thought in your mind was when you answered that question.
- *Please note down their answers in the paper version of the self-report for the chosen instrument.*
- *Please encourage the child with phrases like, "Okay, go on" and affirmative nods.*
- *This is the best practice. However, if the child is too shy, use the following technique.*
- *Only resort to verbal probing if the child does not verbalise. You may ask "So, how did you arrive at that answer?", "Was that hard or easy to answer?", "was there anything that confused you?", "What does the term ______ mean to you?", "Why did you choose that specific answer?"*

- You have answered questions from three different section (show the child the sections in the recording).

(8) Interviewer thanks Child for participating and stops the recording.

For interviewers purpose only:

Categories[8] used to rate child's understanding to determine if the child was able to engage with the questionnaire:

Category 1	☐	The child did not complete the task—was too tired
Category 2	☐	The child did not complete the task—but there is NOT enough evidence to assume that he did not understand enough to finish
Category 3	☐	The child did not complete the task—and there is enough evidence to assume that he did not understand enough to finish
Category 4	☐	The child completed the task, but using comments and questions, there is evidence of poor understanding of the task
Category 5	☐	The child completed the task, and using comments and questions, there is evidence of good/excellent understanding of the task

---

[8] Table as published in the paper by Guerriero, C., Jaume, N.A., Diaz-Ordaz, K., Brown, K.L., Wray, J., Ashworth, J., Abbiss, M. and Cairns, J., 2020: Using Animation to Self-Report Health: A Randomized Experiment with Children, *The Patient-Patient-Centered Outcomes Research*, *13*(2), pp.175-188.

Section A

CHU9D

The CHU9D self-complete version was presented here. Now removed due to copyright restrictions.

Section B

The PedsQLTM Version 4.0 Short Form YOUNG CHILD REPORT for ages 6-7 OR CHILD REPORT ages 8-12 version was presented here. Now removed due to copyright restrictions.

Section C

The EQ-5D-Y (Self- complete version for 8-12-year-old) was presented here. Now removed due to copyright restrictions.

Section D

General health

In general, would you say your health is:

☐ Excellent

☐ Very good

☐ Good

☐ Fair

☐ Poor

10.2.1.2 An interview protocol for the Parent/Guardian

Interviewed simultaneously with the child in the same room.

(1) Greetings and introduction.

(2) The parent would be sent a copy of the information sheet, consent form and brief background questions about their child (see *Information and consent sheet* and *Screening questions*) in advance, electronically or via mail.

(3) If there are any concerns regarding the contents of the information sheet, it will be explained. The signed consent form will be collected from the parent. The interview procedure for both, the child, and the parent, will proceed once the parental consent is obtained.

(4) The parent is requested to wear a noise cancelling headphone so that their responses are not unduly influenced by the conversation between the interviewer and the child (Not compulsory; the parent can decline to wear the headphones. In this case, proceed with the interview without the headphones on the parent).

- Before you start answering the questions, would you be okay to wear these headphones?”
- The reason is twofold. Firstly, we want you to complete the survey with minimal distraction. Secondly, we do not want your answers to be influenced by any discussion that might take place between me and your child.

(5) The parent completes a paper and pen survey comprising four main sections:

Section A: CHU9D, PedsQL and EQ-5D-Y (randomised order for participants) using standard proxy format (proxy 1) from their perspective -paper version.

Section B: Brief socio-demographic questions and EQ-5D (for themselves).

Section C: EQ-5D-Y from the child's perspective (proxy 2) - paper version.

(6) Interviewer thanks both parent and child for participating and stops recording signalling end of interview.

Following are the sections A to D that the parent will complete in a paper format.

Section A: Parent *for* child

In this section we will ask you questions about your child's health and wellbeing. Please read carefully as you answer the following questions.

The CHU9D proxy-complete version was presented here. Now removed due to copyright restrictions.

The PedsQL™ Version 4.0 Short Form (Parent REPORT children for ages 8-12) was presented here. Now removed due to copyright restrictions.

The EQ-5D-Y Proxy version 1 was presented here. Now removed due to copyright restrictions.

Section B: About you

We would like to ask some questions about your health, general demographic characteristics, and your opinion about your child.

The EQ-5D-3L Self-complete version for adults 1 was presented here. Now removed due to copyright restrictions.

1. In general, would you say your health is:

☐ Excellent

☐ Very good

☐ Good

☐ Fair

☐ Poor

2. Are you:

☐ Female

☐ Male

☐ Transgender Female

☐ Transgender Male

☐ Other

3. How old are you? ___________

4. What is the postcode of the area in which you live? ___________

Section C: Parent *as* child

In this final section of the survey, we will ask you to **put yourself in your child's shoes** and answer similar questions, as above, about your child's health and wellbeing. Please read carefully as you answer the following questions.

The EQ-5D-Y Proxy version 2 was presented here. Now removed due to copyright restrictions.

General health

Please tick the **ONE** box that **you think your *child* would choose** to describe his or her health state **TODAY**.

In general, my health is:

☐ Excellent

☐ Very good

☐ Good

☐ Fair

☐ Poor

Comments

If you have any comments that you would like to make about this questionnaire, please write these in the free text box below.

Thank you very much for taking the time to complete this survey!

10.2.1.3 Additional information:

FAQs

Responses to give children if they ask about the following:

1. Explain the meaning of words:

- Once the child has begun the survey and the eye tracking has begun, minimise any interaction with the child.

- It is suggested that **if the respondent asks for clarification, the interviewer can help by re-reading the question verbatim. The interviewer should not try to offer his or her own explanation but suggest that the respondent uses his or her own interpretation.**

- If a child does ask you a question to clarify the meaning of a word in one of the questions, use the definitions (see appendix 2) as a guide.

- If the child still does not understand the meaning of the word even after you have tried to explain it, suggest they skip the question and move onto the next question.

2. Child wants to move:

- Reiterate, politely, that it would be best if he/she doesn't move. Ask if they need anything that you can help with. If the child still wants to move, let the parent know and see if the child can come back and finish the interview. If not, end the interview.

3. Child does not look at the screen:

- Please gently bring back the child's attention to the screen. You can ask the child if he/she is finding any difficulty and offer help.

4. Child is distressed:

- Ask the child if you can help with something. If the child is unable to finish the interview due to distress, please end the interview.

10.2.2 Information sheet and consent form

School of Nursing and Health Sciences

GPO Box 2100

Adelaide SA 5001

Tel: 08 8275 2858

Fax: 08 8275 1130

julie.ratcliffe@flinders.edu.au

http://www.flinders.edu.au/clinical-change/

CRICOS Provider No. 00114A

Important Information

Assessing the validity of self vs proxy assessment of Health-Related Quality of Life in children - a mixed-methods study

Dear Parents and children,

My name is Julie Ratcliffe, and I am a health economist based at Flinders University. I am inviting all parents and children aged 6 to 12 years to participate in a research study. Participation in this study is entirely voluntary.

Purpose of the survey

The purpose of this study is to find out more about children's health-related quality of life from the child's and the parent/guardian's perspective. We want to investigate children's abilities to self-assess their health-related quality of life using questionnaires designed for this purpose. In addition, this research will assess any differences between self (child) and proxy (parent/guardian) reporting of the child's health-related quality of life.

If you and your child agree to participate, you both will take part in a face-to-face interview which will take place in a single meeting room at a central location. Your child will be asked to [1] complete a survey including three brief validated instruments for the measurement of health-related quality of life in children presented in a digital format and [2] to explain their responses for one of the instruments to the interviewer. This conversation will be recorded on a voice recorder. The survey will be presented in a digital format and completed on a laptop computer attached with an eye-tracking equipment. We will use the eye tracking data to track your child's eye movements as they respond to the survey questions and gather data on how children read and respond to the survey questions.

You will be asked to complete a similar survey in hard copy (paper and pen), including the same three brief validated instruments used for the measuring of health-related quality of life in children. The survey will also include questions about your views of your child's current health-related quality of life and some questions about your own quality of life. In addition, you will be asked to wear noise cancelling headphones whilst completing the survey so that you will not hear your child's responses, as both you and your child will be interviewed in the same room.

In the final section of the interview, if your responses and your child's are different, the interviewer will invite both you and your child to discuss them. This part of the interview will only take place if your child agrees to share their responses with you. After you and your child discuss your answers, the interviewer will allow you to alter your responses to any of the questions.

We anticipate that the survey will take between 45 and 60 minutes to complete. The survey will be administered at a convenient venue and time as agreed in advance with yourself and your child. All interviews will be conducted according to current COVID-19 guidelines, including observing stipulated government social distancing guidelines and wiping and disinfecting of all equipment used in this study between each interview.

Potential benefits of the survey

The study is being funded by the Medical Research Future Fund (MRFF). The study is a South Australia wide study that will substantially improve the evidence base on good practices for the collection of health-related quality of life data from children. It will also provide people working in the area of child health care services and health policymakers with important information about the health-related quality of life of Australian children from their and their parent's perspective.

Survey Procedures

The survey involves your child answering a set of questions about their health and their preferences for a series of health states. Some example questions from the survey are as follows:

A. For the child

Pain:

☐ I don't have any pain today

☐ I have a little bit of pain today

☐ I have a bit of pain today

☐ I have quite a lot of pain today

[1] Please click next to the statement which is most like you today:

	PHYSICAL FUNCTIONING (problems with…)	Not at all	Some- times	A lot
1	Is it hard for you to walk more than one block	🙂	😐	😐

[2] Please click next to the statement which is most like you today:

OR

Please click next to the statement which is most like you today:

	ABOUT MY HEALTH AND ACTIVITIES	Never	Almost	Some-	Often	Almost
1	It is hard for me to walk more than one	0	1	2	3	4

[3] Please click next to the statement which is most like you today:

	MOBILITY (walking about)	
1	I have no problems walking about	☐
2	I have some problems walking about	

		☐
3	I have a lot of problems walking about	☐

B. For the parent

[1] Please tick next to the statement which is most like your child today:

Worried

☐ My child doesn't feel worried today

☐ My child feels a little bit worried today

☐ My child feels a bit worried today

☐ My child feels quite worried today

☐ My child feels very worried today

[2] Please tick next to the statement which is most like your child today:

	PHYSICAL FUNCTIONING	Never	Almost	Some-	Often	Almost
	Walking more than one block	0	1	2	3	4

[3] Please tick the one box that you think best describes your child's health today:

MOBILITY (walking about)

☐ He/she has no problems with walking about

☐ He/she has some problems with walking about

☐ He/she has a lot of problems with walking about

Risks or adverse effects

This survey includes questions about how you and your child are feeling. If you or your child feel particularly worried, sad or anxious following completion of this survey, help is available to you from several free counselling services including:

Kids Helpline	**Lifeline**	**Beyond Blue**
Phone - 1800 55 1800	Phone - 13 11 14	Phone – 1300 22 4636
Email -	Website –	Website -

Research Statement

The researchers in this study will gain no direct financial benefit from this survey.

Participation and withdrawal from the study

You and your child's participation in this study is entirely voluntary. The study data is confidential and anonymous. You and your child may withdraw from participation in this research study at any time without any penalty. As a thank you for your time and participation in this research, you will be provided with an e-gift token of $90.

Confidentiality

You and your child will not be identifiable by name in any publication arising from the results of the survey.

Data Storage

The information collected will be stored securely on the Flinders University network throughout the study and will only be accessible via password protected computers. Any identifiable data will be de-identified for data storage purposes. All data will be securely transferred to and stored at Flinders University for at least five years after publication of the results. Following the required data storage period, all data will be securely destroyed according to university protocols.

Additional information

Should you require any further details about this survey at any time, you may contact Ms Diana Khanna on 04 5259 1726 or Dr Christine Mpundu-Kaambwa (08) 8201 2418 from Caring Futures Institute at Flinders University.

Yours Sincerely

Julie Ratcliffe.

Professor Julie Ratcliffe, Caring Futures Institute, Flinders University.

Ms Diana Khanna, PhD student, Caring Futures Institute, Flinders University

Dr Christine Mpundu-Kaambwa, Research Fellow, Caring Futures Institute, Flinders University

This research study has been approved by Flinders University's Human Research Ethics Committee (Project ID 4178). For more information regarding ethical approval of the project the Executive Officer of the Committee can be contacted by telephone (08) 8201 2543 or by email human.researchethics@flinders.edu.au

Informed consent for parents

Flinders Thesis template

1. Details of the interview process and any possible inconveniences and/or risks that are outlined in the attached information sheet have been explained to my satisfaction.

2. I understand that:

   - My child and I may not directly benefit by taking part in this research.
   - My child and I are free to decline to answer particular questions.
   - I understand that there will be no payment to me or my child for taking part in this study.
   - I understand that an audio recording will be made of the session with me and my child. These recordings will be kept confidential. These recordings will be stored securely and only researchers will have access to the recordings. These recordings will be destroyed according to university protocols.
   - I understand that my child and I may withdraw from participation in this research study at any time.
   - While the information gained in this study will be published as explained, my child and I will not be identified, and individual information will remain confidential except where there is a requirement by law for it to be divulged.

I confirm that I have read and understood the information provided.

I, ____, hereby consent to my and my child's participation, as requested, in the research project on assessing validity of self vs proxy assessment of Health-Related Quality of Life in children.

Signed (parent or guardian's name) ______________________________Date ________

10.3 Supplementary information for Chapters 6 and 7

10.3.1 Sample characteristics for the overall qualitative sample

Think aloud interviews were conducted with all 85 child participants. However, data from only 75 children were available for analysis, as data from 10 children (12%) were lost due to a technical error[9]. Of these participants, 39 (52%) completed the EQ-5D-Y-3L interviews while 36 (48%) participated in the CHU9D interviews (discussed in Chapter 7). Table 6.1 presents the socio-demographic characteristics of the overall participant sample included in the qualitative phase. The median age of children in this sample was 9 years (SD=4) and the gender distribution was relatively balanced, with 43% (n=32) being boys and 57% (n=42) girls. Regarding long-term health conditions, 31% (n=23) of the children were reported to have a long-term condition, while 69% (n=52) did not. Among the specific health conditions[6] above reported, the distribution was as follows: 9% (n=2) had attention deficit hyperactivity disorder, 9% (n=2) had autism spectrum disorder, 4% (n=1) had congenital heart disease, 48% (n=11) had asthma, 17% (n=4) had dental caries, and 13% (n=3) reported sleep problems. In terms of general health on the SRH item, 23% (n=17) of the children rated their health as excellent, 52% (n=38) as very good, 21% (n=15) as good, and 4% (n=3) as fair.

---

[9] Occurred due to the technical failure of the recording device used by the interviewers.

Table 10-5 Description of participant characteristics for the overall qualitative sample.

Variable	N	%
Child age		
Mean (SD)	75	9 (2)
Median (IQR)		9 (4)
Child gender		
Boy	32	43
Girl	42	57
Long term condition		
No	52	69
Yes	23	31
Child Health condition		
Attention deficit hyperactivity disorder	2	9
Autism spectrum disorder	2	9
Congenital heart disorder	1	4
Asthma	11	48
Dental caries	4	17
Sleep problems	3	13
Child self-rated general health		
Excellent	17	23
Very good	38	52
Good	15	21
Fair	3	4
Socio-economic condition according to post-code		
Relatively advantaged quintile (SEIFA decile 7,8,9,10)	23	32
Relatively disadvantaged quintile (SEIFA decile 1,2,3,4,5,6)	50	68

SEIFA= Socio-Economic Indexes for Australia

BIBLIOGRAPHY

1. Australian Institute of Family Studies. Australian Legal Definitions: When Is a Child in Need of Protection? ; 2023. Available from: https://aifs.gov.au/resources/resource-sheets/australian-legal-definitions-when-child-need-protection.
2. Brazier J., Ratcliffe J., Saloman J., Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation: Oxford University Press; 2016.
3. Manipis K., Viney, R., De Abreu Lourenço, R., Ng, C., Yu, A., Meshcheriakova, E., de Feria Cardet,, R. C., J., Akanksha, A., Arora, S., Addo, R., Thomas, J., Cronin, P., Vigours, S., Vargas, Parada C., Norris, S., Kim, H., & Goodall, S. Health Technology Assessment methods: Economic Evaluation. Australian Health Technology Assessment Methods and Policy Review. Canberra: Australian Government, Department of Health and Aged Care; 2023.
4. Drummond M. F., Sculpher M. J., Claxton K., Stoddart G. L., Torrance G. W. Methods for the Economic Evaluation of Health Care Programmes: Oxford University Press; 2015.
5. Australian Government Department of Health. Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee (Version 5.0). Department of Health; 2016. Available from: http://www.pbac.pbs.gov.au/home.html.
6. Gold M. R., Siegel J. E., Russell L. B., Weinstein M. C., Weinstein M. C. Cost-Effectiveness in Health and Medicine. Cary, United States: Oxford University Press, Incorporated; 1996.
7. Byford S., Raftery J. Perspectives in Economic Evaluation. BMJ, 1998;316(7143):1529-30. https://doi.org/10.1136/bmj.316.7143.1529.
8. Kim D. D., Silver M. C., Kunst N., Cohen J. T., Ollendorf D. A., Neumann P. J. Perspective and Costing in Cost-Effectiveness Analysis, 1974-2018. PharmacoEconomics, 2020;38(10):1135-1145. https://doi.org/10.1007/s40273-020-00942-2.
9. Bertram M. Y., Stenberg K., Brindley C., Li J., Serje J., Watts R., et al. Disease Control Programme Support Costs: An Update of WHO-CHOICE Methodology, Price Databases and Quantity Assumptions. Cost Effectiveness and Resource Allocation, 2017;15(1):21. https://doi.org/10.1186/s12962-017-0083-6.
10. Avşar T. S., Yang X., Lorgelly P. How is the Societal Perspective Defined in Health Technology Assessment? Guidelines from Around the Globe. PharmacoEconomics, 2023;41(2):123-138. https://doi.org/10.1007/s40273-022-01221-y.
11. Stawasz A., Huang L., Kirby P., Bloom D. Health Technology Assessment for Vaccines Against Rare, Severe Infections: Properly Accounting for Serogroup B Meningococcal Vaccination's Full Social and Economic Benefits. Frontiers in public health, 2020;8:261. https://doi.org/10.3389/fpubh.2020.00261.
12. Culyer A., Chalkidou K., Teerawattananon Y., Santatiwongchai B. Rival Perspectives in Health Technology Assessment and Other Economic Evaluations for Investing in Global and National Health: Who Decides? Who Pays? . F1000Research, 2018;7(72). https://10.12688/f1000research.13284.1.
13. Torrance G. Preferences for Health Outcomes and Cost-Utility Analysis. The American Journal of Managed Care, 1997;3 Suppl:S8-20.
14. Brent R. J. Cost-Benefit Analysis versus Cost-Effectiveness Analysis from a Societal Perspective in Healthcare. International Journal of Environmental Research and Public Health, 2023;20(5). https://doi.org/10.3390/ijerph20054637.
15. Coast J. Is Economic Evaluation in Touch With Society's Health Values? BMJ, 2004;329(7476):1233-6. https://doi.org/10.1136/bmj.329.7476.1233.
16. Dorenkamp M., Boldt J., Leber A. W., Sohns C., Roser M., Boldt L. H., et al. Cost-Effectiveness of Paclitaxel-Coated Balloon Angioplasty in Patients With Drug-Eluting Stent Restenosis. Clinical Cardiology, 2013;36(7):407-13. https://doi.org/10.1002/clc.22130.
17. Kluger S., Obermann K., Tausch V., Chadasch C., Ditscheid E., Thielscher C. Is QALY-Based Rationing Illegal in Countries With a Natural-Law Constitution? A Multidisciplinary Systematic Review. Ethics, Medicine, and Public Health, 2020;14:100484. https://doi.org/10.1016/j.jemep.2020.100484.

18. York Health Economics Consortium. *Incremental Cost-Effectiveness Ratio (ICER) [online]*. 2016.
19. Tsuchiya A., Dolan P. The QALY Model and Individual Preferences for Health States and Health Profiles Over Time: A Systematic Review of the Literature. Medical Decision Making, 2005;25(4):460-7. https://doi.org/10.1177/0272989x05276854.
20. Bush J. W. Health-Related Quality of Life Measurement for Evaluation Research and Policy Analysis. Health Psychology, 1982;1(1):61-80. https://doi.org/10.1037/0278-6133.1.1.61.
21. Brooks R. G. Health Status Measurement: A Perspective on Change: Springer; 1994.
22. Torrance G. W., Thomas W. H., Sackett D. L. A Utility Maximization Model for Evaluation of Health Care Programs. Health Services Insights, 1972;7(2):118.
23. Loomes G., McKenzie L. The Use of QALYs in Health Care Decision Making. Social science & medicine, 1989;28(4):299-308. https://doi.org/10.1016/0277-9536(89)90030-0.
24. Weinstein M. C., Stason W. B. Foundations of Cost-Effectiveness Analysis for Health and Medical Practices. New England Journal of Medicine, 1977;296(13):716-721.
25. Zeckhauser R., Shepard D. Where Now for Saving Lives? Law and Contemporary Problems, 1976;40:5. https://doi.org/10.2307/1191310.
26. Patrick D. L., Erickson P. Health Status and Health Policy: Quality of Life in Health Care Evaluation and Resource Allocation. New York: Oxford University Press; 1993.
27. World Health Organization. *Constitution of the World Health Organization*. 1995 Available from: https://www.who.int/about/governance/constitution.
28. Australian Institute of Health and Welfare. Chronic Condition Multimorbidity. Canberra: Australian Government; 2021. Available from: https://www.aihw.gov.au/reports/chronic-disease/chronic-condition-multimorbidity/contents/chronic-conditions-and-multimorbidity.
29. Prutkin J. M., Feinstein A. R. Quality-of-Life Measurements: Origin and Pathogenesis. The Yale Journal of Biology and Medicine, 2002;75(2):79-93.
30. Spitzer W. O. State of Science 1986: Quality of Life and Functional Status as Target Variables for Research. Journal of Chronic Diseases, 1987;40(6):465-471. https://doi.org/10.1016/0021-9681(87)90002-6.
31. Elkinton J. R. Medicine and the Quality of Life. Annals of Internal Medicine, 1966;64(3):711-4. https://doi.org/10.7326/0003-4819-64-3-711.
32. Romero M., Vivas-Consuelo D., Alvis-Guzman N. Is Health-Related Quality of Life (HRQoL) a Valid Indicator for Health Systems Evaluation? SpringerPlus, 2013;2(1):664-664. https://doi.org/10.1186/2193-1801-2-664.
33. Karimi M., Brazier J. Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? PharmacoEconomics, 2016;34(7):645-649. https://doi.org/10.1007/s40273-016-0389-9.
34. Varni J. W., Burwinkle T. M., Lane M. M. Health-Related Quality of Life Measurement in Pediatric Clinical Practice: An Appraisal and Precept for Future Research and Application. Health and Quality of Life Outcomes, 2005;3(1):34. https://doi.org/10.1186/1477-7525-3-34.
35. Brazier J., Ara R., Rowen D., Chevrou-Severac H. A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. PharmacoEconomics, 2017;35(1):21-31. https://doi.org/10.1007/s40273-017-0545-x.
36. Wille N., Badia X., Bonsel G., Burström K., Cavrini G., Devlin N., et al. Development of the EQ-5D-Y: A Child-Friendly Version of the EQ-5D. Quality of Life Research, 2010;19(6):875-86. https://doi.org/10.1007/s11136-010-9648-y.
37. Feeny D., Furlong W., Torrance G. W., Goldsmith C. H., Zhu Z., Depauw S., et al. Multiattribute and Single-Attribute Utility Functions for the Health Utilities Index Mark 3 System. Medical Care, 2002;40(2):113–128. https://doi.org/10.1097/00005650-200202000-00006.
38. Torrance G. W., Feeny D. H., Furlong W. J., Barr R. D., Zhang Y., Wang Q. Multiattribute Utility Function for a Comprehensive Health Status Classification System: Health Utilities Index Mark 2. Medical Care, 1996;34(7):702-22. https://doi.org/10.1097/00005650-199607000-00004.
39. Torrance G. W. Utility Approach to Measuring Health-Related Quality of Life. Journal of Chronic Diseases, 1987;40(6):593-600. https://doi.org/10.1016/0021-9681(87)90019-1.
40. Feeny D. A Utility Approach to the Assessment of Health-Related Quality of Life. Medical Care, 2000;38(9). https://doi.org/10.1097/00005650-200009002-00022.

41. Gafni A. The Standard Gamble Method: What Is Being Measured and How It Is Interpreted. Health Services Insights, 1994;29(2):207-24.
42. Gudex C. Time Trade-Off User Manual: Props and Self-Completion Methods. Centre for Health Economics, University of York1994.
43. Luce R. D. Individual Choice Behavior: A Theoretical Analysis: Courier Corporation; 2005.
44. McFadden D. Conditional Logit Analysis of Qualitative Choice Behavior. Frontiers in Econometrics, 1973:105-42.
45. Matza L. S., Secnik K., Rentz A. M., Mannix S., Sallee F. R., Gilbert D., et al. Assessment of Health State Utilities for Attention-Deficit/Hyperactivity Disorder in Children Using Parent Proxy Report. Quality of Life Research, 2005;14(3):735-47. https://doi.org/10.1007/pl00022070.
46. Torrance G. W. Social Preferences for Health States: An Empirical Evaluation of Three Measurement Techniques. Socio-Economic Planning Sciences, 1976;10(3):129-136. https://doi.org/10.1016/0038-0121(76)90036-7.
47. Janssen B. M., Oppe M., Versteegh M. M., Stolk E. A. Introducing the Composite Time Trade-Off: A Test of Feasibility and Face Validity. The European Journal of Health Economics, 2013;14 Suppl 1(Suppl 1):S5-13. https://doi.org/10.1007/s10198-013-0503-2.
48. Stolk E., Ludwig K., Rand K., van Hout B., Ramos-Goñi J. M. Overview, Update, and Lessons Learned From the International EQ-5D-5L Valuation Work: Version 2 of the EQ-5D-5L Valuation Protocol. Value in Health, 2019;22(1):23-30. https://doi.org/10.1016/j.jval.2018.05.010.
49. Augustovski F., Rey-Ares L., Irazola V., Oppe M., Devlin N. Lead Versus Lag Time Trade-Off Variants: Does It Make Any Difference? Value in Health, 2013;16(3):A36. https://doi.org/10.1016/j.jval.2013.03.205.
50. Attema A. E., Versteegh M. M., Oppe M., Brouwer W. B. F., Stolk E. A. Lead Time TTO: Leading to Better Health State Valuations? Health Economics, 2013;22(4):376-392. https://doi.org/10.1002/hec.2804.
51. Parkin D., Devlin N. Is There a Case for Using Visual Analogue Scale Valuations in Cost-Utility Analysis? Health Economics, 2006;15(7):653-664. https://doi.org/10.1002/hec.1086.
52. Dolan P., Sutton M. Mapping Visual Analogue Scale Health State Valuations Onto Standard Gamble and Time Trade-Off Values. Social Science & Medicine, 1997;44(10):1519-1530. https://doi.org/10.1016/S0277-9536(96)00271-7.
53. Shields B. J., Palermo T. M., Powers J. D., Grewe S. D., Smith G. A. Predictors of a Child's Ability to Use a Visual Analogue Scale. Child: Care, Health and Development, 2003;29(4):281-290. https://doi.org/10.1046/j.1365-2214.2003.00343.x.
54. Rowen D., Brazier J., Van Hout B. A Comparison of Methods for Converting DCE Values onto the Full Health-Dead QALY Scale. Medical Decision Making, 2015;35(3):328-40. https://doi.org/10.1177/0272989x14559542.
55. Salomon J. A. Reconsidering the Use of Rankings in the Valuation of Health States: A Model for Estimating Cardinal Values from Ordinal Data. Population Health Metrics, 2003;1(1):12. https://doi.org/10.1186/1478-7954-1-12.
56. 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. Social Science & Medicine, 2016;157:48-59. https://doi.org/10.1016/j.socscimed.2016.03.042.
57. Bleichrodt H., Abellan-Perpiñan J. M., Pinto-Prades J. L., Mendez-Martinez I. Resolving Inconsistencies in Utility Measurement Under Risk: Tests of Generalizations of Expected Utility. Management Science, 2007;53(3):469-482. https://doi.org/10.1287/mnsc.1060.0647.
58. Schwappach D. L. B. Resource Allocation, Social Values, and the QALY: A Review of the Debate and Empirical Evidence. Health Expectations, 2002;5(3):210-222. https://doi.org/10.1046/j.1369-6513.2002.00182.x.
59. Dolan P., Kahneman D. Interpretations Of Utility And Their Implications For The Valuation Of Health. The Economic journal (London), 2008;118(525):215-234. 10.1111/j.1468-0297.2007.02110.x.
60. Lipman S. A., Brouwer W. B. F., Attema A. E. QALYs Without Bias? Nonparametric Correction of Time Trade-Off and Standard Gamble Weights Based on Prospect Theory. Health Economics, 2019;28(7):843-854. https://doi.org/10.1002/hec.3895.

61. Bansback N., Brazier J., Tsuchiya A., Anis A. Using a Discrete Choice Experiment to Estimate Health State Utility Values. Journal of Health Economics, 2012;31(1):306-18. https://doi.org/10.1016/j.jhealeco.2011.11.004.
62. Wang X., Zhuo L., Ma Y., Cai T., Must A., Xu L., et al. Similar Responses to EQ-5D-3L by Two Elicitation Methods: Visual Analogue Scale and Time Trade-Off. BMC Medical Research Methodology, 2020;20(1):118. https://doi.org/10.1186/s12874-020-01008-9.
63. Happich M., Moock J., von Lengerke T. Health State Valuation Methods and Reference Points: The Case of Tinnitus. Value in Health, 2009;12(1):88-95. https://doi.org/10.1111/j.1524-4733.2008.00397.x.
64. Coast J. Maximisation in Extra-Welfarism: A Critique of the Current Position in Health Economics. Social Science & Medicine, 2009;69(5):786-92. https://doi.org/10.1016/j.socscimed.2009.06.026.
65. Helgesson G., Ernstsson O., Åström M., Burström K. Whom Should We Ask? A Systematic Literature Review of the Arguments Regarding the Most Accurate Source of Information for Valuation of Health States. Quality of Life Research, 2020;29(6):1465-1482. https://doi.org/10.1007/s11136-020-02426-4.
66. Wittenberg E., Halpern E., Divi N., Prosser L. A., Araki S. S., Weeks J. C. The Effect of Age, Race and Gender on Preference Scores for Hypothetical Health States. Quality of Life Research, 2006;15(4):645-653.
67. Norman R., Cronin P., Viney R., King M., Street D., Ratcliffe J. International Comparisons in Valuing EQ-5D Health States: A Review and Analysis. Value in Health, 2009;12(8):1194-200. https://doi.org/10.1111/j.1524-4733.2009.00581.x.
68. Roudijk B., Donders A. R. T., Stalmeier P. F. M. Cultural Values: Can They Explain Differences in Health Utilities between Countries? Medical Decision Making, 2019;39(5):605-616. https://doi.org/10.1177/0272989x19841587.
69. Dewilde S., Janssen M. F., Lloyd A. J., Shah K. Exploration of the Reasons Why Health State Valuation Differs for Children Compared With Adults: A Mixed Methods Approach. Value in Health, 2022;25(7):1185-1195. https://doi.org/10.1016/j.jval.2021.11.1377.
70. Devlin N. J., Pan T., Sculpher M., Jit M., Stolk E., Rowen D., et al. Using Age-Specific Values for Pediatric HRQoL in Cost-Effectiveness Analysis: Is There a Problem to Be Solved? If So, How? PharmacoEconomics, 2023;41(10):1165-1174. https://doi.org/10.1007/s40273-023-01300-8.
71. Yang Y., Brazier J. E., Tsuchiya A., Young T. A. Estimating a Preference-Based Index for a 5-Dimensional Health State Classification for Asthma Derived from the Asthma Quality of Life Questionnaire. Medical Decision Making, 2010;31(2):281-291. https://doi.org/10.1177/0272989X10379646.
72. King M. T., Costa D. S., Aaronson N. K., Brazier J. E., Cella D. F., Fayers P. M., et al. QLU-C10D: A Health State Classification System for a Multi-Attribute Utility Measure Based on the EORTC QLQ-C30. Quality of Life Research, 2016;25(3):625-36. https://doi.org/10.1007/s11136-015-1217-y.
73. Chiou C. F., Weaver M. R., Bell M. A., Lee T. A., Krieger J. W. Development of the Multi-Attribute Pediatric Asthma Health Outcome Measure (PAHOM). International Journal for Quality in Health Care, 2005;17(1):23-30. https://doi.org/10.1093/intqhc/mzh086.
74. Mokkink L. B., Terwee C. B., Patrick D. L., Alonso J., Stratford P. W., Knol D. L., et al. The COSMIN Study Reached International Consensus on Taxonomy, Terminology, and Definitions of Measurement Properties for Health-Related Patient-Reported Outcomes. Journal of Clinical Epidemiology, 2010;63(7):737-45. https://doi.org/10.1016/j.jclinepi.2010.02.006.
75. 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.
76. Connolly M. A., Johnson J. A. Measuring Quality of Life in Paediatric Patients. PharmacoEconomics, 1999;16(6):605-625. https://doi.org/10.2165/00019053-199916060-00002.

77. 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.
78. Devlin N., Pan T., Kreimeier S., Verstraete J., Stolk E., Rand K., et al. Valuing EQ-5D-Y: The Current State of Play. Health and Quality of Life Outcomes, 2022;20:1-11. https://doi.org/10.1186/s12955-022-01998-8.
79. Mott D. J., Shah K. K., Ramos-Goñi J. M., Devlin N. J., Rivero-Arias O. Valuing EQ-5D-Y-3L Health States Using a Discrete Choice Experiment: Do Adult and Adolescent Preferences Differ? Medical Decision Making, 2021;41(5):584-596. https://doi.org/10.1177/0272989x21999607.
80. Reckers-Droog V., Karimi M., Lipman S., Verstraete J. Why Do Adults Value EQ-5D-Y-3L Health States Differently for Themselves Than for Children and Adolescents: A Think-Aloud Study. Value in Health, 2022;25(7):1174-1184. https://doi.org/10.1016/j.jval.2021.12.014.
81. Lipman S. A., Reckers-Droog V. T., Karimi M., Jakubczyk M., Attema A. E. Self vs. Other, Child vs. Adult: An Experimental Comparison of Valuation Perspectives for Valuation of EQ-5D-Y-3L Health States. The European Journal of Health Economics, 2021;22(9):1507-1518. https://doi.org/10.1007/s10198-021-01377-y.
82. Sieber W., Groessl E., Ph D., Kistler K., Ganiats T., Kaplan R. Quality of Well-Being Self-Administered (QWB-SA) Scale. 2008.
83. Apajasalo M., Sintonen H., Holmberg C., Sinkkonen J., Aalberg V., Pihko H., et al. Quality of Life in Early Adolescence: A Sixteen-Dimensional Health-Related Measure (16D). Quality of Life Research, 1996;5(2):205-11. https://doi.org/10.1007/bf00434742.
84. Apajasalo M., Rautonen J., Holmberg C., Sinkkonen J., Aalberg V., Pihko H., et al. Quality of Life in Pre-Adolescence: a 17-Dimensional Health-Related Measure (17D). Quality of Life Research, 1996;5(6):532-8. https://doi.org/10.1007/bf00439227.
85. Moodie M., Richardson J., Rankin B., Iezzi A., Sinha K. Predicting Time Trade-Off Health State Valuations of Adolescents in Four Pacific Countries Using the Assessment of Quality-of-Life (AQoL-6D) Instrument. Value in Health, 2010;13(8):1014-27. https://doi.org/10.1111/j.1524-4733.2010.00780.x.
86. 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. Quality of Life Research, 2010;19(6):887-897. https://doi.org/10.1007/s11136-010-9649-x.
87. Kreimeier S., Åström M., Burström K., Egmar A. C., Gusi N., Herdman M., et al. EQ-5D-Y-5L: Developing a Revised EQ-5D-Y With Increased Response Categories. Quality of Life Research, 2019;28(7):1951-1961. https://doi.org/10.1007/s11136-019-02115-x.
88. Kreimeier S., Oppe M., Ramos-Goñi J. M., 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 in Health, 2018;21(11):1291-1298. https://doi.org/10.1016/j.jval.2018.05.002.
89. Ramos-Goñi J. M., Oppe M., Stolk E., Shah K., Kreimeier S., Rivero-Arias O., et al. International Valuation Protocol for the EQ-5D-Y-3L. PharmacoEconomics, 2020;38(7):653-663. https://doi.org/10.1007/s40273-020-00909-3.
90. Stevens K. Developing a Descriptive System for a New Preference-Based Measure of Health-Related Quality of Life for Children. Quality of Life Research, 2009;18:1105-13. https://doi.org/10.1007/s11136-009-9524-9.
91. 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 in Health, 2012;15(8):1092-1099. https://doi.org/10.1016/j.jval.2012.07.011.
92. Beusterien K. M., Yeung J. E., Pang F., Brazier J. Development of the Multi-attribute Adolescent Health Utility Measure (AHUM). Health and Quality of Life Outcomes, 2012;10:102. https://doi.org/10.1186/1477-7525-10-102.
93. Weldring T., Smith S. M. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Services Insights, 2013;6:61-8. https://doi.org/10.4137/hsi.S11093.

94. Germain N., Aballéa S., Toumi M. Measuring the Health-Related Quality of Life in Young Children: How Far Have We Come? Journal of Market Access & Health Policy, 2019;7(1):1618661. https://doi.org/10.1080/20016689.2019.1618661.
95. Devlin N L. R., Rowen D. Challenges in Measuring and Valuing Children’s Health-Related Quality of Life. Value & Outcomes Spotlight: ISPOR; 2021. p. 13-14.
96. Bahrampour M., Jones R., Dalziel K., Devlin N., Mulhern B., the Q. T. Comparing Generic Paediatric Health-Related Quality-of-Life Instruments: A Dimensionality Assessment Using Factor Analysis. PharmacoEconomics, 2024;42(1):81-94. 10.1007/s40273-024-01382-y.
97. Petrou S. Methodological Challenges Surrounding QALY Estimation for Paediatric Economic Evaluation. Cost Effectiveness and Resource Allocation, 2022;20(1):10. https://doi.org/10.1186/s12962-022-00345-4.
98. Matza L. S., Patrick D. L., Riley A. W., Alexander J. J., Rajmil L., Pleil A. M., 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 in Health, 2013;16(4):461-479. https://doi.org/10.1016/j.jval.2013.04.004.
99. Bevans K. B., Ahuvia I. L., Hallock T. M., Mendonca R., Roth S., Forrest C. B., et al. Investigating Child Self-Report Capacity: A Systematic Review and Utility Analysis. Quality of Life Research, 2020;29(5):1147-1158. https://doi.org/10.1007/s11136-019-02387-3.
100. White-Koning M., Arnaud C., Dickinson H. O., 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.
101. Varni J. W., Limbers C. A., Burwinkle T. M. 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 and Quality of Life Outcomes, 2007;5:2. https://doi.org/10.1186/1477-7525-5-2.
102. Ravens-Sieberer U., Erhart M., Wille N., Wetzel R., Nickel J., Bullinger M. Generic Health-Related Quality-of-Life Assessment in Children and Adolescents: Methodological Considerations. PharmacoEconomics, 2006;24(12):1199-220. https://doi.org/10.2165/00019053-200624120-00005.
103. Riley A. W. Evidence That School-Age Children Can Self-Report on Their Health. Ambulatory Pediatrics, 2004;4(4):371-376. https://doi.org/10.1367/A03-178R.1.
104. Kwon J., Kim S. W., Ungar W. J., Tsiplova K., Madan J., Petrou S. A Systematic Review and Meta-analysis of Childhood Health Utilities. Medical Decision Making, 2017;38(3):277-305. https://doi.org/10.1177/0272989X17732990.
105. Roddenberry A., Renk K. Quality of Life in Pediatric Cancer Patients: The Relationships Among Parents' Characteristics, Children's Characteristics, and Informant Concordance. Journal of Child and Family Studies, 2008;17(3):402-426. https://doi.org/10.1007/s10826-007-9155-0.
106. Eiser C., Varni J. W. Health-Related Quality of Life and Symptom Reporting: Similarities and Differences Between Children and Their Parents. European Journal of Pediatrics, 2013;172(10):1299-1304. https://doi.org/10.1007/s00431-013-2049-9.
107. Otero S., Eiser C., Wright N., Butler G. Implications of Parent and Child Quality of Life Assessments for Decisions About Growth Hormone Treatment in Eligible Children. Child: Care, Health and Development, 2013;39(6):782-788. https://doi.org/10.1111/cch.12021.
108. Griebsch I., Coast J., Brown J. Quality-Adjusted Life-Years Lack Quality in Pediatric Care: A Critical Review of Published Cost-Utility Studies in Child Health. Pediatrics, 2005;115(5):e600-14. https://doi.org/10.1542/peds.2004-2127.
109. Eiser C., Morse R. The Measurement of Quality of Life in Children: Past and Future Perspectives. Journal of Developmental & Behavioral Pediatrics, 2001;22(4).
110. Petrou S. Methodological Issues Raised by Preference-Based Approaches to Measuring the Health Status of Children. Health Economics, 2003;12(8):697-702. https://doi.org/10.1002/hec.775.
111. Thorrington D., Eames K. Measuring Health Utilities in Children and Adolescents: A Systematic Review of the Literature. PLoS One, 2015;10(8):e0135672. https://doi.org/10.1371/journal.pone.0135672.

112. Wolstenholme J. L., Bargo D., Wang K., Harnden A., Räisänen U., Abel L. Preference-Based Measures to Obtain Health State Utility Values for Use in Economic Evaluations with Child-Based Populations: A Review and UK-Based Focus Group Assessment of Patient and Parent Choices. Quality of Life Research, 2018;27(7):1769-1780. https://doi.org/10.1007/s11136-018-1831-6.
113. Mascarenhas E., Miguel L. S., Oliveira M. D., Fernandes R. M. Economic Evaluations of Medical Devices in Paediatrics: A Systematic Review and a Quality Appraisal of the Literature. Cost Effectiveness and Resource Allocation, 2024;22(1):33. https://doi.org/10.1186/s12962-024-00537-0.
114. Bailey C., Howell M., Raghunandan R., Salisbury A., Chen G., Coast J., et al. Preference Elicitation Techniques Used in Valuing Children's Health-Related Quality-of-Life: A Systematic Review. PharmacoEconomics, 2022;40(7):663-698. https://doi.org/10.1007/s40273-022-01149-3.
115. 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. Social Science & Medicine, 2019;240:112543. https://doi.org/10.1016/j.socscimed.2019.112543.
116. Jiang M., Ma Y., Li M., Meng R., Ma A., Chen P. A comparison of self-reported and proxy-reported health utilities in children: a systematic review and meta-analysis. Health and Quality of Life Outcomes, 2021;19(1):45. https://doi.org/10.1186/s12955-021-01677-0.
117. Shoukri M. M. Measurement of Agreement In: N. Balakrishnan T. C., B. Everitt, W. Piegorsch, F. Ruggeri and J.L. Teugels, editor. Wiley StatsRef: Statistics Reference Online. New Jersey: John Wiley & Sons, Ltd; 2014.
118. Kottner J., Audigé L., Brorson S., Donner A., Gajewski B. J., Hróbjartsson A., et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Journal of Clinical Epidemiology, 2011;64(1):96-106. https://doi.org/10.1016/j.jclinepi.2010.03.002.
119. Roydhouse J. K., Cohen M. L., Eshoj H. R., Corsini N., Yucel E., Rutherford C., et al. The Use of Proxies and Proxy-Reported Measures: A Report of the International Society for Quality of Life Research (ISOQOL) Proxy Task Force. Quality of Life Research, 2022;31(2):317-327. https://doi.org/10.1007/s11136-021-02937-8.
120. Hernández J. D., Spir M. A., Payares K., Posada A. M., Salinas F. A., Garcia H. I., et al. Assessment by Proxy of the SF-36 and WHO-DAS 2.0: A Systematic Review. Journal of Rehabilitation Medicine, 2023;55:jrm4493. https://doi.org/10.2340/jrm.v55.4493.
121. Dolan P., Olsen J. A., Menzel P., Richardson J. An Inquiry Into the Different Perspectives That Can Be Used When Eliciting Preferences in Health. Health Economics, 2003;12(7):545-51. https://doi.org/10.1002/hec.760.
122. Pickard A. S., Knight S. J. Proxy Evaluation of Health-Related Quality of Life: A Conceptual Framework for Understanding Multiple Proxy Perspectives. Medical Care, 2005;43(5):493-499. https://doi.org/10.1097/01.mlr.0000160419.27642.a8.
123. Wyrwich K. W., Metz S. M., Kroenke K., Tierney W. M., Babu A. N., Wolinsky F. D. Triangulating Patient and Clinician Perspectives on Clinically Important Differences in Health-Related Quality of Life Among Patients with Heart Disease. Health Services Insights, 2007;42(6 Pt 1):2257-74; discussion 2294-323. https://doi.org/10.1111/j.1475-6773.2007.00733.x.
124. 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. Medical Care, 2008;46(11):1140-1148. https://doi.org/10.1097/MLR.0b013e31817d69a6.
125. Tol M., Kuipers J. P., Willigenburg N. W., Willems H. C., Poolman R. W. 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 and Quality of Life Outcomes, 2021;19(1):35. https://doi.org/10.1186/s12955-021-01679-y.
126. Cremeens J., Eiser C., Blades M. Factors Influencing Agreement Between Child Self-Report and Parent Proxy-Reports on the Pediatric Quality of Life Inventory 4.0 (PedsQL) Generic Core Scales. Health and Quality of Life Outcomes, 2006;4:58. https://doi.org/10.1186/1477-7525-4-58.

127. Bulamu N. B., Kaambwa B., Ratcliffe J. A Systematic Review of Instruments for Measuring Outcomes in Economic Evaluation Within Aged Care. Health and Quality of Life Outcomes, 2015;13:179. https://doi.org/10.1186/s12955-015-0372-8.
128. Varni J. W., Limbers C. A. The Pediatric Quality of Life Inventory: Measuring Pediatric Health-Related Quality of Life from the Perspective of Children and Their Parents. Pediatric Clinics of North America, 2009;56(4):843-863. https://doi.org/10.1016/j.pcl.2009.05.016.
129. Golicki D., Młyńczak K. Measurement Properties of the EQ-5D-Y: A Systematic Review. Value in Health, 2022;25(11):1910-1921. https://doi.org/10.1016/j.jval.2022.05.013.
130. 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.
131. Jones R., Mulhern B., McGregor K., Yip S., O'Loughlin R., Devlin N., et al. Psychometric Performance of HRQoL Measures: An Australian Paediatric Multi-Instrument Comparison Study Protocol (P-MIC). Children (Basel, Switzerland), 2021;8(8). https://doi.org/10.3390/children8080714.
132. Ngwira L. G., Maheswaran H., Verstraete J., Petrou S., Niessen L., Smith S. C. Psychometric Performance of the Chichewa Versions of the EQ-5D-Y-3L and EQ-5D-Y-5L Among Healthy and Sick Children and Adolescents in Malawi. Journal of Patient-Reported Outcomes, 2023;7(1):22. https://doi.org/10.1186/s41687-023-00560-4.
133. Verstraete J., Amien R., Scott D. Comparing Measurement Properties of the English EQ-5D-Y 3-Level Version With the 5-Level Version in South Africa. Value in Health Regional Issues, 2022;30:140-147. https://doi.org/10.1016/j.vhri.2021.12.004.
134. Scott D., Ferguson G. D., Jelsma J. The Use of the EQ-5D-Y Health-Related Quality of Life Outcome Measure in Children in the Western Cape, South Africa: Psychometric Properties, Feasibility, and Usefulness - A Longitudinal, Analytical Study. Health and Quality of Life Outcomes, 2017;15(1):12. https://doi.org/10.1186/s12955-017-0590-3.
135. Wu X. Y., Ohinmaa A., Johnson J. A., Veugelers P. J. Assessment of Children's Own Health Status Using Visual Analogue Scale and Descriptive System of the EQ-5D-Y: Linkage Between Two Systems. Quality of Life Research, 2014;23(2):393-402. https://doi.org/10.1007/s11136-013-0479-5.
136. Wu X., Veugelers P. J., Ohinmaa A. Health Behavior, Health-Related Quality of Life, and Mental Health Among Canadian Children: A Population-Based Cohort Study. Frontiers in nutrition, 2021;8:638259. https://doi.org/10.3389/fnut.2021.638259.
137. Åström M., Persson C., Lindén-Boström M., Rolfson O., Burström K. Population Health Status Based on the EQ-5D-Y-3L Among Adolescents in Sweden: Results by Sociodemographic Factors and Self-Reported Comorbidity. Quality of Life Research, 2018;27(11):2859-2871. https://doi.org/10.1007/s11136-018-1985-2.
138. Palacios-Cartagena R. P., Adsuar J. C., Hernández-Mocholí M., Carlos-Vivas J., Barrios-Fernández S., Garcia-Gordillo M. A., et al. Health-Related Quality of Life Norm Data of the Peruvian Adolescents: Results Using the EQ-5D-Y. International Journal of Environmental Research and Public Health, 2021;18(16). https://doi.org/10.3390/ijerph18168735.
139. Fraser H., Thompson L., Crawley E., Ridd M., Brigden A. "Not One Size Fits All": The Challenges of Measuring Paediatric Health-Related Quality of Life and the Potential Role of Digital Ecological Momentary Assessment: A Qualitative Study. Quality of Life Research, 2023;33. https://doi.org/10.1007/s11136-023-03535-6.
140. Cremeens J., Eiser C., Blades M. A Qualitative Investigation of School-Aged Children's Answers to Items From a Generic Quality of Life Measure. Child: Care, Health and Development, 2007;33(1):83-89. https://doi.org/10.1111/j.1365-2214.2006.00665.x.
141. Solans M., Pane S., Estrada M.-D., Serra-Sutton V., Berra S., Herdman M., et al. Health-Related Quality of Life Measurement in Children and Adolescents: A Systematic Review of Generic and Disease-Specific Instruments. Value in Health, 2008;11(4):742-764. https://doi.org/10.1111/j.1524-4733.2007.00293.x.
142. Foundation E. R. *EQ-5D-Y User Guide*. 2020.
143. Amien R., Scott D., Verstraete J. Performance of the EQ-5D-Y Interviewer Administered Version in Young Children. Children (Basel, Switzerland), 2022;9(1):93. https://doi.org/10.3390/children9010093.

144. 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. The European Journal of Health Economics, 2021;22(7):1103-1113. https://doi.org/10.1007/s10198-021-01309-w.
145. Shiroiwa 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. Quality of Life Research, 2019;28(11):3093-3105. https://doi.org/10.1007/s11136-019-02238-1.
146. Canaway A. G., Frew E. J. 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. Quality of Life Research, 2013;22(1):173-183. https://doi.org/10.1007/s11136-012-0119-5.
147. Gusi N., Perez-Sousa M. A., Gozalo-Delgado M., Olivares P. R. [Validity and Reliability of the Spanish EQ-5D-Y Proxy Version]. Anales de pediatria (Barcelona, Spain : 2003), 2014;81(4):212-9. https://doi.org/10.1016/j.anpedi.2013.11.028.
148. Bray N., Noyes J., Harris N., Edwards R. T. Measuring the Health-Related Quality of Life of Children With Impaired Mobility: Examining Correlation and Agreement Between Children and Parent Proxies. BMC research notes, 2017;10(1):377. https://doi.org/10.1186/s13104-017-2683-9.
149. Babineau J. Product Review: Covidence (Systematic Review Software). The Journal of the Canadian Health Libraries Association, 2014;35(2):68-71. http://dx.doi.org/10.5596/c14-016.
150. Page M. J., McKenzie J. E., Bossuyt P. M., Boutron I., Hoffmann T. C., Mulrow C. D., et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. Systematic Reviews, 2021;10(1):89. https://doi.org/10.1186/s13643-021-01626-4.
151. Wongpakaran N., Wongpakaran T., Wedding D., Gwet K. L. A Comparison of Cohen's Kappa and Gwet's AC1 When Calculating Inter-Rater Reliability Coefficients: A Study Conducted with Personality Disorder Samples. BMC Medical Research Methodology, 2013;13:61. https://doi.org/10.1186/1471-2288-13-61.
152. Lin L. I. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics, 1989;45(1):255-268. https://doi.org/10.2307/2532051.
153. Ranganathan P., Pramesh C. S., Aggarwal R. Common Pitfalls in Statistical Analysis: Measures of Agreement. Perspectives in Clinical Research, 2017;8(4):187-191. https://doi.org/10.4103/picr.PICR_123_17.
154. Liu J., Tang W., Chen G., Lu Y., Feng C., Tu X. M. Correlation and Agreement: Overview and Clarification of Competing Concepts and Measures. Shanghai archives of psychiatry, 2016;28(2):115-120. https://doi.org/10.11919/j.issn.1002-0829.216045.
155. Stemler S. E. A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability. Practical Assessment, Research and Evaluation, 2004;9:4.
156. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 1960;20(1):37-46. https://doi.org/10.1177/001316446002000104.
157. Koo T. K., Li M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine, 2016;15(2):155-163. https://doi.org/10.1016/j.jcm.2016.02.012.
158. Juniper E. F., Guyatt G. H., Jaeschke R. How to Develop and Validate a New Health-Related Quality of Life Instrument. 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1996.
159. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed: Academic press; 1988. p. 82.
160. Landis J. R., Koch G. G. The Measurement of Observer Agreement for Categorical Data. Biometrics, 1977;33(1):159-74. https://doi.org/10.2307/2529310.
161. Gwet K. L. Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters. 4 ed: Advanced Analytics, LLC; 2014.
162. Campbell M., McKenzie J. E., Sowden A., Katikireddi S. V., Brennan S. E., Ellis S., et al. Synthesis Without Meta-Analysis (SWiM) in Systematic Reviews: Reporting Guideline. BMJ, 2020;368:l6890. https://doi.org/10.1136/bmj.l6890.
163. Popay J., Roberts H., Sowden A., Petticrew M., Arai L., Rodgers M., et al. *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme Version*. 2006. **1**, b92 DOI: https://doi.org/10.13140/2.1.1018.4643.

164. Konstantinidis M., Le L. W., Gao X. An Empirical Comparative Assessment of Inter-Rater Agreement of Binary Outcomes and Multiple Raters. Symmetry, 2022;14(2):262. https://doi.org/10.3390/sym14020262.

165. Raudenbush S. W. Analyzing Effect Sizes: Random-Effects Models. The Handbook of Research Synthesis and Meta-Analysis, 2nd Ed. New York, NY, US: Russell Sage Foundation; 2009. p. 295-315.

166. Sánchez-Meca J., López-López J. A., López-Pina J. A. Some Recommended Statistical Analytic Practices When Reliability Generalization Studies Are Conducted. The British Journal of Mathematical and Statistical Psychology, 2013;66(3):402-25. https://doi.org/10.1111/j.2044-8317.2012.02057.x.

167. Sun S. Meta-Analysis of Cohen's Kappa. Health Services and Outcomes Research Methodology, 2011;11(3):145-163. https://doi.org/10.1007/s10742-011-0077-3.

168. Tufanaru C., Munn Z., Stephenson M., Aromataris E. Fixed or Random Effects Meta-Analysis? Common Methodological Issues in Systematic Reviews of Effectiveness. International Journal of Evidence-based Healthcare, 2015;13(3):196-207. https://doi.org/10.1097/xeb.0000000000000065.

169. Kmet L. M., Cook L. S., Lee R. C. Standard Quality Assessment Criteria for Evaluating Primary Research Papers From a Variety of Fields. 2004. https://doi.org/10.7939/R37M04F16.

170. Papaioannou D., Brazier J., Paisley S. Systematic Searching and Selection of Health State Utility Values from the Literature. Value in Health, 2013;16(4):686-95. https://doi.org/10.1016/j.jval.2013.02.017.

171. Barr R. D., Simpson T., Whitton A., Rush B., Furlong W., Feeny D. H. 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. European journal of cancer, 1999;35(2):248-55. https://doi.org/10.1016/s0959-8049(98)00366-9.

172. Jelsma J., Ramma L. How Do Children at Special Schools and Their Parents Perceive Their HRQoL Compared to Children at Open Schools? Health and Quality of Life Outcomes, 2010;8:72-72. https://doi.org/10.1186/1477-7525-8-72.

173. Verrips G. H., Stuifbergen M. C., den Ouden A. L., Bonsel G. J., Gemke R. J., Paneth N., et al. Measuring Health Status Using the Health Utilities Index: Agreement Between Raters and Between Modalities of Administration. Journal of Clinical Epidemiology, 2001;54(5):475-81. https://doi.org/10.1016/s0895-4356(00)00317-6.

174. Morrow A. M., Hayen A., Quine S., Scheinberg A., Craig J. C. 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 and Development, 2012;38(2):186-95. https://doi.org/10.1111/j.1365-2214.2011.01240.x.

175. Wolke D., Chernova J., Eryigit-Madzwamuse S., Samara M., Zwierzynska K., Petrou S. Self and Parent Perspectives on Health-Related Quality of Life of Adolescents Born Very Preterm. The Journal of Pediatrics, 2013;163(4):1020-6.e2. https://doi.org/10.1016/j.jpeds.2013.04.030.

176. Fitriana T. S., Purba F. D., Stolk E., Busschbach J. J. V. EQ-5D-Y-3L and EQ-5D-Y-5L Proxy Report: Psychometric Performance and Agreement With Self-Report. Health and Quality of Life Outcomes, 2022;20(1):88. https://doi.org/10.1186/s12955-022-01996-w.

177. Glaser A., Kennedy C., Punt J., Walker D. Standardized Quantitative Assessment of Brain Tumor Survivors Treated Within Clinical Trials in Childhood. International journal of cancer, 1999;12:77-82. https://doi.org/10.1002/(sici)1097-0215(1999)83:12+<77::aid-ijc14>3.0.co;2-x.

178. Glaser A. W., Furlong W., Walker D. A., Fielding K., Davies K., Feeny D. H., et al. Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the U.K. European Journal of Cancer, 1999;35(2):256-61. https://doi.org/10.1016/s0959-8049(98)00367-0.

179. Sung L., Young N. L., Greenberg M. L., McLimont M., Samanta T., Wong J., et al. Health-Related Quality of Life (HRQL) Scores Reported from Parents and Their Children with Chronic Illness Differed Depending on Utility Elicitation Method. Journal of Clinical Epidemiology, 2004;57(11):1161-6. https://doi.org/10.1016/j.jclinepi.2004.05.003.

180. Fu L., Talsma D., Baez F., Bonilla M., Moreno B., Ah-Chu M., et al. Measurement of Health-Related Quality of Life in Survivors of Cancer in Childhood in Central America: Feasibility, Reliability, and Validity. Journal of Pediatric Hematology/Oncology, 2006;28(6):331-41. https://doi.org/10.1097/00043426-200606000-00003.
181. Banks B. A., Barrowman N. J., Klaassen R. Health-Related Quality of Life: Changes in Children Undergoing Chemotherapy. Journal of Pediatric Hematology/Oncology, 2008;30(4):292-7. https://doi.org/10.1097/MPH.0b013e3181647bda.
182. Fluchel M., Horsman J. R., Furlong W., Castillo L., Alfonz Y., Barr R. D. Self- and Proxy-Reported Health Status and Health-Related Quality of Life in Survivors of Childhood Cancer in Uruguay. Pediatric Blood & Cancer, 2008;50(4):838-43. https://doi.org/10.1002/pbc.21299.
183. Penn A., Lowis S. P., Stevens M. C., Shortman R. I., Hunt L. P., McCarter R. J., et al. A Detailed Prospective Longitudinal Assessment of Health Status in Children with Brain Tumors in the First Year After Diagnosis. Journal of Pediatric Hematology/Oncology, 2011;33(8):592-9. https://doi.org/10.1097/MPH.0b013e31821388c0.
184. Hetherington K., Wakefield C. E., Kunalan K. P. K., Donoghoe M. W., McGill B. C., Fardell J. E., et al. Quality of Life (QoL) of Children and Adolescents Participating in a Precision Medicine Trial for High-Risk Childhood Cancer. Cancers, 2022;14(21):5310. https://doi.org/10.3390/cancers14215310.
185. Czyzewski D. I., Mariotto M. J., Bartholomew L. K., LeCompte S. H., Sockrider M. M. Measurement of Quality of Well Being in a Child and Adolescent Cystic Fibrosis Population. Medical Care, 1994;32(9):965-972. https://doi.org/10.1097/00005650-199409000-00007.
186. Brunner H. I., Maker D., Grundland B., Young N. L., Blanchette V., Stain A. M., et al. Preference-Based Measurement of Health-Related Quality of Life (HRQL) in Children With Chronic Musculoskeletal Disorders (MSKDs). Medical Decision Making, 2003;23(4):314-22. https://doi.org/10.1177/0272989x03256008.
187. Belfort M. B., Zupancic J. A., Riera K. M., Turner J. H., Prosser L. A. Health State Preferences Associated With Weight Status in Children and Adolescents. BMC Pediatrics, 2011;11:12. https://doi.org/10.1186/1471-2431-11-12.
188. Lee J. M., Rhee K., O'Grady M. J., Basu A., Winn A., John P., et al. Health Utilities for Children and Adults with Type 1 Diabetes. Medical care, 2011;49(10):924-931. https://doi.org/10.1097/MLR.0b013e318216592c.
189. Rhodes E. T., Prosser L. A., Lieu T. A., Songer T. J., Ludwig D. S., Laffel L. M. Preferences for Type 2 Diabetes Health States Among Adolescents with or at Risk of Type 2 Diabetes Mellitus. Pediatric Diabetes, 2011;12(8):724-732. https://doi.org/10.1111/j.1399-5448.2011.00772.x.
190. Ungar W. J., Boydell K., Dell S., Feldman B. M., Marshall D., Willan A., et al. A Parent-Child Dyad Approach to the Assessment of Health Status and Health-Related Quality of Life in Children with Asthma. PharmacoEconomics, 2012;30(8):697-712. https://doi.org/10.2165/11597890-000000000-00000.
191. Kulpeng W., Sornsrivichai V., Chongsuvivatwong V., Rattanavipapong W., Leelahavarong P., Cairns J., et al. Variation of Health-Related Quality of Life Assessed by Caregivers and Patients Affected by Severe Childhood Infections. BMC Pediatrics, 2013;13:122. https://doi.org/10.1186/1471-2431-13-122.
192. Sims-Williams H. J., Sims-Williams H. P., Mbabazi Kabachelor E., Warf B. C. Quality of Life Among Children with Spina Bifida in Uganda. Archives of Disease in Childhood, 2017;102(11):1057-1061. https://doi.org/10.1136/archdischild-2016-312307.
193. Bharij A., Neighbors K., Alonso E. M., Mohammad S. Health utility and quality of life in pediatric liver transplant recipients. Pediatric Transplantation, 2020;24(4):e13720. https://doi.org/10.1111/petr.13720.
194. Perez Sousa M., Olivares Sánchez-Toledo P. R., Gusi Fuerte N. Parent-Child Discrepancy in the Assessment of Health-Related Quality of Life Using the EQ-5D-Y Questionnaire. Archivos Argentinos de Pediatria, 2017;115(6):541-546. https://doi.org/10.5546/aap.2017.eng.541.
195. Perez-Sousa M. A., Olivares P. R., Garcia-Hermoso A., Gusi N. Does Anthropometric and Fitness Parameters Mediate the Effect of Exercise on the HRQoL of Overweight and Obese

Children/Adolescents? Quality of Life Research, 2018;27(9):2305-2312. https://doi.org/10.1007/s11136-018-1893-5.

196. van Summeren J., Klunder J. W., Holtman G. A., Kollen B. J., Berger M. Y., Dekker J. H. Parent-child Agreement on Health-Related Quality of Life in Children With Functional Constipation in Primary Care. Journal of Pediatric Gastroenterology and Nutrition, 2018;67(6):726-731. https://doi.org/10.1097/mpg.0000000000002124.

197. Rogers H. J., Vermaire J. H., Gilchrist F., Schuller A. A. The Relationship between Caries-Specific Quality of Life and Generic Wellbeing in a Dutch Pediatric Population. Dentistry Journal (Basel), 2019;7(3). https://doi.org/10.3390/dj7030067.

198. Sinlapamongkolkul P., Surapolchai P. Health-Related Quality of Life in Thai Children with Thalassemia as Evaluated by PedsQL and EQ-5D-Y: A Single-Center Experience. Mediterranean Journal of Hematology and Infectious Diseases, 2020;12(1):e2020036. https://doi.org/10.4084/mjhid.2020.036.

199. Kate Monica Isabel R., Bull K., Trotter C., Wood S., Jones C. E. Paediatric Health-Related Quality of Life in Congenital Cytomegalovirus. Archives of Disease in Childhood, 2022;107(11):1004-1011. https://doi.org/10.1136/archdischild-2022-324007.

200. Abraham S., Edginton E., Cottrell D., Tubeuf S. Measuring health-related quality of life measures in children: lessons from a pilot study. Research in Psychotherapy: Psychopathology, Process, and Outcome, 2022;25(1).

201. Lin J., Wong C. K. H., Cheung P. W. H., Luo N., Cheung J. P. Y. 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-e807. https://doi.org/10.1097/brs.0000000000003431.

202. Horsman J., Furlong W., Feeny D., Torrance G. The Health Utilities Index (HUI): Concepts, Measurement Properties, and Applications. Health and Quality of Life Outcomes, 2003;1:54-54. https://doi.org/10.1186/1477-7525-1-54.

203. Janssen M. F., Bonsel G. J., Luo N. Is EQ-5D-5L Better Than EQ-5D-3L? A Head-to-Head Comparison of Descriptive Systems and Value Sets from Seven Countries. PharmacoEconomics, 2018;36(6):675-697. https://doi.org/10.1007/s40273-018-0623-8.

204. Sprangers M. A. G., Aaronson N. K. The Role of Health Care Providers and Significant Others in Evaluating the Quality of Life of Patients with Chronic Disease: A Review. Journal of Clinical Epidemiology, 1992;45(7):743-760. https://doi.org/10.1016/0895-4356(92)90052-O.

205. Weinfurt K. P., Trucco S. M., Willke R. J., Schulman K. A. Measuring Agreement Between Patient and Proxy Responses to Multidimensional Health-Related Quality-of-Life Measures in Clinical Trials: An Application of Psychometric Profile Analysis. Journal of Clinical Epidemiology, 2002;55(6):608-18. https://doi.org/10.1016/s0895-4356(02)00392-x.

206. Sneeuw K. C., Aaronson N. K., Sprangers M. A., Detmar S. B., Wever L. D., Schornagel J. H. Value of Caregiver Ratings in Evaluating the Quality of Life of Patients with Cancer. Journal of Clinical Oncology, 1997;15(3):1206-1217. https://doi.org/10.1200/JCO.1997.15.3.1206.

207. Raley S., Bianchi S. M., Wang W. When Do Fathers Care? Mothers' Economic Contribution and Fathers' Involvement in Child Care. American Journal of Sociology, 2012;117(5):1422-1459. https://doi.org/10.1086/663354.

208. Singh I., Asnani M., Harrison A. Health-Related Quality of Life in Adolescents With Chronic Illness in Jamaica: Adolescent and Parent Reports. Journal of Adolescent Health, 2022;72. https://doi.org/10.1016/j.jadohealth.2022.07.013.

209. Creswell C., Violato M., Fairbanks H., White E., Parkinson M., Abitabile G., et al. Clinical Outcomes and Cost-Effectiveness of Brief Guided Parent-Delivered Cognitive Behavioural Therapy and Solution-Focused Brief Therapy for the Treatment of Childhood Anxiety Disorders: A Randomised Controlled Trial. Lancet Psychiatry, 2017;4(7):529-539. https://doi.org/10.1016/s2215-0366(17)30149-9.

210. Vermeulen K. M., Jansen D. E., Knorth E. J., Buskens E., Reijneveld S. A. Cost-Effectiveness of Multisystemic Therapy Versus Usual Treatment for Young People with Antisocial Problems. Criminal Behaviour and Mental Health, 2017;27(1):89-102. https://doi.org/10.1002/cbm.1988.

211. Welie A. G., Stolk E., Mukuria C., Belay Y. B., Krahn M. D., Sander B., et al. Reliability and Validity of Using EQ-5D-5L Among Healthy Adolescents and Adolescents with Major Mental Health Disorders in Ethiopia. The European Journal of Health Economics, 2022;23(7):1105-1119. https://doi.org/10.1007/s10198-021-01412-y.
212. Le L. K., Richards-Jones S., Chatterton M. L., Engel L., Lawrence D., Stevenson C., et al. Australian Adolescent Population Norms for the Child Health Utility Index 9D: Results from the Young Minds Matter Survey. Quality of Life Research, 2021;30(10):2895-2906. https://doi.org/10.1007/s11136-021-02864-8.
213. Wolf R. T., Jeppesen P., Pedersen M. M. A., Puggaard L. B., Thastum M., Bilenberg N., et al. Cost-Effectiveness of a Transdiagnostic Psychotherapy Program for Youth with Common Mental Health Problems. BMC Health Services Research, 2022;22(1):819. https://doi.org/10.1186/s12913-022-08187-9.
214. Radez J., Reardon T., Creswell C., Lawrence P. J., Evdoka-Burton G., Waite P. Why Do Children and Adolescents (Not) Seek and Access Professional Help for Their Mental Health Problems? A Systematic Review of Quantitative and Qualitative Studies. European Child & Adolescent Psychiatry, 2021;30(2):183-211. https://doi.org/10.1007/s00787-019-01469-4.
215. De Los Reyes A., Cook C. R., Gresham F. M., Makol B. A., Wang M. Informant Discrepancies in Assessments of Psychosocial Functioning in School-Based Services and Research: Review and Directions for Future Research. Journal of School Psychology, 2019;74:74-89. https://doi.org/10.1016/j.jsp.2019.05.005.
216. Rajmil L., López A. R., López-Aguilà S., Alonso J. Parent-Child Agreement on Health-Related Quality of Life (HRQOL): A Longitudinal Study. Health and Quality of Life Outcomes, 2013;11:101. https://doi.org/10.1186/1477-7525-11-101.
217. Tain Y.-L., Lu P.-C., Kuo H.-C., Hsu C.-N. Differences in Health-Related Quality of Life in Children with Chronic Kidney Disease as Reported by Children and Parent Proxies. Pediatric Nephrology, 2023;38(2):519-528. https://doi.org/10.1007/s00467-022-05621-2.
218. Lopez-Espejo M. A., Nuñez A. C., Ruz M., Saez V., Moscoso O. C., Escobar R. G. Caregiver-Child Agreement in Health-Related Quality of Life of Autistic Children and Adolescents. Journal of Autism and Developmental Disorders, 2023;53(3):977-986. https://doi.org/10.1007/s10803-022-05689-7.
219. Gothwal V. K., Bharani S., Mandal A. K. Parent-Child Agreement on Health-Related Quality of Life in Congenital Glaucoma. Translational Vision Science & Technology, 2018;7(4):15. https://doi.org/10.1167/tvst.7.4.15.
220. Seid M., Varni J. W., Jacobs J. R. Pediatric Health-Related Quality-of-Life Measurement Technology: Intersections between Science, Managed Care, and Clinical Care. Journal of Clinical Psychology in Medical Settings, 2000;7(1):17-27. https://doi.org/10.1023/A:1009541218764.
221. Fischer K. I., Barthel D., Otto C., Ravens-Sieberer U., Thyen U., Klein M., et al. Minimal Associations Between Clinical Data and Children's Self-Reported Health-Related Quality of Life in Children With Chronic Conditions—A Cross-Sectional Study. Frontiers in Pediatrics, 2019;7. https://doi.org/10.3389/fped.2019.00017.
222. 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):1043-1067. https://doi.org/10.1007/s40273-022-01177-z.
223. Upton P., Lawford J., Eiser C. Parent–Child Agreement Across Child Health-Related Quality of Life Instruments: A Review of the Literature. Quality of Life Research, 2008;17(6):895. https://doi.org/10.1007/s11136-008-9350-5.
224. Padilla J.-L., Leighton J. P. Cognitive Interviewing and Think Aloud Methods. In: Zumbo B. D.,Hubley A. M., editors. Understanding and Investigating Response Processes in Validation Research. Cham: Springer International Publishing; 2017. p. 211-228.
225. Ericsson K. A., Simon H. A. Verbal Reports as Data. Psychological Review, 1980;87:215-251. https://doi.org/10.1037/0033-295X.87.3.215.
226. Rogers H. J., Gilchrist F., Marshman Z., Rodd H. D., Rowen D. Selection and Validation of a Classification System for a Child-Centered Preference-Based Measure of Oral Health-

Related Quality of Life Specific to Dental Caries. Journal of Patient-Reported Outcomes, 2020;4(1):105. https://doi.org/10.1186/s41687-020-00268-9.
227. Nutakki K., Varni J. W., Steinbrenner S., Draucker C. B., Swigonski N. L. Development of the Pediatric Quality of Life Inventory Neurofibromatosis Type 1 Module Items for Children, Adolescents, and Young Adults: Qualitative Methods. Journal of Neuro-Oncology, 2017;132(1):135-143. https://doi.org/10.1007/s11060-016-2351-2.
228. Panepinto J. A., Torres S., Varni J. W. Development of the PedsQL™ Sickle Cell Disease Module Items: Qualitative Methods. Quality of Life Research, 2012;21(2):341-357. https://doi.org/10.1007/s11136-011-9941-4.
229. Varni J. W., Kay M. T., Limbers C. A., Franciosi J. P., Pohl J. F. PedsQL Gastrointestinal Symptoms Module Item Development: Qualitative Methods. Journal of Pediatric Gastroenterology and Nutrition, 2012;54(5). https://doi.org/10.1097/MPG.0b013e31823c9b88.
230. Liu Y., Hinds P. S., Wang J., Correia H., Du S., Ding J., et al. Translation and Linguistic Validation of the Pediatric Patient-Reported Outcomes Measurement Information System Measures into Simplified Chinese Using Cognitive Interviewing Methodology. Cancer Nursing, 2013;36(5):368-76. https://doi.org/10.1097/NCC.0b013e3182962701.
231. Sepúlveda P. R., Molina G. T., Molina C. R., Martínez N. V., González A. E., L M. G., et al. [Validation of an Instrument to Measure Health-Related Quality of Life in Chilean Children and Adolescents]. Revista Medica de Chile, 2013;141(10):1283-92. https://doi.org/10.4067/s0034-98872013001000007.
232. García-Martínez de Bartolomé R., Barrio-Torres J., Cilleruelo-Pascual M. L. [Transcultural Adaptation of the Pediatric Eosinophilic Esophagitis Quality of Life Module Questionnaire]. Anales de pediatria, 2020;92(6):332-338. https://doi.org/10.1016/j.anpedi.2020.01.005.
233. Lawford J., Volavka N., Eiser C. A Generic Measure of Quality of Life for Children Aged 3-8 Years: Results of Two Preliminary Studies. Pediatric Rehabilitation, 2001;4(4):197-207. https://doi.org/10.1080/13638490210124033.
234. Hunter L. C., Molitor F., Chafetz R. S., Jane Mulcahey M., Vogel L. C., Betz R., et al. Development and Pilot Test of the Shriners Pediatric Instrument for Neuromuscular Scoliosis (SPNS): A Quality of Life Questionnaire for Children With Spinal Cord Injuries. The Journal of Spinal Cord Medicine, 2007;30(sup1):S150-S157. https://doi.org/10.1080/10790268.2007.11754594.
235. Eddy L., Khastou L., Cook K. F., Amtmann D. Item Selection in Self-Report Measures for Children and Adolescents with Disabilities: Lessons from Cognitive Interviews. Journal of Pediatric Nursing, 2011;26(6):559-565. https://doi.org/10.1016/j.pedn.2010.07.004.
236. Amien R., Scott D., Verstraete J. The Validity and Reliability of the Interviewer-Administered EQ-5D-Y-3L Version in Young Children. Health and Quality of Life Outcomes, 2023;21(1):19. https://doi.org/10.1186/s12955-023-02100-6.
237. Tourangeau R., Rips L. J., Rasinski K. The Psychology of Survey Response. Cambridge: Cambridge University Press; 2000.
238. Engel L., Bucholc J., Mihalopoulos C., Mulhern B., Ratcliffe J., Yates M., et al. A Qualitative Exploration of the Content and Face Validity of Preference-Based Measures Within the Context of Dementia. Health and Quality of Life Outcomes, 2020;18(1):178. https://doi.org/10.1186/s12955-020-01425-w.
239. Lay K., Crocker M., Engel L., Ratcliffe J., Charlton S., Hutchinson C. How Do Older Adults Receiving Aged Care Services Understand and Respond to the EQ-5D-5L? A Think-Aloud Study in Residential Care. Quality of Life Research, 2023. https://doi.org/10.1007/s11136-023-03466-2.
240. Nwankwo H., Coast J., Hewison A., Kinghorn P., Madathil S., Bailey C. A Think-Aloud Study of the Feasibility of Patients with End-Stage Organ Failure Completing the ICECAP-SCM. Palliative Medicine, 2022;36(10):1559-1569. https://doi.org/10.1177/02692163221122979.
241. Wright J., Moghaddam N., Dawson D. L. Cognitive Interviewing in Patient-Reported Outcome Measures: A Systematic Review of Methodological Processes. Qualitative Psychology, 2021;8:2-29. https://doi.org/10.1037/qup0000145.
242. Coast J., De Allegri M. Qualitative Methods in Health Economics: Oxford University Press; 2018.

243. Viney R., Norman R., King M. T., Cronin P., Street D. J., Knox S., et al. Time Trade-Off Derived EQ-5D Weights for Australia. Value in Health, 2011;14(6):928-36. https://doi.org/10.1016/j.jval.2011.04.009.
244. Kreimeier S., Mott D., Ludwig K., Greiner W. EQ-5D-Y Value Set for Germany. PharmacoEconomics, 2022;40(Suppl 2):217-229. https://doi.org/10.1007/s40273-022-01143-9.
245. Devlin N., Roudijk B., Viney R., Stolk E. EQ-5D-Y-3L Value Sets, Valuation Methods and Conceptual Questions. PharmacoEconomics, 2022;40(2):123-127. https://doi.org/10.1007/s40273-022-01226-7.
246. Varni J. W., Burwinkle T. M., Seid M. The PedsQL as a Pediatric Patient-Reported Outcome: Reliability and Validity of the PedsQL Measurement Model in 25,000 Children. Expert Review of Pharmacoeconomics & Outcomes Research, 2005;5(6):705-19. https://doi.org/10.1586/14737167.5.6.705.
247. Australian Bureau of Statistics. Survey of Income and Housing, User Guide. Australia; 2019-2020. Available from: https://www.abs.gov.au/statistics/detailed-methodology-information/concepts-sources-methods/survey-income-and-housing-user-guide-australia/2019-20/income#low-and-lower-income-households.
248. Australian Institute of Health and Welfare. Health of Children. Canberra: AIHW; 2020. Available from: https://www.aihw.gov.au/reports/australias-health/health-of-children.
249. Teddlie C., Yu F. Mixed Methods Sampling: A Typology With Examples. Journal of Mixed Methods Research, 2007;1(1):77-100. https://doi.org/10.1177/1558689806292430.
250. Shiroiwa 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):3093-3105. https://doi.org/10.1007/s11136-019-02238-1.
251. Creswell J. W., Clark V. L. P. Designing and Conducting Mixed Methods Research: Sage publications; 2017.
252. Idler E. L., Benyamini Y. Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. Journal of Health and Social Behavior, 1997;38(1):21-37.
253. Harris P. A., Taylor R., Thielke R., Payne J., Gonzalez N., Conde J. G. Research Electronic Data Capture (REDCap): A Metadata-Driven Methodology and Workflow Process for Providing Translational Research Informatics Support. Journal of Biomedical Informatics, 2009;42(2):377-381. https://doi.org/10.1016/j.jbi.2008.08.010.
254. Harris P. A., Taylor R., Minor B. L., Elliott V., Fernandez M., O'Neal L., et al. The REDCap Consortium: Building an International Community of Software Platform Partners. Journal of Biomedical Informatics, 2019;95:103208. https://doi.org/10.1016/j.jbi.2019.103208.
255. van den Haak M., De Jong M., Jan Schellens P. Retrospective vs. Concurrent Think-Aloud Protocols: Testing the Usability of an Online Library Catalogue. Behaviour & Information Technology, 2003;22(5):339-351. https://doi.org/10.1080/0044929031000.
256. Peute L. W., de Keizer N. F., Jaspers M. W. The Value of Retrospective and Concurrent Think Aloud in Formative Usability Testing of a Physician Data Query Tool. Journal of Biomedical Informatics, 2015;55:1-10. https://doi.org/10.1016/j.jbi.2015.02.006.
257. Guerriero C., Jaume N. A., Diaz-Ordaz K., Brown K. L., Wray J., Ashworth J., et al. Using Animation to Self-Report Health: A Randomized Experiment with Children. The Patient 2020;13(2):175-188. https://doi.org/10.1007/s40271-019-00392-9.
258. Moum T. Mode of Administration and Interviewer Effects in Self-Reported Symptoms of Anxiety and Depression. Social Indicators Research, 1998;45(1):279-318. https://doi.org/10.1023/A:1006958100504.
259. Muehlhausen W., Doll H., Quadri N., Fordham B., O'Donohoe P., Dogar N., et al. Equivalence of Electronic and Paper Administration of Patient-Reported Outcome Measures: A Systematic Review and Meta-Analysis of Studies Conducted Between 2007 and 2013. Health and Quality of Life Outcomes, 2015;13:167. https://doi.org/10.1186/s12955-015-0362-x.
260. Young N. L., Varni J. W., Snider L., McCormick A., Sawatzky B., Scott M., et al. The Internet is Valid and Reliable for Child-Report: An Example Using the Activities Scale for Kids (ASK) and the Pediatric Quality of Life Inventory (PedsQL). Journal of Clinical Epidemiology, 2009;62(3):314-320. https://doi.org/10.1016/j.jclinepi.2008.06.011.

261. Australian Bureau Of Statistics. Socio-Economic Indexes for Areas (SEIFA) 2033.0.55.001 Canberra; 2018. Available from: https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/756EE3DBEFA869EFCA258259000BA746/$File/SEIFA%202016%20Technical%20Paper.pdf.
262. Lin L., Hedayat A. S., Sinha B., Yang M. Statistical Methods in Assessing Agreement: Models, Issues, and Tools. Journal of the American Statistical Association, 2002;97(457):257-270. https://doi.org/10.1198/016214502753479392.
263. Chen C.-C., Barnhart H. X. Comparison of ICC and CCC for Assessing Agreement for Data Without and With Replications. Computational Statistics & Data Analysis, 2008;53(2):554-564. https://doi.org/10.1016/j.csda.2008.09.026.
264. Klein D. Implementing a General Framework for Assessing Interrater Agreement in Stata. Stata Journal, 2018;18(4):871-901.
265. StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC; 2019.
266. Nowell L. S., Norris J. M., White D. E., Moules N. J. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. International Journal of Qualitative Methods, 2017;16(1):1609406917733847. https://doi.org/10.1177/1609406917733847.
267. World Health Organization. ICD-10: International Statistical Classification of Diseases and Related Health Problems: Tenth Revision. 5th ed ed. Geneva: World Health Organization; 2016.
268. Australian Bureau of Statistics. 2071.0 - Census of Population and Housing: Reflecting Australia - Stories from the Census, 2016. Canberrra; 2018. Available from: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2071.0~2016~Main%20Features~Socio-Economic%20Advantage%20and%20Disadvantage~123.
269. 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 PedsQL(TM). Applied Health Economics and Health Policy, 2023;21(6):937-947. https://doi.org/10.1007/s40258-023-00831-7.
270. Conijn J. M., Smits N., Hartman E. E. Determining at What Age Children Provide Sound Self-Reports: An Illustration of the Validity-Index Approach. Assessment, 2019;27(7):1604-1618. https://doi.org/10.1177/1073191119832655.
271. Woolley M. E., Bowen G. L., Bowen N. K. Cognitive Pretesting and the Developmental Validity of Child Self-Report Instruments: Theory and Applications. Research on Social Work Practice, 2004;14(3):191-200. https://doi.org/10.1177/1049731503257882.
272. Qadeer R. A., Ferro M. A. Child–Parent Agreement on Health-Related Quality of Life in Children with Newly Diagnosed Chronic Health Conditions: A Longitudinal Study. International Journal of Adolescence and Youth, 2018;23(1):99-108. https://doi.org/10.1080/02673843.2017.1297242.
273. Petsios K., Priftis K. N., Tsoumakas C., Hatziagorou E., Tsanakas J. N., Galanis P., et al. Level of Parent–Asthmatic Child Agreement on Health-Related Quality of Life. The Journal of Asthma, 2011;48(3):286-297. https://doi.org/10.3109/02770903.2011.555031.
274. Kirkham F. J., 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 & Behavior, 2020;102:106340. https://doi.org/10.1016/j.yebeh.2019.04.058.
275. Vilagut G. Test-Retest Reliability. In: Michalos A. C., editor. Encyclopedia of Quality of Life and Well-Being Research. Dordrecht: Springer Netherlands; 2014. p. 6622-6625.
276. Hutchinson C., Worley A., Khadka J., Milte R., Cleland J., Ratcliffe J. Do We Agree or Disagree? A Systematic Review of the Application of Preference-Based Instruments in Self- and Proxy-Reporting of Quality of Life in Older People. Social Science & Medicine, 2022;305:115046. https://doi.org/10.1016/j.socscimed.2022.115046.
277. Arias J. A., Williams C., Raghvani R., Aghajani M., Baez S., Belzung C., et al. The Neuroscience of Sadness: A Multidisciplinary Synthesis and Collaborative Review. Neuroscience & Biobehavioral Reviews, 2020;111:199-228. https://doi.org/10.1016/j.neubiorev.2020.01.006.

278. Janse A. J., Sinnema G., Uiterwaal C. S. P. M., Kimpen J. L. L., Gemke R. J. B. J. Quality of Life in Chronic Illness: Children, Parents, and Paediatricians Have Different but Stable Perceptions. Acta Paediatrica, 2008;97(8):1118-1124. https://doi.org/10.1111/j.1651-2227.2008.00847.x.
279. Hemmingsson H., Ólafsdóttir L. B., Egilson S. T. Agreements and Disagreements Between Children and Their Parents in Health-Related Assessments. Disability and Rehabilitation, 2017;39(11):1059-1072. https://doi.org/10.1080/09638288.2016.1189603.
280. Varni J. W., Seid M., Rode C. A. The PedsQL™: Measurement Model for the Pediatric Quality of Life Inventory. Medical Care, 1999;37(2). https://doi.org/10.1097/00005650-199902000-00003.
281. Petersen K. D., 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 PedsQLTM 4.0 Short Form 15. The Patient 2018;11(1):29-37. https://doi.org/10.1007/s40271-017-0265-5.
282. Petersen S., Hägglöf B., Stenlund H., Bergström E. Psychometric Properties of the Swedish PedsQL, Pediatric Quality of Life Inventory 4.0 Generic Core Scales. Acta Paediatrica, 2009;98(9):1504-1512. https://doi.org/10.1111/j.1651-2227.2009.01360.x.
283. Gkoltsiou K., Dimitrakaki C., Tzavara C., Papaevangelou V., Varni J. W., Tountas Y. Measuring Health-Related Quality of Life in Greek Children: Psychometric Properties of the Greek Version of the Pediatric Quality of Life Inventory™ 4.0 Generic Core Scales. Quality of Life Research, 2008;17(2):299-305. https://doi.org/10.1007/s11136-007-9294-1.
284. Kook S. H., Varni J. W. Validation of the Korean Version of the Pediatric Quality of Life Inventory 4.0 (PedsQL) Generic Core Scales in School Children and Adolescents Using the Rasch Model. Health and Quality of Life Outcomes, 2008;6:41. https://doi.org/10.1186/1477-7525-6-41.
285. Hijkoop A., Ten Kate C. A., Madderom M. J., I. Jsselstijn H., Reuser J. A., Koopman H., et al. Sex Differences in Children's Health Status as Measured by the Pediatric Quality of Life Inventory (PedsQL)™: Cross-Sectional Findings From a Large School-Based Sample in the Netherlands. BMC Pediatrics, 2021;21(1):580. https://doi.org/10.1186/s12887-021-03059-3.
286. Pakpour A. H. Psychometric Properties of the Iranian Version of the Pediatric Quality of Life Inventory™ Short Form 15 Generic Core Scales. Singapore Medical Journal, 2013;54(6):309-14. https://doi.org/10.11622/smedj.2013123.
287. Ungar W. J., Mirabelli C., Cousins M., Boydell K. M. A Qualitative Analysis of a Dyad Approach to Health-Related Quality of Life Measurement in Children with Asthma. Social Science & Medicine, 2006;63(9):2354-2366. https://doi.org/10.1016/j.socscimed.2006.06.016.
288. Annett R. D., Bender B. G., DuHamel T. R., Lapidus J. Factors Influencing Parent Reports on Quality of Life for Children With Asthma. The Journal of Asthma, 2003;40(5):577-87. https://doi.org/10.1081/jas-120019030.
289. Ronen G. M., Streiner D. L., Rosenbaum P., Canadian Pediatric Epilepsy N. Health-related Quality of Life in Children with Epilepsy: Development and Validation of Self-report and Parent Proxy Measures. Epilepsia, 2003;44(4):598-612. https://doi.org/10.1046/j.1528-1157.2003.46302.x.
290. Collins D. Pretesting Survey Instruments: An Overview of Cognitive Methods. Quality of Life Research, 2003;12(3):229-238. https://doi.org/10.1023/A:1023254226592.
291. April K. T., Feldman D. E., Platt R. W., Duffy C. M. Comparison Between Children With Juvenile Idiopathic Arthritis (JIA) and Their Parents Concerning Perceived Quality of Life. Quality of Life Research, 2006;15(4):655-61. https://doi.org/10.1007/s11136-005-3690-1.
292. Britto M. T., Kotagal U. R., Chenier T., Tsevat J., Atherton H. D., Wilmott R. W. Differences Between Adolescents' and Parents' Reports of Health-Related Quality of Life in Cystic Fibrosis. Pediatric Pulmonology, 2004;37(2):165-71. https://doi.org/10.1002/ppul.10436.
293. Rebok G., Riley A., Forrest C., Starfield B., Green B., Robertson J., et al. Elementary School-Aged Children's Reports of Their Health: A Cognitive Interviewing Study. Quality of Life Research, 2001;10(1):59-70. https://doi.org/10.1023/a:1016693417166.
294. Brazier J., McCabe C. 'Is There a Case for Using Visual Analogue Scale Valuations in CUA?' by Parkin and Devlin: A Response—'Yes, There Is a Case, but What Does It Add to Ordinal Data?'. Health Economics, 2007;16(6):645-647. https://doi.org/10.1002/hec.1246.

295. Centers for Disease Control Prevention. Measuring Healthy Days: Population Assessment of Health-Related Quality of Life. Atlanta, Georgia; 2001. Available from: https://stacks.cdc.gov/view/cdc/6406/cdc_6406_DS1.pdf.

296. Arbuckle R., Abetz-Webb L. "Not Just Little Adults": Qualitative Methods to Support the Development of Pediatric Patient-Reported Outcomes. The Patient 2013;6(3):143-59. https://doi.org/10.1007/s40271-013-0022-3.

297. Varni J. W., Limbers C. A., Burwinkle T. M. How Young Can Children Reliably and Validly Self-Report Their Health-Related Quality of Life? An Analysis of 8,591 Children Across Age Subgroups with the PedsQL 4.0 Generic Core Scales. Health and Quality of Life Outcomes, 2007;5:1-1. https://doi.org/10.1186/1477-7525-5-1.

298. Amien R., Scott D., Verstraete J. Performance of the EQ-5D-Y Interviewer Administered Version in Young Children. Children, 2022;9(1):93. https://doi.org/10.3390/children9010093.

299. Stevens K., Palfreyman S. The Use of Qualitative Methods in Developing the Descriptive Systems of Preference-Based Measures of Health-Related Quality of Life for Use in Economic Evaluation. Value in Health, 2012;15(8):991-998. https://doi.org/10.1016/j.jval.2012.08.2204.

300. Mpundu-Kaambwa C., Bulamu N. B., Lines L., Chen G., Whitehurst D. G. T., Dalziel K., et al. Exploring the Use of Pictorial Approaches in the Development of Paediatric Patient-Reported Outcome Instruments: A Systematic Review. PharmacoEconomics, 2024;42(1):9-37. https://doi.org/10.1007/s40273-024-01357-z.

301. Nordlof J. Vygotsky, Scaffolding, and the Role of Theory in Writing Center Work. The Writing Center Journal, 2014;34(1):45-64.

302. Rowen D., Rivero-Arias O., Devlin N., Ratcliffe J. Review of Valuation Methods of Preference-Based Measures of Health for Economic Evaluation in Child and Adolescent Populations: Where are We Now and Where are We Going? PharmacoEconomics, 2020;38(4):325-340. https://doi.org/10.1007/s40273-019-00873-7.

303. Lipman S. A., Reckers-Droog V. T., Kreimeier S. Think of the Children: A Discussion of the Rationale for and Implications of the Perspective Used for EQ-5D-Y Health State Valuation. Value in Health, 2021;24(7):976-982. https://doi.org/10.1016/j.jval.2021.01.011.

304. Kind P., Klose K., Gusi N., Olivares P. R., Greiner W. Can Adult Weights Be Used to Value Child Health States? Testing the Influence of Perspective in Valuing EQ-5D-Y. Quality of Life Research, 2015;24(10):2519-39. https://doi.org/10.1007/s11136-015-0971-1.

305. Shah K. K., Ramos-Goñi J. M., Kreimeier S., Devlin N. J. An Exploration of Methods for Obtaining 0 = Dead Anchors for Latent Scale EQ-5D-Y Values. The European Journal of Health Economics, 2020;21(7):1091-1103. https://doi.org/10.1007/s10198-020-01205-9.

306. Powell P. A., Rowen D., Rivero-Arias O., Tsuchiya A., Brazier J. E. Valuing Child and Adolescent Health: A Qualitative Study on Different Perspectives and Priorities Taken by the Adult General Public. Health and Quality of Life Outcomes, 2021;19(1):222. https://doi.org/10.1186/s12955-021-01858-x.

307. Lipman S. A., Reckers-Droog V. T. Comparing Heuristic Valuation Processes Between Health State Valuation from Child and Adult Perspectives. The European Journal of Health Economics, 2024. https://doi.org/10.1007/s10198-023-01668-6.

308. EUROQOL. *Value Sets*. 2024 Available from: https://euroqol.org/information-and-support/resources/value-sets/.

309. Bailey C., Dalziel K., Cronin P., Devlin N., Viney R. How are Child-Specific Utility Instruments Used in Decision Making in Australia? A Review of Pharmaceutical Benefits Advisory Committee Public Summary Documents. PharmacoEconomics, 2022;40(2):157-182. https://doi.org/10.1007/s40273-021-01107-5.

310. Hill H., Rowen D., Pennington B., Wong R., Wailoo A. A Review of the Methods Used to Generate Utility Values in NICE Technology Assessments for Children and Adolescents. Value in Health, 2020;23(7):907-917. https://doi.org/10.1016/j.jval.2020.02.011.

311. Hu J., Zhu L., Bao H., Liu Y., Xing H., Kang Q., et al. Utility Estimations of Different Health States of Patients With Type I, II, and III Spinal Muscular Atrophy in China: A Mixed Approach Study With Patient- and Proxy-Reported Data. Frontiers in public health, 2022;10:1054931. https://doi.org/10.3389/fpubh.2022.1054931.

312. Ryan J. M., McKay E., Anokye N., Noorkoiv M., Theis N., Lavelle G. Comparison of the CHU-9D and the EQ-5D-Y Instruments in Children and Young People with Cerebral Palsy: A Cross-Sectional Study. BMJ Open, 2020;10(9):e037089. https://doi.org/10.1136/bmjopen-2020-037089.
313. 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 in Health, 2015;18(4):432-8. https://doi.org/10.1016/j.jval.2015.02.014.
314. Khanna D., Khadka J., Mpundu-Kaambwa C., Chen G., Dalziel K., Devlin N., et al. 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. PharmacoEconomics, 2024. https://doi.org/10.1007/s40273-024-01356-0.
315. Jones R., O'Loughlin R., Xiong X., Bahrampour M., McGregor K., Yip S., et al. Collecting Paediatric Health-Related Quality of Life Data: Assessing the Feasibility and Acceptability of the Australian Paediatric Multi-Instrument Comparison (P-MIC) Study. Children (Basel, Switzerland), 2023;10(10). https://doi.org/10.3390/children10101604.
316. Jones R., Mulhern B., Devlin N., Hiscock H., Chen G., O'Loughlin R., et al. *Australian Paediatric Multi-Instrument Comparison (P-MIC) Study: Technical Methods Paper, Version 4 (27/06/2023), Data Cut 3 (27/06/2023)*. 2023.
317. Devlin N P. D., Janssen B. Methods for Analysing and Reporting EQ-5D Data [Internet]. Cham (CH): Springer; 2020.
318. Devlin N. J., Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. Applied Health Economics and Health Policy, 2017;15(2):127-137. https://doi.org/10.1007/s40258-017-0310-5.
319. Norman R., Mulhern B., Lancsar E., Lorgelly P., Ratcliffe J., Street D., et al. The Use of a Discrete Choice Experiment Including Both Duration and Dead for the Development of an EQ-5D-5L Value Set for Australia. PharmacoEconomics, 2023;41(4):427-438. https://doi.org/10.1007/s40273-023-01243-0.
320. Prevolnik Rupel V., Ramos-Goñi J. M., Ogorevc M., Kreimeier S., Ludwig K., Greiner W. Comparison of Adult and Adolescent Preferences Toward EQ-5D-Y-3L Health States. Value in Health, 2021;24(9):1350-1359. https://doi.org/10.1016/j.jval.2021.03.019.
321. Yang Z., Jiang J., Wang P., Jin X., Wu J., Fang Y., et al. Estimating an EQ-5D-Y-3L Value Set for China. PharmacoEconomics, 2022;40(2):147-155. https://doi.org/10.1007/s40273-022-01216-9.
322. Roudijk B., Sajjad A., Essers B., Lipman S., Stalmeier P., Finch A. P. A Value Set for the EQ-5D-Y-3L in the Netherlands. PharmacoEconomics, 2022;40(Suppl 2):193-203. https://doi.org/10.1007/s40273-022-01192-0.
323. Shiroiwa T., Ikeda S., Noto S., Fukuda T., Stolk E. Valuation Survey of EQ-5D-Y Based on the International Common Protocol: Development of a Value Set in Japan. Medical Decision Making, 2021;41(5):597-606. https://doi.org/10.1177/0272989x211001859.
324. Fitriana T. S., Roudijk B., Purba F. D., Busschbach J. J. V., Stolk E. Estimating an EQ-5D-Y-3L Value Set for Indonesia by Mapping the DCE onto TTO Values. PharmacoEconomics, 2022;40(Suppl 2):157-167. https://doi.org/10.1007/s40273-022-01210-1.
325. Ramos-Goñi J. M., Oppe M., Estévez-Carrillo A., Rivero-Arias O., Wolfgang G., Simone K., et al. Accounting for Unobservable Preference Heterogeneity and Evaluating Alternative Anchoring Approaches to Estimate Country-Specific EQ-5D-Y Value Sets: A Case Study Using Spanish Preference Data. Value in Health, 2022;25(5):835-843. https://doi.org/10.1016/j.jval.2021.10.013.
326. Dewilde S., Roudijk B., Tollenaar N. H., Ramos-Goñi J. M. An EQ-5D-Y-3L Value Set for Belgium. PharmacoEconomics, 2022;40(Suppl 2):169-180. https://doi.org/10.1007/s40273-022-01187-x.
327. Prevolnik Rupel V., Ogorevc M. EQ-5D-Y Value Set for Slovenia. PharmacoEconomics, 2021;39(4):463-471. https://doi.org/10.1007/s40273-020-00994-4.
328. Rencz F., Ruzsa G., Bató A., Yang Z., Finch A. P., Brodszky V. Value Set for the EQ-5D-Y-3L in Hungary. PharmacoEconomics, 2022;40(2):205-215. https://doi.org/10.1007/s40273-022-01190-2.

329. Dłużniak-Gołaska K., Szostak-Węgierek D., Panczyk M., Szypowska A., Sińska B. May Gender Influence the Quality of Life in Children and Adolescents With Type 1 Diabetes? Patient Preference and Adherence, 2019;13:1589-1597. https://doi.org/10.2147/ppa.S206969.
330. Langeland I. O., Sollesnes R., Nilsen R. M., Almenning G., Langeland E. Examining Boys' and Girls' Health-Related Quality of Life from the First to the Third Year of Upper Secondary School: A Prospective Longitudinal Study. Nursing Open, 2019;6(4):1606-1614. https://doi.org/10.1002/nop2.366.
331. Michel G., Bisegger C., Fuhr D. C., Abel T., The K. g. Age and Gender Differences in Health-Related Quality of Life of Children and Adolescents in Europe: A Multilevel Analysis. Quality of Life Research, 2009;18(9):1147-1157. https://doi.org/10.1007/s11136-009-9538-3.
332. Helseth S., Haraldstad K., Christophersen K. A. A Cross-Sectional Study of Health-Related Quality of Life and Body Mass Index in a Norwegian School Sample (8–18 Years): A Comparison of Child and Parent Perspectives. Health and Quality of Life Outcomes, 2015;13:47. https://doi.org/10.1186/s12955-015-0239-z.
333. Wang J., Jin W., Shi L., Geng Y., Zhu X., Hu W. Health-Related Quality of Life in Children: The Roles of Age, Gender and Interpersonal Trust. International Journal of Environmental Research and Public Health, 2022;19(22). https://doi.org/10.3390/ijerph192215408.
334. Bolton K., Kremer P., Rossthorn N., Moodie M., Gibbs L., Waters E., et al. The Effect of Gender and Age on the Association Between Weight Status and Health-Related Quality of Life in Australian Adolescents. BMC Public Health, 2014;14(1):898. https://doi.org/10.1186/1471-2458-14-898.
335. Gusi N., Perez-Sousa M. A., Gozalo-Delgado M., Olivares P. R. [Validity and reliability of the spanish EQ-5D-Y proxy version]. An Pediatr (Barc), 2014;81(4):212-9. https://doi.org/10.1016/j.anpedi.2013.11.028.
336. Australian Government Department of Health and Aged Care. About Health Technologies and Digital Health. 2022. Available from: https://www.health.gov.au/topics/health-technologies-and-digital-health/about.
337. You X., Okunade A. A. Income and Technology as Drivers of Australian Healthcare Expenditures. Health Economics, 2017;26(7):853-862. https://doi.org/10.1002/hec.3403.
338. Australian Institute of Health and Welfare. Health System Overview. Canberrra: Australian Government; 2024. Available from: https://www.aihw.gov.au/reports/australias-health/health-system-overview.
339. Australian Institute of Health and Welfare. *Health Expenditure*. 2023 Available from: https://www.aihw.gov.au/reports/health-welfare-expenditure/health-expenditure#:~:text=Australia%20spent%20an%20estimated%20%24241.3,more%20than%20in%202020%E2%80%9321.
340. Eiser C., Morse R. Can Parents Rate Their Child's Health-Related Quality of Life? Results of a Systematic Review. Quality of Life Research, 2001;10(4):347-57. https://doi.org/10.1023/a:1012253723272.
341. Lancaster G. A. Statistical Issues in the Assessment of Health Outcomes in Children: A Methodological Review. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2009;172(4):707-727. https://doi.org/10.1111/j.1467-985X.2009.00590.x.
342. Bakker A., Streefkerk N., Bakker A., van Gorp M., van Litsenburg R., Grootenhuis M., et al. A Systematic Review of Health-Related Quality of Life in Children and Adolescents During Treatment for Cancer. EJC Paediatric Oncology, 2023;2. https://doi.org/10.1016/j.ejcped.2023.100134.
343. Theunissen N. C., Vogels T. G., Koopman H. M., Verrips G. H., Zwinderman K. A., Verloove-Vanhorick S. P., et al. The Proxy Problem: Child Report Versus Parent Report in Health-Related Quality of Life Research. Quality of Life Research, 1998;7(5):387-97. https://doi.org/10.1023/a:1008801802877.
344. Stevens K. Valuation of the Child Health Utility 9D Index. PharmacoEconomics, 2012;30(8):729-47. https://doi.org/10.2165/11599120-000000000-00000.
345. Mpundu-Kaambwa C., Bulamu N., Lines L., Chen G., Dalziel K., Devlin N., et al. A Systematic Review of International Guidance for Self-Report and Proxy Completion of

Child-Specific Utility Instruments. Value in Health, 2022;25(10):1791-1804. https://doi.org/10.1016/j.jval.2022.04.1723.

346. Blackmore A. M., Mulhern B., Norman R., Reddihough D., Choong C. S., Jacoby P., et al. How Well Does the EQ-5D-Y-5L Describe Children With Intellectual Disability?: "There's a Lot More to My Child Than That She Can't Wash or Dress Herself.". Value in Health, 2024;27(2):190-198. https://doi.org/10.1016/j.jval.2023.11.005.

347. Park E., Kwon M. Health-Related Internet Use by Children and Adolescents: Systematic Review. Journal of Medical Internet Research, 2018;20(4):e120. https://doi.org/10.2196/jmir.7731.

348. Logan D., Claar R., Scharff L. Social Desirability Response Bias and Self-Report of Psychological Distress in Pediatric Chronic Pain Patients. Pain, 2008;136:366-72. https://doi.org/10.1016/j.pain.2007.07.015.

PUBLICATIONS FROM THIS THESIS

PharmacoEconomics
https://doi.org/10.1007/s40273-022-01177-z

SYSTEMATIC REVIEW

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

Diana Khanna[1] · Jyoti Khadka[1,2] · Christine Mpundu-Kaambwa[1] · Kiri Lay[1] · Remo Russo[3,4] · Julie Ratcliffe[1] on behalf of The Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team

Accepted: 11 July 2022

Abstract

Objective The aim of this study was to examine the level of agreement between self- and proxy-reporting of health-related quality of life (HRQoL) in children (under 18 years of age) using generic preference-based measures.

Methods A systematic review of primary studies that reported agreement statistics for self and proxy assessments of overall and/or dimension-level paediatric HRQoL using generic preference-based measures was conducted. Where available, data on intraclass correlation coefficients (ICCs) were extracted to summarise overall agreement levels, and Cohen's kappa was used to describe agreement across domains. A meta-analysis was also performed to synthesise studies and estimate the level of agreement between self- and proxy-reported paediatric overall and domain-level HRQoL.

Results Of the 30 studies included, 25 reported inter-rater agreement for overall utilities, while 17 reported domain-specific agreement. Seven generic preference-based measures were identified as having been applied: Health Utilities Index (HUI) Mark 2 and 3, EQ-5D measures, Child Health Utility 9 Dimensions (CHU9D), and the Quality of Well-Being (QWB) scale. A total of 45 dyad samples were included, with a total pooled sample of 3084 children and 3300 proxies. Most of the identified studies reported a poor inter-rater agreement for the overall HRQoL using ICCs. In contrast to more observable HRQoL domains relating to physical health and functioning, the inter-rater agreement was low for psychosocial-related domains, e.g., 'emotion' and 'cognition' attributes of both HUI2 and HUI3, and 'feeling worried, sad, or unhappy' and 'having pain or discomfort' domains of the EQ-5D. Parents demonstrated a higher level of agreement with children relative to health professionals. Child self- and proxy-reports of HRQoL showed lower agreement in cancer-related studies than in non-cancer-related studies. The overall ICC from the meta-analysis was estimated to be 0.49 (95% confidence interval 0.34–0.61) with poor inter-rater agreement.

Conclusion This study provides evidence from a systematic review of studies reporting dyad assessments to demonstrate the discrepancies in inter-rater agreement between child and proxy reporting of overall and domain-level paediatric HRQoL using generic preference-based measures. Further research to drive the inclusion of children in self-reporting their own HRQoL wherever possible and limiting the reliance on proxy reporting of children's HRQoL is warranted.

✉ Diana Khanna
khan0420@flinders.edu.au

Extended author information available on the last page of the article

Key Points for Decision Makers

The application of child-specific preference-based measures enables the calculation of utilities for cost utility analysis of health technologies targeted for paediatric populations.

Proxy reports (e.g., parent/guardian or a health professional), used in lieu of child self-reports in circumstances when self-reports are not feasible, can often diverge from the child's assessment of their own HRQoL.

This review examined the agreement between the child self- and proxy-reported overall and domain-level HRQoL using generic preference-based measures.

In general, the inter-rater agreement was poor for overall utilities across the measure/s applied and/or the context of the application. In addition, the agreement between children and proxy respondents within the domains of the respective measures was lower for psychosocial-related attributes compared with physical attributes.

1 Introduction

Evidence from economic evaluation is increasingly being utilised by regulatory bodies such as the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia and the National Institute for Health and Care Excellence (NICE) in parts of the UK to evaluate the cost effectiveness of health technologies targeted for paediatric populations [1]. PBAC, for example, considers evidence derived from measures of health-related quality of life (HRQoL) when recommending medicines eligible for government subsidies under the Pharmaceutical Benefits Scheme (PBS) [2]. Economic evaluations involving cost-utility analysis (CUA) have become the most prevalent approach for providing health economic evidence to assess the cost effectiveness of new health technologies for adult and paediatric populations. Within CUA, outcomes are most typically presented as quality-adjusted life-years (QALYs). The QALY combines 'utility' indexed on a 0–1 scale (where 0 is equivalent to being dead and 1 is equivalent to full health) and length of life into a single generic measure of health outcome, thereby facilitating comparisons of the health gains generated from alternative interventions [3, 4].

The application of child-specific preference-based measures enables the derivation of utilities (preference weights) for incorporating into CUA of health technologies targeted for paediatric populations [5]. In a previous review of validated measures, Chen and Ratcliffe identified nine generic preference-based measures that have been applied to measure and value HRQoL in children and adolescents: Quality of Well-Being Scale (QWB), Health Utilities Index Mark 2 (HUI2), Health Utilities Index Mark 3 (HUI3), Sixteen-dimensional measure of health-related quality of life (HRQoL) [16D], Seventeen-dimensional measure of HRQoL (17D), Assessment of Quality of Life 6-Dimension (AQoL-6D) Adolescent, Child Health Utility 9 Dimensions (CHU9D), EQ-5D Youth version (EQ-5D-Y) and Adolescent Health Utility Measure (AHUM). Preference-based measures comprise two main components: a descriptive system for measuring HRQoL, and a preference-based scoring algorithm for generating utilities. The descriptive systems of the identified nine generic preference-based measures that have been applied to measure and value HRQoL in children and adolescents differ in the content, type, absolute number of HRQoL dimensions (domains/attributes) and/or response levels included. Similarly, the preference weighted scoring algorithms (value sets) for these measures also differ according to the methods used to generate the value set, e.g., time-trade off (TTO), standard gamble (SG) or discrete choice experiments (DCEs) and the population from whom the value set was derived, e.g. adults or young people [3].

Ideally, the individual themselves should be the principal source of information about their own HRQoL [1]; however, self-assessment of HRQoL is challenging in the paediatric population. According to the Professional Society for Health Economics and Outcomes Research (ISPOR) Good Research Practices Patient-Reported Outcomes (PRO) Task Force Report, there is insufficient evidence to determine whether self-reporting of HRQoL by children under 8 years of age is reliable or valid [6]. Furthermore, older children with conditions associated with neurodevelopmental delays may be unable to self-assess their own HRQoL due to limited cognitive abilities. Such circumstances may require relying on an adult proxy such as a parent/guardian or a health professional to assess the child's HRQoL [7].

It is well-documented that proxy assessments of HRQoL in any population group tend to differ from self-assessments, with proxy assessors typically reporting lower HRQoL than the person themselves [1, 6, 8, 9]. Two previous systematic reviews by Khadka et al. and Jiang et al. of child self- and proxy-reported child utilities found that utilities tended to differ, with proxies often underestimating the child's HRQoL [10, 11]. In child populations, there is some evidence to indicate that proxy assessment of the child's HRQOL may be influenced by external factors, e.g. mother's assessment of the child's HRQoL may be influenced by their own HRQoL [12].

In their systematic review, Jiang et al. examined the difference in self- and proxy-reported utilities [11]. Child

HRQoL ratings obtained by two different observers, the child self and the proxy, are likely to differ owing to the differences in their perspectives. Therefore, it is also important to determine the extent to which the two raters agree or assign the same rating for an item being measured, i.e., to report inter-rater agreement measures that estimate the strength of agreement between raters [13, 14]. This systematic review sought to add to the existing evidence by focusing on reported measures of agreement in child and proxy assessments of paediatric HRQoL using established generic preference-based measures, highlighting individual domain-level differences in agreement, in addition to overall utilities. This study also presents the methods and findings from a meta-analysis of reported agreement statistics to provide an overall indication of the extent of agreement in child self and proxy assessments of paediatric HRQoL according to the available evidence.

2 Methods

2.1 Search Strategy

The literature search strategy was adapted from a previous study undertaken by Khadka et al., and the search keywords were reproduced [10]. The time frame covered by the previous search was from inception to 30 July 2017. To reflect the latest publications during the 4-year period since the initial search undertaken by Khadka and colleagues, this review incorporated peer-reviewed articles published in electronic journals between 30 June 2017 and 19 May 2021. The online databases searched included PubMed, The Cochrane Library, Web of Science, EconLit, Embase, PsycINFO and CINAHL (via EBSCOhost). Key words such as ‘utility’, ‘quality-adjusted life years’, ‘children’, ‘adolescents’, and ‘preference-based measure of HRQoL’ as well as related Medical Subject Headings (MeSH) terms were used for the systematic literature search. A detailed account of the search terms and the strategy is presented in Appendix 1 (see electronic supplementary material [ESM]). The identified studies were screened using the web-based systematic review software Covidence [15]. This review is registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration number CRD42021256815). The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement guidelines were used for reporting this review (Appendix 2, see ESM) [16].

2.2 Inclusion and Exclusion Criteria

All studies published in English with full-text availability were included. Eligible studies included primary studies applying generic preference-based measures to derive health utilities amenable to QALY calculations in a paediatric population as assessed by the child (from hereon, *child* or *children* refer to all school-age children and adolescents, i.e., between 5 and 18 years of age unless stated otherwise) and proxy dyads. Inclusion criteria were studies reporting the agreement level for overall and/or domain-level paediatric HRQoL by both children and the proxies reporting on behalf of the children. Those studies that reported the paediatric health state utilities as assessed by child (self) and proxy respondents but did not include the agreement statistics were excluded. Additionally, as this systematic review focused on studies applying generic preference-based measures to derive health utilities, primary studies conducted among the paediatric populations were excluded if the utilities were obtained (1) directly using SG, TTO and VAS, or (2) indirectly using condition-specific (as opposed to generic) HRQoL measures.

2.3 Article Screening

Article screening was carried out in three steps. In the first step, two independent reviewers (DK and KL) screened the titles and abstracts based on the inclusion and exclusion criteria. Records with conflicting decisions were deferred to a third reviewer to reach a consensus. Articles selected at the screening stage were then included for a full-text review in the second step. The same two reviewers reviewed all the articles included in this stage. Simultaneously, two other reviewers (JK and CMK) independently assessed 10% of the articles in total to confirm the decisions of the former pair of reviewers. Following a discussion with the initial reviewing pair and the other reviewers (JR, JK, CMK) to reach a consensus, full-text articles that met the criteria were included. In the final step of this process, all eligible articles were subsequently consolidated and information relevant to the study was extracted.

2.4 Data Extraction

Data extraction was performed by the first author (DK). Each article was assessed to retrieve the following information: bibliographic details, geographic setting, study design, health state experienced, the generic preference-based measure used, target sample size, age range of the children included, sample gender composition, proxy type and sample size, mode of administration for both individuals in the dyad, statistical test(s) that report the overall and/or domain level of agreement between self- and proxy-reported HRQoL, and any reported methodological concerns. A Microsoft Excel (Version 2019; Microsoft Corporation, Redmond, WA, USA) database was used to enter and store the extracted data.

2.4.1 Extraction and Interpretation of Agreement Statistics

Inter-rater agreement is the degree to which the assessments of two or more individuals (raters) are identical using the same measure and assessing the same subject. There are multiple methods to measure inter-rater agreement based on the type of variable (continuous or categorical) and the number of raters. Agreement measures such as the intraclass correlation coefficient (ICC), Cohen's kappa (κ), Bland–Altman plots, percentage agreement and Gwet's agreement coefficient (AC1) assess the degree to which the assessments by the individual raters are identical or in agreement based on the type of data (e.g., nominal or continuous) [14, 17]. Correlation coefficients, also commonly reported to indicate agreement, determine the linear relationship between two continuous variables (Pearson's product-moment correlation or Pearson's r) or two ranked variables (Spearman's rho) [18].

It is important to note that in statistical analysis, correlation coefficients (e.g., Pearson's r) are considered as suboptimal measures of inter-rater agreement. They only provide a measure of the strength of a linear association between scores by raters and may indicate strong correlations even in the presence of a significant difference between the HRQoL assessments if the scores by both raters vary similarly. As a result, correlation coefficients may over- or underestimate the true level of agreement and inaccurately reflect the degree of agreement between raters [14, 18–20]. Inter-rater agreement is also often estimated using the percentage agreement approach [20]. However, percentage agreement does not correct for the level of agreement resulting from a random decision made by the raters. Cohen's kappa accounts for this random agreement and is more robust [21]. Therefore, percentage agreement is excluded from this review as a measure of child and proxy agreement. Only two studies reported the inter-rater agreement using the Bland–Altman plot and were thus not included in this review.

Thus, in the present study, to examine the concordance in the paediatric HRQoL obtained by self and proxy reports, we treat the ICC and kappa values as primary evidence. In addition, the results of the correlation coefficients, both Pearson's r and Spearman's rho, are presented as supplementary evidence.

ICC's can take a value between 0 and 1, whereas kappa and correlation coefficient statistics range from -1 to 1. Values for ICCs < 0.5 indicate poor agreement between raters, whereas values between 0.5 and 0.75, 0.75 and 0.9, and > 0.9 indicate moderate, good, and excellent agreement, respectively [22]. Spearman's correlation coefficients with a value < 0.20 represent no correlation, values between 0.20 and 0.35 represent weak correlation, values between 0.35 and 0.50 represent moderate correlation, and values ≥ 0.50 represent strong correlation [23]. Pearson's r coefficients are interpreted using Cohen's conventions. The correlation is small if the coefficient is 0.30 or less, medium if it is 0.50 or less, and large if it is > 0.50 [24]. Cohen's kappa and Gwet's AC1 have similarly defined thresholds, with classifications defined as slight (poor), fair, moderate, substantial (good) and almost perfect (very good) correlation for values ≤ 0.2, 0.4, 0.6, 0.8 and 1, respectively [17, 25].

2.5 Data Synthesis and Analysis

The estimates of the agreement level between child self- and proxy-reported HRQoL were described using a textual approach in the form of a narrative synthesis [26, 27]. Several studies did not report the mean age of participating children in the dyad, and hence only the age range was analysed. Studies that included children with cancer along with other chronic illnesses were identified as non-cancer-related studies. Caregivers reporting as proxies on behalf of children were grouped under parents. When the type of correlation was not mentioned in the study, it was assumed to be Pearson's r.

A meta-analysis was performed on a subset of the studies to synthesise the quantitative information and estimate the overall and domain-level agreement between child self- and proxy-reported HRQoL. To obtain an average estimate of inter-rater agreement, we synthesised the ICCs for overall utilities as they are reported on a continuous scale. Similarly, considering the ordinal nature of the responses within the attributes, kappa statistic was used to estimate the domain-level inter-rater agreement. Studies reporting only the correlation coefficients were excluded from the meta-analysis.

The meta-analysis was conducted using Stata 16.1 (Stata Corp LLC, College Station, TX, USA). Since the assumption of homogeneity is not reasonable for the present data due to the diverse nature of the target samples in consideration, we used a random-effects model to allow for between-study variability in effect sizes. The weights were estimated using a restricted maximum likelihood (REML) method [28]. A Fisher's z-transformation was applied to obtain an approximately normal sampling distribution in order to calculate the 95% confidence intervals (CIs) for each ICC for the overall utilities. The z-scores were then transformed back into correlations for ease of interpretation [29].

For the domain level meta-analysis, the standard errors (se) for kappa values ($\hat{\kappa}$) were calculated using the following formula (Eq. 1):

$$\mathrm{se}_{\kappa} = \sqrt{\frac{p(1-p)}{n(1-p_{c})^{2}}}, \tag{1}$$

where p is the observed percentage agreement, n is the number of rater pairs and p_{c} is the agreement expected by chance.

However, since no study reported the values for p_c, but did report p and $\hat{\kappa}$, p_c was calculated as shown in Eq. (2) [30]:

$$p_c = \sqrt{\frac{p - \hat{\kappa}}{1 - \hat{\kappa}}}. \tag{2}$$

A forest plot was used to depict the results of the meta-analysis (overall agreement). Heterogeneity was assessed using a forest plot as well as Cochran's test of homogeneity (Q statistic) and the I^2 statistic. Each sample was considered unique if any of the following variables relevant to the analysis were unique: type of proxy, measure, health condition, or age group composition (i.e., if children below 8 years of age were included in the sample). An exploratory meta-analysis (assuming a random-effects model) was conducted to estimate the moderation by these variables. A random-effect meta-regression was used to supplement the findings of the meta-analysis, as the studies were not considered sufficiently similar for a fixed-effects model [31]. The sample was also considered to be unique if the same sample was examined in a different time period for longitudinal studies. Publication bias was evaluated using funnel plots and a regression-based funnel plot asymmetry test.

2.6 Risk of Bias and Quality Assessment

Two independent reviewers (DK and JK) appraised the quality and suitability of the included studies. The overall reporting quality score was calculated using a checklist for quantitative studies as given by Kmet et al., and was used to assess the risk of bias [32]. From each of the selected articles that met the inclusion criteria, we extracted information for 14 quality indicator variables (details provided in ESM Appendix 3). Two points were assigned to each of these variables if they were appropriately reported in the article, one if the item was incompletely reported, and none if not reported at all. The sum of all the points indicated the overall reporting quality score of the article, with 28 being the maximum. The summary scores were rescaled between 0 and 1, with 1 denoting the highest quality. If the item was not applicable to a particular study, scores were adjusted by excluding the total possible scores of those items from the summary score. The minimum threshold for inclusion of studies based on quality scores was set at 0.6. The results of a sensitivity analysis carried out using the criteria by Papaioannou and colleagues to confirm the conclusions from the former appraisal are reported in Appendix 4 (see ESM) [33].

3 Results

3.1 Search Results

A PRISMA flow diagram illustrates the selection process (Fig. 1). An extensive literature search of seven databases was conducted using the search strategy described above. 43,522 records published between 30 June 2017 and 19 May 2021 were identified and were subsequently imported into Covidence; 19,309 records were deduplicated by Covidence, leaving 24,213 records for title and abstract screening. Of these, the vast majority (23,547) were excluded. Reasons for exclusion were (1) non-primary studies; (2) non-paediatric target population; (3) no health state utilities reported; (4) inaccessible articles; and (5) English was not the main language of publication. Subsequently, 666 records were included in the full-text review stage. At this stage, in addition to the previously specified exclusion criteria, studies were excluded if agreement statistics between the child self- and proxy-reported health state utilities and/or at domain level were not reported. In total, 30[1] studies fully met the inclusion criteria and were thus included in the final review.

3.2 Main Characteristics of the Studies

Table 1 presents an overview of the studies included in this systematic review. All the studies appraised for quality of reporting were of high quality, scoring 0.7 and over. The following study designs were employed: cross-sectional (83%), longitudinal (23%), and case-control (3%). HRQoL measures applied to obtain health state utilities either independently or in combination with other measures included the HUI3 (57%), EQ-5D measures (EQ-5D-Y-3L, EQ-5D-Y-5L, EQ-5D-3L, and the EQ VAS; 37%), HUI2 (33%), CHU9D (7%), and the QWB scale (3%). Cancer or history of cancer was the most common condition for which HRQoL was assessed (27%), predominantly blood and brain malignancies. Some studies (30%) also included children from the general population as the target sample or as the comparator/control group. The proxy respondent was exclusively a parent (mother, father, or a caregiver) in most of the identified studies (83%). Several studies (17%) used health professionals (nurses, physicians, and physiotherapists) or teachers as proxies, together with parents. The only exception was the study by Barr et al., which used only nurses and

---

[1] The two papers by Glaser et al., i.e. 'Standardized quantitative assessment of brain tumor survivors treated within clinical trials in childhood' [36] and 'Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the U.K.' [37], were published in two different journals but used the same sample to report different results. To prevent double counting, these two papers were considered as one.

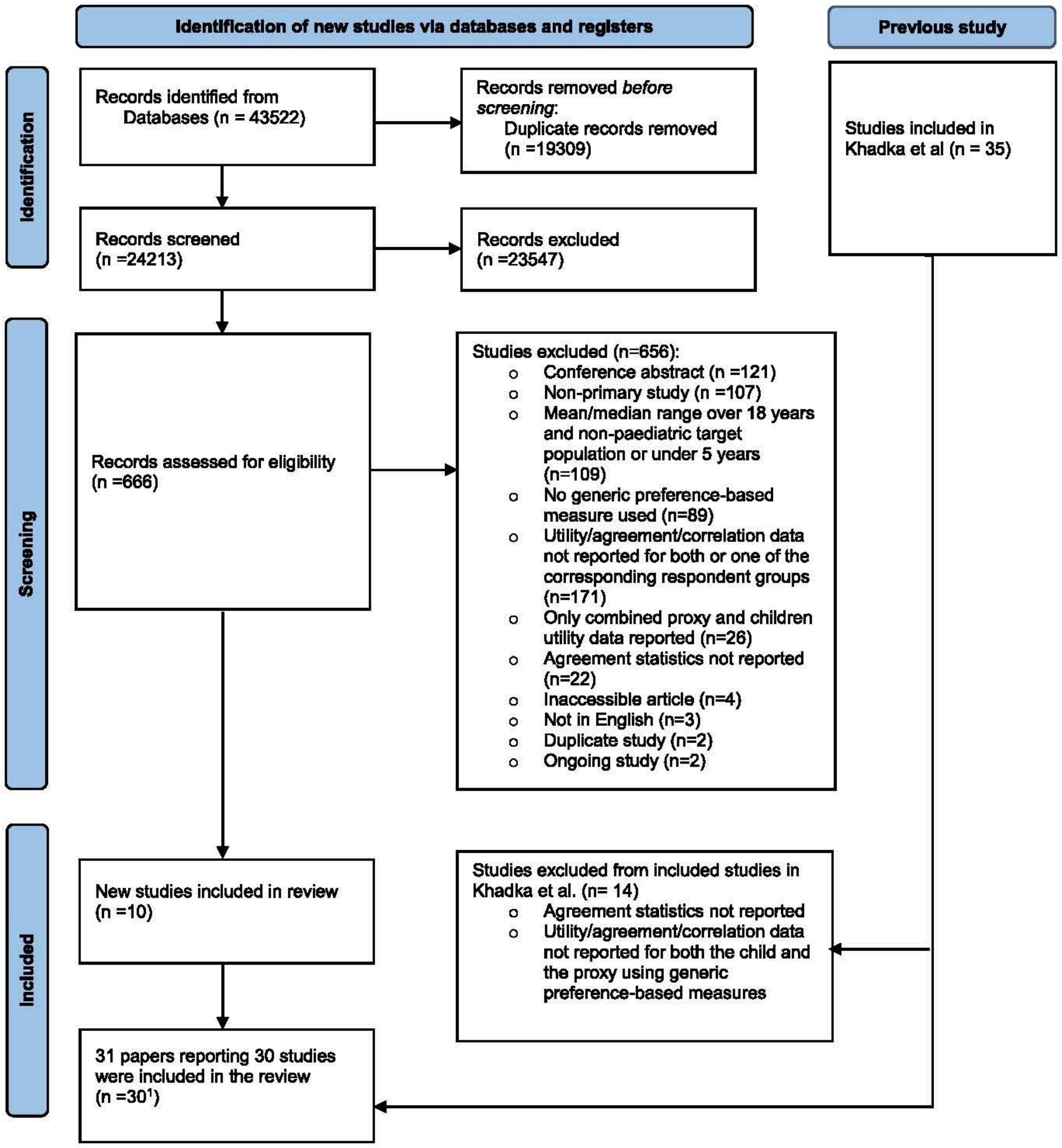


Fig. 1 Literature search flow diagram using the PRISMA checklist. *PRISMA* Preferred Reporting Items for Systematic Reviews and Meta-Analyses. [1]Thirty studies were included in the final review. The two papers by Glaser et al., i.e. 'Standardized quantitative assessment of brain tumor survivors treated within clinical trials in childhood' [36] and 'Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the UK' [37], were published in two different journals but used the same sample to report different results. To prevent double counting, these two papers were considered as one

physicians for proxy-reported utilities using HUI2 and 3 in cancer survivors [34]. Each study administered the proxy version of the measures adopting a proxy/proxy perspective, except one [35], which used a proxy/patient perspective (asking the proxy to rate the child's HRQoL from the child's perspective).

The measures were either administered by a trained interviewer (50%) or self-completed by the children (47%).

Table 2 (continued)

Author Country Year	Health state experienced	Mean/median age (range) of the child in the dyad (where available) or study	Child sample (male %) included in the dyad (where available) or study	Proxy type (*n*)	Measure	Administration mode child	Quality score
Zhou et al. [54] China 2021	Haematological malignancies	10.5 (8.0–17.0)	96 (64.6)	Caregiver (96)	EQ-5D-3L-Y/VAS, EQ-5D-5L-Y	Interviewer administered	0.95

HRQoL health-related quality of life, *HUI2* Health Utilities Index Mark 2, *HUI3* Health Utilities Index Mark 3, *VAS* visual analogue scale, *CNS* central nervous system, *ALL* acute lymphoblastic leukaemia

3.3.1 Inter-Rater Agreement Based on the Type of Measure

HUI2 and 3 The inter-rater agreement between children and proxies for nine studies as indicated by the ICCs was poor for overall utilities [34, 36, 37, 42, 43, 48–50, 55, 56]. The overall ICC for HUI2 was slightly higher than that of HUI3. In contrast to HUI2, which showed good to excellent agreement for the overall utilities for one-quarter of the samples in the studies, the agreement using HUI3 was moderate at best. The correlation coefficients obtained from 10 studies indicated moderate associations between child self and proxy reports [36, 37, 44, 45, 55–60, 63].

Across the HUI2 attributes of ‘emotion’, ‘cognition’ and ‘pain’, the overall kappa values indicated fair agreement for those domains with a moderate agreement for ‘sensation’. Overall, the kappa values suggested a substantial agreement for ‘mobility’, the highest level of agreement among all attributes, and a moderate agreement for ‘self-care’ between the child/proxy dyad [34, 36, 37, 39]. The lowest kappa values were reported for ‘emotion’ and ‘cognition’ in the assessment of HRQoL by children and proxies. For the ‘pain’ attribute, both slight and substantial levels of agreement were reported equally among the samples.

For HUI3, the overall agreement using kappa values was fair for ‘cognition’, ‘emotion’, ‘speech’ and ‘pain’; moderate for ‘hearing’, ‘dexterity’ and ‘ambulation’; and substantial for ‘vision’ [36–39, 41]. Similar to HUI2, the lowest agreement between children and proxies for HUI3 attributes was reported for ‘emotion’ and ‘cognition’. In contrast, high kappa values were frequently reported for the attributes of ‘vision’, ‘ambulation’ and ‘dexterity’, with the agreement level ranging from substantial to almost perfect.

The ICC values demonstrated a poor agreement for subjective domains (‘emotion’, ‘cognition’, and ‘pain’) with some even reporting negative values. The agreement was between good to moderate for the observable domains of sensation, mobility, self-care, vision, hearing, and dexterity, with the notable exception of ‘ambulation’ and ‘speech’, which showed poor inter-rater agreement [42, 43, 49]. The agreement within the ‘ambulation’ and ‘speech’ attributes was moderate only in one instance between cancer survivors and their parents [43].

EQ-5D measures and the EQ VAS None of the studies reported the ICCs for the overall utilities or the summary scores using EQ-5D measures. Of the six studies reporting the ICCs for the EQ VAS scores, the majority showed poor agreement between child/proxy dyads [46, 47, 51–54]. However, an improvement in the inter-rater agreement was noted from baseline to follow-up [51, 54]. Kappa statistics reported for five studies indicated, on average, fair agreement between children and parents for all domains of EQ-5D [35, 40, 46, 47, 51]. The agreement was the lowest for the ‘feeling worried, sad, or unhappy’ and ‘having pain or discomfort’

Table 2 Details of the cancer studies that reported dyad self and proxy HRQoL using preference-based quality-of-life instruments

Author Country Year	Health state experienced	Mean/median age (range) of the child in the dyad (where available) or study	Child sample (male %) included in the dyad (where available) or study	Proxy type (*n*)	Measure	Administration mode child	Quality score
Barr et al. [34] Canada 1999	Cancer survivors: CNS tumours	13.5 (9.5–17.9)	15 (46.3)	Nurses (15), physicians (12)	HUI2/3	Self-administered	0.7
Glaser et al. [36, 37] UK 1999	Cancer survivors: CNS tumours	10.5 (6.0–16.0)	28	Physiotherapist (30), parents (30), physicians (27)	HUI2/3	Self and Interviewer administered	0.85
Sung et al. [56] Canada 2004	Cancer patients: rheumatic diseases, haemophilia, conditions requiring bone marrow transplant	13.7 (12.0–18.0)	22 (55)	Parents	HUI2	Self-administered	0.9
Fu et al. [42] El Salvador, Honduras, Nicaragua, Panama 2006	Cancer survivors: leukaemia, lymphoma, renal tumours, germ cell tumours, retinoblastoma, malignant bone tumours, CNS tumours, sympathetic nervous system tumours, soft tissue sarcomas, carcinomas, others	12.8 (5–25.8)	211 (52.6)	Parents (180), physicians (201)	HUI2/3	Interviewer administered	0.8
Banks et al. [48] Canada 2008	Cancer including leukaemia, lymphoma, and brain tumour	9.5 (10.0–18.0)	11 (65)	Parents (22)	HUI2/3	Self-administered	0.85
Fluchel et al. [43] Uruguay 2008	Cancer survivors: ALL, brain tumours, Wilms tumour, retinoblastoma, Hodgkin disease, non-Hodgkin lymphoma, acute myeloid leukaemia, rhabdomyosarcoma, neuroblastoma, Ewing sarcoma, ovarian sarcoma, osteogenic sarcoma	13.6 (7.0–28.0)	95 (49.5)	Parents (95)	HUI3	Interviewer administered	0.95
	General health (control)	12.2 (8.0–17.0)	96 (33.3)	Parents (91)	HUI3	Interviewer administered	
Penn et al. [59] UK 2011	Cancer patients: brain tumour	12.4 (8.0–17.6)	29 (48.3)	Parents (29)	HUI3	Interviewer administered	0.85
	General health (control)	10.7 (8.0–18.9)	32 (50)	Parents (32)	HUI3	Interviewer administered	

Table 1 An overview of the included studies

Description	No. of studies
Total studies included	30
Child-specific preference-based measures used	
HUI2	10
HUI3	17
EQ-5D-Youth, EQ-5D and EQ VAS	11
CHU9D	2
QWB	1
Health conditioned studied	
Cancer or history of cancer	8
Other health conditions (including general health)	22
Child proxy pairs (with some studies using more than one proxy type)	
Child/parent	29
Child/health professionals (nurses, physicians, physiotherapists) or teachers	5
Self-mode of administration for child in the age range	
6–7 years	3
8 years and above	11
Interviewer mode of administration for child in the age range	
6–7 years	6
8 years and above	10
Level of agreement statistics reported	
For overall utilities	25
For attribute-level utilities	17

HUI2 Health Utilities Index Mark 2, *HUI3* Health Utilities Index Mark 3, *VAS* visual analogue scale, *CHU9D* Child Health Utility 9 Dimensions, *QWB* Quality of Well-Being scale

One study used both an interviewer administration mode for children below 8 years of age and self-completion for the older children [36, 37]. The majority of the studies (83%) reported the inter-rater agreement for overall utilities. Five studies only reported the domain-level agreement [35, 38–41]. When reported, ICCs were slightly more commonly represented (60%) than correlation coefficients in measuring the overall child/proxy agreement level. Cohen's kappa (59%) was the most frequently used measure of agreement at the attribute level, followed by ICC (18%) and Gwet's AC1 (12%).

A summary of the included studies is presented in Tables 2 and 3 grouped into cancer- and non-cancer-related conditions, respectively. All the included studies were published between 1994 and 2021 and used primary data to obtain child health state utilities by employing generic preference-based measures. Majority of the studies were published in North America (USA and Canada; 33%) and Europe (UK, Spain, Netherlands, and Germany; 33%), followed by Asia (Thailand, Japan, Hong Kong, and China; 17%). Forty-five unique dyad samples based on the proxy type were included in the studies, with a total pooled sample of 3084 children and 3300 proxies. The age range for children in the included studies was between 5 and 18 years. Eight studies reported children younger than 8 years of age completing a self-report questionnaire either independently or with some assistance [35–37, 40, 42–46].

3.3 Proxy/Child Agreement

Table 4 presents a summary of reported agreement statistics for overall utilities using ICCs or correlation coefficients, i.e., Pearson's r and Spearman's rho. The studies used all the identified measures except for the EQ-5D-Y-5L, and employed both caregivers and health professionals as proxies. The sample size of the dyad ranged from 11 [45] to 654 [47]. From a total of 26 studies (58 samples), 12 studies reported only the ICCs [34, 42, 43, 46–54], and three studies reported ICCs alongside the correlation coefficients [36, 37, 55, 56]. Six studies reported only Spearman's rho [45, 57–61], whereas four studies reported only Pearson's r [44, 62–64]. Details of the included studies reporting the domain-level agreement statistics are presented in Table 5. The domain-level agreement was reported for 17 studies (40 samples), of which 10 studies used Cohen's kappa [34–41, 46, 47, 51], three studies used ICC [42, 43, 49], and two used Gwet's AC1 [53, 54]. No study reported the domain-level agreement for the CHU9D and QWB measures.

Table 3 Details of the studies with health conditions other than cancer that reported dyad self and proxy HRQoL using preference-based quality-of-life instruments

Author Country Year	Health state experienced	Mean/median age (range) of the child in the dyad (where available) or study	Child sample (male %) included in the dyad (where available) or study	Proxy type (*n*)	Measure (proxy measure)	Administration mode child	Quality score
Czyzewski et al. [62] USA 1994	Cystic fibrosis	(12–17.9)	55	Parents (199)	QWB	Self-administered	0.8
Verrips et al. [38] Netherlands 2001	Very low birth weight (VLBW): Mail Telephone Face-to-face Repeat mail	14.2 (14.0) 14.3 (14.0) 14.3 (14.0) 14.2 (14.0)	486 (49) 100 (54) 103 (51) 203 (52)	Parents (481) Parents (100) Parents (103) Parents (203)	HUI3 HUI3 HUI3 HUI3	Self-administered Self-administered Self-administered Self-administered	0.85
Brunner et al. [55] Canada 2003	Musculoskeletal disorders	9 (8.0–18.0)	55	Parents (68)	HUI3	Interviewer administered	0.8
Jelsma and Ramma [35] South Africa 2010	Children with functional impairment General health (control)	(7.0–12.0) (7.0–12.0)	61 (74) 567 (45)	Mother (57) Mother (530)	EQ-5D-Y/VAS (EQ-5D-Y Proxy 2) EQ-5D-Y/VAS	Self-administered Self-administered	0.85
Belfort et al. [57] Germany 2016	Overweight or obese General health (control)	10.3 (8.0–17.0) 11.5 (8.0–18.0)	76 (52.6)	Parents (63)	HUI3	Interviewer administered	0.95
Lee et al. [58] USA 2011	Type 1 diabetes mellitus. Complications: hypertension, hypercholesterolaemia, cardiovascular disease, renal disease, neurological disease, retinopathy	13.7 (8.0–18.0)	231 (48.5)	Parents (223)	HUI3	Interviewer administered	0.95
Morrow et al. [39] Australia 2012	Chronic illness: any cancer, cystic fibrosis, type 1 diabetes, cerebral palsy (GMFCS V), any chronic neurological condition, liver transplant, inflammatory bowel disease, chronic kidney disease, autism	12.2 (12.0–18.0)	69 (54.2)	Parents (129) Physicians (34)	HUI2/3	Self-administered	0.85
Rhodes et al. [60] USA 2012	Obesity; type 2 diabetes mellitus; prediabetes; insulin resistance	15.5 (12.0–18.0)	108	Parents (108)	HUI3	Interviewer administered	0.85

Table 3 (continued)

Author Country Year	Health state experienced	Mean/median age (range) of the child in the dyad (where available) or study	Child sample (male %) included in the dyad (where available) or study	Proxy type (*n*)	Measure (proxy measure)	Administration mode child	Quality score
Rogers et al. [64] Netherlands 2019	Dental caries	11 (11.0)	486 (48)	Parents (486)	CHU9D (NL)	Self-administered	1
Shiroiwa et al. [47] Japan 2019	General health	11 (8.0–15.0)	654 (50)	Parents (654)	EQ-5D-Y/VAS	Self-administered	0.9
Sinlapamongkolkul et al. [61] Thailand 2020	Thalassaemia	9.1 (8.0–18.0)	85 (54)	Parents (85)	EQ VAS	Self and interviewer administered	0.95
Lin et al. [53] Hong Kong 2020	Adolescent/Juvenile idiopathic sclerosis (AIS/JIS)	14 (10.0–12.0)	125 (9.4)	Caregiver (125)	EQ-5D-Y/VAS	Self-administered	0.95

HRQoL health-related quality of life, *HUI2* Health Utilities Index Mark 2, *HUI3* Health Utilities Index Mark 3, *EQ-5D-Y* EQ-5D Youth version, *VAS* visual analogue scale, *CHU9D* Child Health Utility 9 Dimensions, *CHU9D (NL)* CHU9D Dutch version, *QWB* Quality of Well-Being scale, *GMFCS* Gross Motor Function Classification System, *AOM* acute otitis media, *MMR* mild mental retardation, *SMR* severe mental retardation, *MR* mental retardation, *VLBW* very low birth weight, *VP* very preterm, *AIS* adolescent idiopathic scoliosis, *JIS* juvenile idiopathic scoliosis

domains, followed by, 'doing usual activities', 'looking after myself' and the highest for 'walking about'.

The inter-rater agreement between children and proxies within the EQ-5D domains using Gwet's AC1 ranged from moderate to very good [53, 54]. Children and adolescents with haematological malignancies were assessed using both 3L and 5L versions of the EQ-5D-Y in the study by Zhou et al. They found moderate to good agreement between the self- and caregiver-reported HRQoL for the five dimensions. The agreement improved from baseline to follow-up for all except the 'having pain or discomfort' domain in the 3L version and the 'walking about' and 'looking after myself' domains in the 5L version. However, no significant difference between the 3L and 5L versions was reported [54]. Among children with Adolescent/Juvenile idiopathic scoliosis (AIS/JIS), Lin et al. showed very good agreement with the caregivers in all domains except the 'having pain or discomfort' and 'feeling worried, sad, or unhappy' domains [53].

CHU9D and QWB The only study that reported the ICC using CHU9D showed moderate inter-rater agreement [50]. Using a large sample of 384 child/parent dyads, Rogers et al. reported a weak but significant correlation between the child self and proxy reports using CHU9D [64]. In their study, Czyzewski et al. reported a moderate correlation between the self- and proxy-reported utilities using QWB [62].

3.3.2 Inter-Rater Agreement Based on the Type of Proxy

Both types of proxies (parents and health professionals) showed poor inter-rater agreement, although parents showed higher agreement overall, regardless of measures and/or health conditions. All studies using health professionals as proxies assessed the HRQoL of children with cancer or child cancer survivors. Among these, Fluchel and colleagues used physicians and teachers as proxies for the children in the control group with no health condition [43]. A negative ICC (− 0.31, 95% CI − 0.22 to 0.262) was noted, indicating poor inter-rater agreement between the pair [43]. Only one study showed good to excellent agreement between cancer survivors and health professionals (nurses and physicians) using HUI2 [34]. Glaser and colleagues compared the inter-rater agreement between children with a history of cancer and their parents, physicians, and physiotherapists. Both the agreement (ICC) and correlation (Pearson's *r*) values were better for parents, closely followed by physiotherapists, and worst for physicians [36, 37]. In the study by Ungar et al., the authors found a poor inter-rater agreement when children and parents reported paediatric HRQoL separately using the HUI2 and 3; however, the agreement was found to be statistically significant and moderate using a consensus-based dyad approach [49].

Table 3 (continued)

Author Country Year	Health state experienced	Mean/median age (range) of the child in the dyad (where available) or study	Child sample (male %) included in the dyad (where available) or study	Proxy type (*n*)	Measure (proxy measure)	Administration mode child	Quality score
Ungar et al. [49] Canada 2012	Asthma	10.9 (8.0–17.0)	91 (55)	Parents (91)	HUI2/3	Interviewer administered	1
Kulpeng et al. [44] Thailand 2013	Common pneumococcal infections and sequelae: meningitis, bacteraemia, pneumonia, AOM, hearing loss, chronic lung disease, epilepsy, MMR, SMR, and MR combined with epilepsy	10 (7.0–14.0)	74	Caregiver (74)	HUI2/3, EQ-5D	Interviewer administered	0.85
Wolke et al. [41] Germany 2013	VLBW/VP	13 (13.0)	260 (52)	Parents (260)	HUI3	Self-administered	0.85
	General health (control)	13 (13.0)	282 (49)	Parents (282)	HUI3	Self-administered	
Gusi et al. [40] Spain 2014	General health	(6.0–17.0)	442	Mother (442) Father (266)	EQ-5D-Y	Self and interviewer administered	0.9
Sims-Williams [63] Uganda 2017	Open spina bifida; associated complications	(10.0–14.0)	66 (56)	Caregiver (66)	HUI3	Interviewer administered	0.95
Bharij et al. [50] USA 2017	Paediatric liver transplant recipients	13.6 (12.0–21.7)	108 (44.4)	Parents (108)	HUI2/3, CHU9D	Interviewer administered	0.9
Bray et al. [45] UK 2017	Long-term mobility impairment: cerebral palsy, hemiplegia, muscular dystrophy	(6.0–18.0)	13 (61.5)	Parents (13)	HUI2/3, EQ-5D-Y/VAS	Self-administered	0.9
Perez Sousa et al. [46] Spain 2017	Cerebral palsy	10.9 (6.0–17.0)	62 (65.4)	Mother (62)	EQ-5D-Y/VAS	Interviewer administered	0.95
Perez Sousa et al. [51] Spain 2018	Obesity: exercise	9.6 (6.0–14.0)	106 (55)	Parents (106)	EQ-5D-Y/VAS	Interviewer administered	0.95
	Obesity: control	8.7 (6.0–13.0)	45 (47)	Parents (45)	EQ-5D-Y/VAS	Interviewer administered	
van Summeren et al. [52] The Netherlands 2018	Functional constipation	10 (8.0–18.0)	56 (43)	Parents (56)	EQ VAS	Self-administered	0.95

Table 4 Details of the included studies of level of agreement by overall utilities between self- and proxy-reported HRQoL using preference-based quality-of-life instruments

Authors (intervention)	Measure	Proxy type	Sample size dyad	Correlation test	Correlation coefficient (p value)	95% CI
Barr et al. [34]	HUI2	Nurses	15	ICC	0.85	
		Physicians	12		0.95	
Glaser et al. [36]	HUI2	Physiotherapist	25	ICC	0.4	
		Parents	24		0.57	
		Physicians	19		0.15	
Glaser et al. [37]	HUI2	Physiotherapist	25	Pearson	0.54 (< 0.01)	
		Parents	24		0.59 (< 0.01)	
		Physicians	19		0.37 (0.12)	
Sung et al. [56]	HUI2	Parents	19	ICC	0.11 (0.3)	− 0.35, 0.53
				Spearman	0.14	− 0.34, 0.55
	HUI3	Parents	19	ICC	− 0.01	− 0.45, 0.44
				Spearman	0.11	0.35, 0.55
Fu et al. [42]	HUI2	Parents	120	ICC	0.389	0.227, 0.531
		Physicians	156		0.379	0.237, 0.506
	HUI3	Parents	156	ICC	0.433	0.297, 0.552
		Physicians	166		0.341	0.200, 0.469
Banks et al. [48]	HUI2	Parents	11	ICC	0.74	0.29, 0.92
	HUI3	Parents	11	ICC	0.42	− 0.21, 0.80
Fluchel et al. [43]	HUI3	Parents	92	ICC	0.3087	0.1125, 0.4818
		Physicians	91		0.066	− 0.1402, 0.2669
Fluchel et al. [43] (control)	HUI3	Physicians/teachers	89	ICC	− 0.3103	− 0.4857, − 0.1106
Penn et al. [59]	HUI3	Parents	21	Spearman	0.76 (< 0.001)	
Penn et al. [59] (control)	HUI3	Parents	22	Spearman	0.31	
Zhou et al. [54] (baseline)	EQ VAS	Caregiver	96	ICC	0.22	
Zhou et al. [54] (follow-up)	EQ VAS	Caregiver	96	Yes	0.556	
Czyzewski et al. [62]	QWB	Parents	55	Pearson	0.39	
Brunner et al. [55]	HUI3	Parents	45	ICC	0.43	
				Pearson	0.57	
Belfort et al. [57] (overall)	HUI3	Parents	63	Spearman	0.47 (0.0002)	
Lee et al. [58]	HUI3	Parents	223	Spearman	0.34	0.22, 0.45
Rhodes et al. [60]	HUI3	Parents	96	Spearman	0.24 (< 0.05)	
Ungar et al. [49]	HUI2	Parents	72	ICC	0.021	− 0.22, 0.262
	HUI3	Parents	75	ICC	0.169	− 0.070, 0.389
Ungar et al. [49] (Dyad)	HUI2	Parent with child	72	ICC	0.545 (< 0.0001)	0.360, 0.689
	HUI3	Parent with child	75	ICC	0.735 (< 0.0001)	0.611, 0.824
Kulpeng et al. [44]	HUI2	Caregiver	74	Pearson	0.58 (< 0.05)	
	HUI3	Caregiver	74	Pearson	0.67 (< 0.05)	
	EQ-5D	Caregiver	74	Pearson	0.77 (< 0.05)	
	EQ VAS	Caregiver	74	Pearson	0.5 (< 0.05)	
Sims-Williams et al. [63]	HUI3	Caregiver	62	Pearson	0.848	
Bharij et al. [50]	HUI2	Parents	61	ICC	0.9 (< 0.001)	
	HUI3	Parents	60	ICC	0.75 (< 0.001)	
	CHU9D	Parents	96	ICC	0.69 (< 0.001)	
Bray et al. [45]	HUI2	Parents	13	Spearman	0.728 (0.005)	
	HUI3	Parents	13	Spearman	0.842 (< 0.001)	
	EQ-5D-Y	Parents	11	Spearman	0.665 (0.026)	
	EQ VAS	Parents	13	Spearman	0.545 (0.054)	
Perez Sousa et al. [46]	EQ VAS	Mother	62	ICC	0.389 (0.029)	
		Father	62		0.581 (0.962)	
Perez Sousa et al. [51] (overall: baseline)	EQ VAS	Parents	151	ICC	0.5 (< 0.0001)	

Table 4 (continued)

Authors (intervention)	Measure	Proxy type	Sample size dyad	Correlation test	Correlation coefficient (p value)	95% CI
Perez Sousa et al. [51] (overall: follow-up: post treatment)	EQ VAS	Parents	151	ICC	0.7 (< 0.0001)	
van Summeren et al. [52]	EQ VAS	Parents	56	ICC	0.78	0.65, 0.87
Rogers et al. [64]	CHU9D	Parents	184	Pearson	0.156 (0.02)	
Rogers et al. [64] (control)	CHU9D	Parents	302	Pearson	0.183 (0.01)	
Rogers et al. [64] (overall)	CHU9D	Parents	386	Pearson	0.183 (< 0.001)	
Shiroiwa et al. [47]	EQ VAS	Parents	654	ICC	0.06	
Sinlapamongkolkul et al. [61]	EQ VAS	Caregiver	85	Spearman	0.334 (0.001)	
Lin et al. [53] (overall)	EQ VAS	Caregiver	125	Yes	0.29	

HRQoL health-related quality of life, *CI* confidence interval, *HUI2* Health Utilities Index Mark 2, *HUI3* Health Utilities Index Mark 3, *EQ-5D-Y* EQ-5D Youth version, *VAS* visual analogue scale, *CHU9D* Child Health Utility 9 Dimensions, *QWB* Quality of Well-Being scale, *ICC* intraclass correlation coefficient

The agreement between children and physiotherapists was generally low with the exception of one study where physiotherapists reported higher agreement than parents and physicians within the HUI3 attributes of 'vision' and 'speech' [36, 37]. Overall, physicians reported excellent agreement when assessing the functional attributes, e.g., 'mobility' and 'ambulation', whereas the subjective attributes of 'emotion', 'pain' and 'cognition' lacked sufficient agreement [36, 37, 39, 42, 43].

Parents followed a similar suit and reported slight to fair agreement within the 'emotion and 'cognition' attributes of HUI2 and 3. In the assessment of 'emotion', the only exception was reported in a study of children with very low birth weight by Wolke et al., which showed moderate agreement with the parents in the study population [41]. Moreover, father/child pairs agreed only slightly within all domains of EQ-5D-Y. In comparison, a better agreement was reported with mothers for the domains 'walking about', 'doing usual activity' and 'having pain or discomfort' [46].

3.3.3 Inter-Rater Agreement Based on the Type of Condition

Within the cancer-related studies, children with a history of cancer showed a much better agreement (ICC 0.44, 95% CI 0.26–0.62) with the proxy reports than those with active cancer (ICC 0.34, 95% CI 0.04–0.64). In addition to the higher agreement level, correlations observed were also large for the former cohort (0.52, 95% CI 0.31–0.68), whereas cancer patients showed weak associations (0.40, 95% CI − 0.15 to 0.76) with the proxy reports of their HRQoL. It is unclear if cancer-related studies showed an overall lower agreement between the child self and proxy reports of HRQoL, than studies with conditions other than cancer. For instance, in a longitudinal study of cancer patients, Penn and colleagues found strong associations between the HUI3 generated overall utilities as reported by children and proxies in the study population, but weak correlations for those in the control group [59]. Conditions such as respiratory (asthma) and musculoskeletal diseases assessed using HUI2 and 3 showed poor inter-rater agreement between child self- and proxy-reported utilities [49, 55]. Using the EQ VAS, van Summeren and colleagues found good inter-rater agreement in children with functional constipation [52]. Additionally, in a longitudinal study of children with obesity, the agreement between children and parents for EQ VAS scores was found to be moderate at baseline and at follow-up [51]. Strong associations (Spearman's rho) were noted between the utilities reported by children with cerebral palsy, hemiplegia, and/or muscular dystrophy and their parents using both EQ-5D-Y and EQ VAS [45], while the correlation between children with thalassaemia and their caregivers using the EQ VAS was weak [61]. Kulpeng et al. also indicated a large correlation (Pearson's r) between self- and proxy-derived utilities using EQ-5D and EQ VAS in children with severe childhood infections [44].

The agreement and correlation between child self- and proxy-reported overall HRQoL observed between healthy children and proxies, including parents, physicians, and teachers, was, on average, low [43, 47]; however, evidence for the domain-level agreement was inconsistent. Kappa values in the study by Wolke et al., suggested moderate to almost perfect agreement between children with no specific health condition and parents across all HUI3 attributes [41]. In contrast, another study observed perfect agreement only within the 'hearing', 'ambulation', and 'dexterity' attributes, while the remaining attributes showed poor or no agreement [43]. Notably, this study used physicians/teachers as proxies rather than parents, which could potentially account for the contrasting findings. Similarly, one of the two studies using the EQ-5D-Y reported a moderate to almost perfect agreement across all domains except 'having pain or

Table 5 Details of the included studies' level of agreement by domains (attributes) between self- and proxy-reported HRQoL using preference-based quality-of-life instruments

Authors (intervention)	Measure	Proxy type	Attribute	Statistic reported	Agreement statistic (*p* value)	95% CI
Barr et al. [34]	HUI2	Nurses	Sensation	Cohen's kappa	0.05	
			Emotion		0.13	
			Cognition		0.54	
			Pain		0.71	
	HUI2	Physicians	Sensation	Cohen's kappa	0.42	
			Emotion		0.13	
			Cognition		0.37	
			Pain		0.73	
Fu et al. [42]	HUI2	Parents	Sensation	ICC	0.773	0.706, 0.826
			Mobility		0.67	0.584, 0.742
			Emotion		0.104	− 0.058, 0.262
			Cognition		0.121	− 0.026, 0.263
			Self-care		0.422	0.298, 0.532
			Pain		0.14	− 0.002, 0.277
	HUI2	Physicians	Sensation	ICC	0.829	0.778, 0.870
			Mobility		0.569	0.465, 0.657
			Emotion		0	− 0.143, 0.143
			Cognition		0.102	− 0.045, 0.245
			Self-care		0.754	0.686, 0.810
			Pain		0.08	− 0.063, 0.219
Morrow et al. [39]	HUI2	Parents	Sensation	Cohen's kappa	0.51	0.23, 0.78
			Mobility		0.59	0.31, 0.86
			Emotion		0.32	0.10, 0.53
			Cognition		0.29	0.35, 0.54
			Pain		0.44	0.23, 0.64
	HUI2	Physicians	Sensation	Cohen's kappa	0.27	− 0.26, 0.56
			Mobility		0.62	0.37, 0.88
			Emotion		0.18	− 0.03, 0.88
			Cognition		0.07	− 0.16, 0.30
			Pain		0.11	− 0.11, 0.34
	HUI3	Parents	Ambulation	Cohen's kappa	0.52	0.29, 0.77
			Dexterity		0.12	− 0.11, 0.34
			Emotion		0.27	0.04, 0.51
			Cognition		0.32	0.09, 0.55
			Pain		0.43	0.25, 0.62
	HUI3	Physicians	Ambulation	Cohen's kappa	0.56	0.31, 0.82
			Dexterity		0.11	− 0.12, 0.33
			Emotion		0.16	− 0.05, 0.37
			Cognition		0.05	− 0.11, 0.20
			Pain		0.36	0.17, 0.55
Glaser et al. [36]	HUI2	Physiotherapist	Sensation	Cohen's kappa	0.32	
			Mobility		NS	
			Emotion		0.37	
			Cognition		0.7	
			Self-care		0.43	
			Pain		NS	
	HUI2	Parents	Sensation	Cohen's kappa	0.54	
			Mobility		0.72	

Table 5 (continued)

Authors (intervention)	Measure	Proxy type	Attribute	Statistic reported	Agreement statistic (*p* value)	95% CI
			Emotion		0.37	
			Cognition		NS	
			Self-care		0.47	
			Pain		0.62	
	HUI2	Physicians	Sensation	Cohen's kappa	0.38	
			Mobility		0.77	
			Emotion		NS	
			Cognition		NS	
			Self-care		0.78	
			Pain		NS	
Glaser et al. [37]	HUI3	Physiotherapist	Vision	Cohen's kappa	0.62	
			Hearing		0.12	
			Speech		0.64	
			Ambulation		0.19	
			Dexterity		0.77	
			Emotion		0.4	
			Pain		0.33	
	HUI3	Parents	Vision	Cohen's kappa	0.62	
			Hearing		0.49	
			Speech		0.47	
			Ambulation		0.73	
			Dexterity		0.82	
			Emotion		0.28	
			Pain		0.56	
	HUI3	Physicians	Vision	Cohen's kappa	0.6	
			Hearing		0.67	
			Speech		0.14	
			Ambulation		0.77	
			Dexterity		0.48	
			Emotion		0.14	
			Pain		0.14	
Ungar et al. [49]	HUI2	Parents	Mobility	ICC	0.108	− 0.101, 0.308
			Emotion		0.065	− 0.155, 0.278
	HUI2	Parent with child	Mobility	ICC	0.713	0.593, 0.802
			Emotion		0.468	0.281, 0.621
Verrips et al. [38]: Mail	HUI3	Parents	Vision	Cohen's kappa	0.87	
			Hearing		0.33	
			Speech		0.23	
			Ambulation		0.66	
			Dexterity		0.63	
			Emotion		0.29	
			Cognition		0.36	
			Pain		0.43	
Verrips et al. [38]: Telephone	HUI3	Parents	Vision	Cohen's kappa	0.69	
			Speech		0.21	
			Ambulation		0.73	
			Dexterity		0.61	
			Emotion		0.2	
			Cognition		0.17	

Table 5 (continued)

Authors (intervention)	Measure	Proxy type	Attribute	Statistic reported	Agreement statistic (p value)	95% CI
			Pain		0.22	
Verrips et al. [38]: Face-to-face	HUI3	Parents	Vision	Cohen's kappa	0.75	
			Hearing		0	
			Speech		0.19	
			Ambulation		0.39	
			Dexterity		0.8	
			Emotion		0.07	
			Cognition		0.09	
			Pain		0.08	
Wolke et al. [41]	HUI3	Parents	Vision	Cohen's kappa	0.87	0.88, 0.86
			Hearing		0.59	0.59, 0.59
			Speech		0.22	0.22, 0.22
			Ambulation		0.78	0.78, 0.78
			Dexterity		0.67	0.68, 0.66
			Emotion		0.41	0.42, 0.4
			Cognition		0.32	0.32, 0.32
			Pain		0.48	0.49, 0.47
Wolke et al. [41]: General health (control)	HUI3	Parents	Vision	Cohen's kappa	0.82	0.81, 0.83
			Hearing		1	0.99, 1.01
			Speech		0.23	0.23, 0.23
			Dexterity		0.67	0.66, 0.68
			Emotion		0.37	0.36, 0.38
			Cognition		0.2	0.2, 0.2
			Pain		0.46	0.45, 0.47
Gusi et al. [40]			Pain or discomfort		0.68 (< 0.05)	
			Worried, sad, or unhappy		0.221 (< 0.05)	
Jelsma and Ramma [35]	EQ-5D-Y	Mother	Mobility	Cohen's kappa	0.15	
			Self-care		0.08	
			Doing usual activities		0.01	
			Pain or discomfort		0.2	
			Worried, sad, or unhappy		0.21	
Jelsma and Ramma [35]: General health (control)	EQ-5D-Y	Mother	Mobility	Cohen's kappa	0.6	
			Self-care		0.33	
			Doing usual activities		0.34	
			Pain or discomfort		0.41	
			Worried, sad, or unhappy		0.22	
Perez Sousa et al. [46]	EQ-5D-Y	Mother	Mobility	Cohen's kappa	0.713 (< 0.001)	
			Self-care		0.057 (0.536)	
			Doing usual activities		0.436 (< 0.001)	
			Pain or discomfort		0.128 (0.183)	
			Worried, sad, or unhappy		0.165 (0.14)	
	EQ-5D-Y	Father	Mobility	Cohen's kappa	0.042 (0.653)	
			Self-care		0.044 (0.622)	
			Doing usual activities		0.019 (0.841)	
			Pain or discomfort		0.067 (0.469)	
			Worried, sad, or unhappy		0.016 (0.854)	

Table 5 (continued)

Authors (intervention)	Measure	Proxy type	Attribute	Statistic reported	Agreement statistic (p value)	95% CI
Perez Sousa et al. [51]	EQ-5D-Y	Parents	Mobility Self-care Doing usual activities Pain or discomfort Worried, sad, or unhappy Mobility	Cohen's kappa	0.51 (< 0.001) 0.36 (< 0.001) 0.22 (< 0.001) 0.27 (< 0.001) 0.42 (< 0.001)	
Perez Sousa et al. [51]: control	EQ-5D-Y	Parents		Cohen's kappa	0.15 (0.03)	
			Self-care		0.13 (0.04)	
			Doing usual activities		0.09 (0.19)	
			Pain or discomfort		0.26 (< 0.001)	
			Worried, sad, or unhappy		0.37 (< 0.001)	
Shiroiwa et al. [47]	EQ-5D-Y	Parents	Mobility	Cohen's kappa	0.5	
			Self-care		0.91	
			Doing usual activities		0.78	
			Pain or discomfort		0.15	
			Worried, sad, or unhappy		0.12	

HRQoL health-related quality of life, *CI* confidence interval, *HUI2* Health Utilities Index Mark 2, *HUI3* Health Utilities Index Mark 3, *EQ-5D-Y* EQ-5D Youth version, *ICC* intraclass correlation coefficient, *NS* non-significant

discomfort' and 'feeling worried, sad or unhappy', while the other reported lower agreement ranging from slight to fair across all domains [35, 47].

3.4 Meta-Analysis Results

In the following, results for the meta-analysis are provided for studies that reported the ICC (95% CI) for the overall utilities and Cohen's kappa for the domain-level HRQoL. Nine studies were included in the analysis to estimate the ICC for overall utilities elicited using child-specific generic preference-based measures [34, 36, 37, 42, 43, 48–50, 55, 56]. Six studies that reported the ICCs for EQ VAS scores were excluded as there is some debate in the literature about VAS scores and the extent to which they can be interpreted as utilities [46, 47, 51–54]. Kappa statistics for the domain-level agreement were reported for 10 studies employing HUI2 and 3 (five studies) [34, 36–39, 41] and EQ-5D-Y (five studies) [35, 40, 46, 47, 51]. However, since four of five studies using EQ-5D-Y did not report the standard errors of the kappa values or the percentage agreement values, the EQ-5D measure was excluded from the domain-level meta-analysis of agreement.

3.4.1 Inter-Rater Agreement for Overall Utilities

The overall ICC for all 24 samples using HUI2 and 3 with CHU9D was 0.49 (0.34–0.61) and without CHU9D was 0.48 (0.32–0.61). Figure 2 depicts the study-specific and overall estimates of ICC, their respective 95% CIs and the study weights (%). The test for homogeneity resulted in a Q test statistic of 196.18 ($p < 0.001$). The heterogeneity in the studies was high ($I^2 = 91\%$) due to the presence of high variability between studies.

Exploratory moderators such as type of measure, health condition, proxy, and the age composition of the children in the sample were used to potentially explain this heterogeneity. The moderators were categorised according to the (1) type of measure used—HUI2 (12 samples) or HUI3 (11 samples) or CHU9D (1 sample); (2) health condition assessed—cancer- (15 samples) or non-cancer-related (9 samples); (3) type of proxy used—parent/caregiver (16 samples) or health professional/teacher (8 samples); and (4) lower age limit of the sample—below 8 years (10 samples) or 8 years and above (14 samples).

HUI3 had an estimated ICC of 0.37 (0.18–0.53), much lower than HUI2, which had an estimated ICC of 0.58 (0.34–0.75). The overall ICC for cancer-related samples was 0.43 (0.27–0.57), whereas for samples with conditions other than cancer, including general health, it was 0.54 (0.28–0.73). The ICC estimate for parent proxies was 0.49 (0.31–0.63), whereas for health professionals it was only marginally lower at 0.47 (0.11–0.72). Samples that also included younger children had an ICC of 0.39 (0.33–0.44), which was lower than the ICC of 0.5 (0.44–0.56) with older children. However, none of the group differences were statistically significant and therefore did not suggest moderation by any of the included variables.

The results of the meta-regression showed that none of the explanatory variables were statistically significant, thus showing no significant differences in child and proxy agreement according to the type of measure, health condition

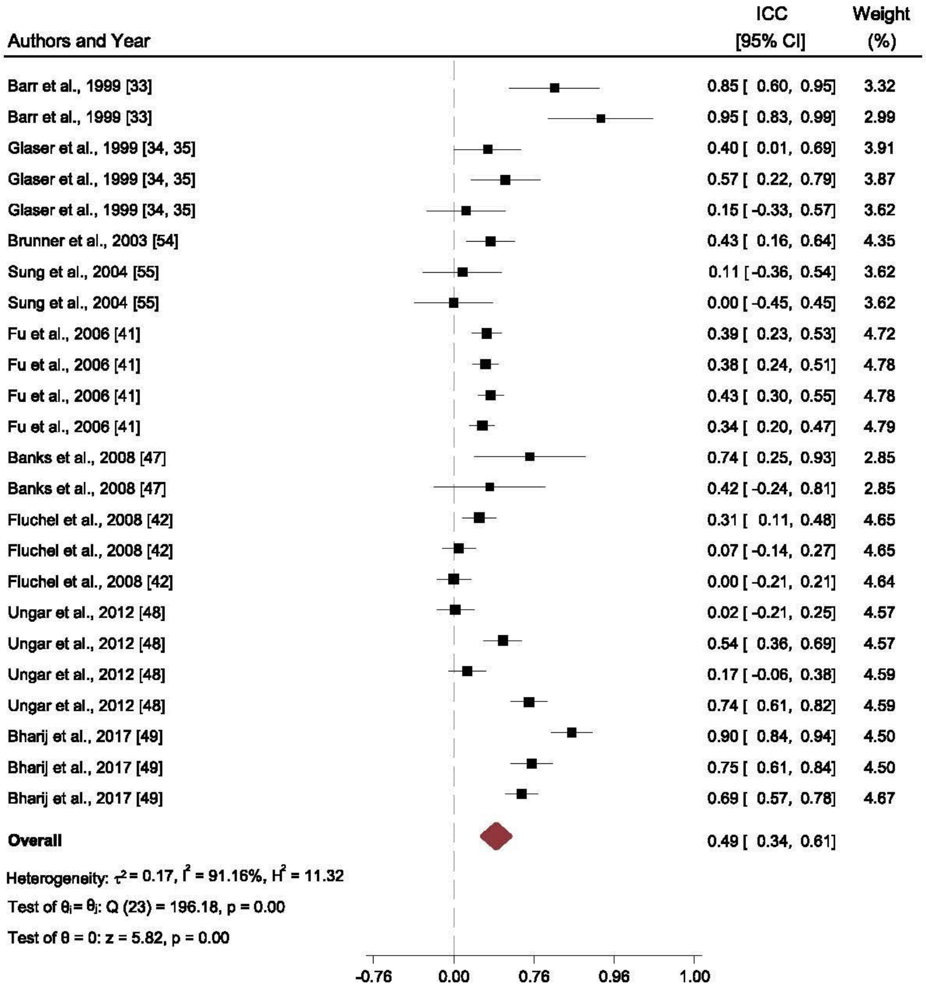


Fig. 2 Summary of the interrater reliability across studies. The forrest plot depicts the study-specific and overall estimates of ICCs, their respective 95% CIs and the study weight (%) for 24 studies obtained using a random effects model. *ICCs* intraclass correlation coefficients, *CIs* confidence intervals

experienced, proxy type and the inclusion of children below 8 years in the sample. The funnel plot and the funnel-plot test for asymmetry ($p = 0.133$) did not suggest any publication bias.

3.4.2 Inter-Rater Agreement for Domain-Level Health-Related Quality of Life

The estimated kappa and its 95% CI for HUI2 and 3 attributes is summarised in Table 6. In total, 36 samples for HUI2 and 68 samples for HUI3 were synthesised for the meta-analysis. The estimated kappa values for HUI2 attributes of ‘emotion’ (0.25), ‘cognition’ (0.3) and ‘pain’ (0.38), and the HUI3 attributes of ‘cognition’ (0.23), ‘emotion’ (0.27), ‘speech’ (0.3) and ‘pain’ (0.36) were the lowest. In contrast, there was higher agreement for the more easily observable physical- or function-related attributes such as ‘mobility’ (0.61) for HUI2 and ‘ambulation’ (0.64), ‘dexterity’ (0.65) and ‘vision’ (0.78) for HUI3. The heterogeneity was lower for HUI2 studies ($I^2 = 75\%$) than for HUI3 studies ($I^2 = 90\%$). Although no small-study bias was present in the analysis of HUI3 samples ($p = 0.327$), there was a possibility of such a bias using the HUI2 samples ($p = 0.003$).

4 Discussion

To our knowledge, this is the first study to comprehensively examine the evidence relating to the level of agreement between child- and proxy-reported paediatric HRQoL using generic preference-based measures across health conditions. This study systematically reviewed the papers reporting agreement measures to describe the inter-rater agreement in the assessment of paediatric HRQoL by child self and proxy reports.

Thirty studies were identified that reported the agreement statistics between child self- and proxy-reported overall and/or domain-level HRQoL. Most of these studies showed poor inter-rater agreement for overall utilities. At the domain level, there were some important differences common to all the generic preference-based measures. In particular, the agreement between children and proxy respondents was weaker for psychosocial-related HRQoL domains and stronger for physical HRQoL domains. No studies that reported agreement measures between self- and proxy-reported overall utilities over time were identified. This is an important omission as repeated HRQoL assessments over time form critical inputs for the calculation of QALYs for CUA. Divergences in self- and proxy-reported childhood utilities over time may impact, potentially substantially, upon the results of economic evaluations and regulatory decision making for the recommendation of new pharmaceuticals/medical technologies.

It is unclear if the preference-based measure/s applied in the identified studies have any influence on the level of agreement between self- and proxy-reported paediatric HRQoL. In this review, we found a greater agreement with HUI2 than HUI3. There are two main differences between the measures. First, the two measures differ in their response levels. HUI3 has 5–6 response levels whereas HUI2 has 3–5 [65]. Intuitively, a higher inter-rater agreement would be expected with measures with fewer response levels if the inter-rater agreement depended on the response levels within the measure. However, a study evaluating the child and proxy agreement using the EQ-5D-Y-3L and -5L versions found a higher agreement with the five-response-level version than with three [66]. Second, HUI2 and HUI3 have different underlying constructs for the attributes with the same name. For example, in HUI2 the ‘emotion’ attribute assesses distress and anxiety, while the HUI3 frames ‘emotion’ in terms of happiness rather than depression [65]. Currently, there is insufficient evidence to investigate whether the discrepancy reflects this difference or is a coincidental finding.

Table 6 Domain (attribute)-level overall kappa estimates with their 95% CIs for HUI2 and 3

Attribute	Agreement ($\hat{\kappa}$)	Lower 95% CI	Upper 95% CI
HUI2			
Self-care	0.576	0.347	0.806
Cognition	0.296	0.088	0.505
Emotion	0.250	0.158	0.342
Mobility	0.615	0.463	0.767
Pain	0.385	0.148	0.622
Sensation	0.409	0.306	0.512
HUI3			
Ambulation	0.641	0.535	0.747
Cognition	0.229	0.145	0.313
Dexterity	0.646	0.541	0.751
Emotion	0.272	0.190	0.353
Hearing	0.497	0.232	0.762
Pain	0.361	0.265	0.457
Speech	0.300	0.174	0.427
Vision	0.782	0.713	0.850

CIs confidence intervals, *HUI2* Health Utilities Index Mark 2, *HUI3* Health Utilities Index Mark 3, $\hat{\kappa}$ estimated kappa value

The agreement for EQ VAS was lower than for the EQ-5D-Y domains. This may be attributed to the fact that the VAS and the domain-level responses are elicited using different response scales. The VAS has a response scale from 0 to 100, whereas each of the five domains are described using a 3- or 5-level response scale [3]. Hence, a higher discrepancy may be expected with VAS due to the much larger range for its response scale.

Proxy type used was found to have some influence on the level of agreement between self- and proxy-reported paediatric HRQoL. The findings of HRQoL studies conducted in a paediatric oncology setting suggest that the information obtained from the child, the parent and the health professional are generally complementary and valid [67]. However, Sprangers and Aronson concluded that health professionals generally tend to underestimate the pain and also, conversely, the overall HRQoL of the individual [68]. While able to accurately assess the patient's physical condition, health professionals often failed to consider the emotional and social components of HRQoL [69]. In line with previous studies in adult cancer patients where agreement was higher with close companions, the child/parent agreement in this review was also found to be higher compared with child/health professional agreement [70]. Moreover, mothers demonstrated a higher agreement than fathers. This gender disparity may be associated with their degree of involvement in childcare [71].

The level of inter-rater agreement decreases with more severe conditions [69]. A recent study in paediatric patients found that the agreement between children and caregivers was higher when their condition improved compared with when they were ill [66]. We found that cancer-related cohorts had a lower overall agreement than cohorts with or without health conditions other than cancer. Interestingly, a low inter-rater agreement was seen between children with no obvious health conditions and their parents. One study showed worse correlations between parents and healthy children than children with a history of cancer [43]. These findings should be explored in more detail to determine whether this is a demonstrable trend. Self and proxy agreement data in the assessment of mental illnesses remains scarce. Studies have examined HRQoL in children with mental or behavioural disorders using preference-based measures, but none have assessed the level of child/proxy agreement [72, 73].

Self-report using the EQ-5D-Y has been prescribed for children aged 8 years and older [3]. The use of HUI2/3 was not recommended for self-report in children under 12 years of age [65]; however, studies have reportedly used these measures for self-completion in children younger than the recommended age group [35, 45, 48]. The minimum age at which children can reliably and accurately self-report has not been conclusively identified yet and is likely to be influenced by a variety of factors (including the reading and comprehension abilities of the child, the measure/s being applied and the mode of completion) [6]. There also remains a gap in the literature exploring the potential for differential levels of agreement between proxies and children by age groups. A previous study in a sample of children aged 8–18 years has shown that agreement decreases with age [74]. In this review, one study reported the agreement statistics (Gwet's AC1) for children (10–12 years) and adolescents (13–15 years) separately. In both groups, the correlation between child self- and proxy-reported domain-level HRQoL was strong and positive, with a marginally stronger association reported between adolescents and caregivers than children and caregivers [53]. Due to these inconsistent findings, further research is needed to determine if an age differential exists in the level of child/proxy agreement.

We found that 33% of the studies reported only the correlation coefficients that were synthesised to describe the inter-rater agreement in this review. The difference between agreement and correlation has been addressed in literature [19]. However, until recently, standalone correlation coefficients have been employed to assess agreement between child self and proxy report [75]. Correlation and agreement both measure the strength of association between two the variables of interest; however, the key difference is that agreement coefficients, in addition, account for the absolute agreement between the raters. Correlations may be high even if the ratings are not equal but only vary similarly. On the other hand, a perfect agreement would imply that all ratings, by each rater, are the same [14, 18]. Thus, correlation coefficients, if used, presented along with agreement statistics may provide a more comprehensive picture of the level of agreement.

This study has several limitations that are important to highlight. The inter-rater agreement for overall utilities and for the respective domains was quantitatively examined for only HUI2 and 3 for the following reasons. (1) HUI measures were widely used among the studies included in this analysis, with HUI3 being the most dominant. (2) Despite its relatively wide application, the majority of the identified studies using the EQ-5D-Y did not report the overall utilities, potentially due to the absence of an established preference-based scoring algorithm for the EQ-5D-Y to date. When reported, only the correlation (using Pearson's r or Spearman's rho) between the child self- and proxy-reported utilities was examined. While agreement was reported for the EQ VAS scores, they were not pooled due to paucity of evidence demonstrating the comparability of the VAS scores with the index scores. The EQ VAS scores were therefore not included in the meta-analysis. Furthermore, due to a lack of studies reporting the domain-level agreement between self and proxy reports of paediatric HRQoL, along with percentage agreement, the meta-analysis of the EQ-5D-Y domains was not feasible. (3) The analysis of the agreement level using the CHU9D and the QWB was also limited due to inadequate reporting of agreement statistics. Interpretation of the results of the meta-analysis is bounded by the presence of high heterogeneity between studies, which could not be explained by the subgroup analysis. Furthermore, due to practical resource constraints, we were only able to include articles published in the English language.

5 Conclusion

This systematic review summarising the agreement between child self and proxy rating of HRQoL using established generic preference-based measures generally found a poor inter-rater agreement. Convergence with child self-rating was more likely in the proxy assessment of paediatric HRQoL within domains with observable attributes e.g., physical health domains, than with less-observable attributes e.g., psychosocial domains. Further research to drive the inclusion of children in self-reporting their own HRQoL wherever possible and limiting the reliance on proxy reporting of children's HRQoL is warranted.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40273-022-01177-z.

Acknowledgements The authors would like to thank Shannon Brown and Pawel Skuza from Flinders University. Shannon Brown provided helpful advice on strategies for the literature search, and Pawel Skuza provided statistical consultation.

QUOKKA Project Team—Collaborating investigators: Nancy Devlin, Richard Norman, Rosalie Viney, Julie Ratcliffe, Kim Dalziel, Brendan Mulhern, Harriet Hiscock, Deborah Street, Gang Chen; Research fellows and PhD students: Tessa Peasgood, Cate Bailey, Christine Mpundu-Kaambwa, Alice Yu, Mina Bahrampour, Renee Jones, Rachel O'Loughlin, Yiting Luo, Alex van Heusden, Xiuqin Xiong, Diana Khanna, Ashwini De Silva.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Data availability statement All data generated and/or analysed during this study are included in this published article (and its supplementary information file).

Author contributions JR, JK, CM-K and DK conceptualised this study. DK led the systematic review with contributions from JK, CM-K and JR. DK and KL performed abstract and title screening. DK, KL, CM-K, JK and JR screened the full text of the articles for inclusion. DK wrote the first draft. JR, JK and CM-K provided feedback on the first draft and agreed on the final draft. All authors reviewed and approved the final amendments. DK and JR act as guarantors of the review.

Funding DK is supported by a PhD 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. The funder was not involved in developing this protocol.

Conflict of interest Diana Khanna, Jyoti Khadka, Christine Mpundu-Kaambwa, Kiri Lay, Remo Russo and Julie Ratcliffe declare no conflicts of interest.

References

1. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press; 2016.
2. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (version 5.0). Australian Government, Department of Health; 2016.
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. Virgili G, Koleva D, Garattini L, Banzi R, Gensini GF. Utilities and QALYs in health economic evaluations: glossary and introduction. Intern Emerg Med. 2010;5(4):349–52. https://doi.org/10.1007/s11739-010-0420-7.
5. 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.
6. 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.
7. Petrou S. Methodological issues raised by preference-based approaches to measuring the health status of children. Health Econ. 2003;12(8):697–702. https://doi.org/10.1002/hec.775.
8. Kwon J, Kim SW, Ungar WJ, Tsiplova K, Madan J, Petrou S. A systematic review and meta-analysis of childhood health utilities. Med Decis Mak. 2017;38(3):277–305. https://doi.org/10.1177/0272989X17732990.
9. Roddenberry A, Renk K. Quality of life in pediatric cancer patients: the relationships among parents' characteristics, children's characteristics, and informant concordance. J Child Fam Stud. 2008;17(3):402–26. https://doi.org/10.1007/s10826-007-9155-0.
10. 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.
11. Jiang M, Ma Y, Li M, Meng R, Ma A, Chen P. A comparison of self-reported and proxy-reported health utilities in children: a systematic review and meta-analysis. Health Qual Life Outcomes 2021;19(1):45. https://doi.org/10.1186/s12955-021-01677-0.
12. Otero S, Eiser C, Wright N, Butler G. Implications of parent and child quality of life assessments for decisions about growth hormone treatment in eligible children. Child Care Health Dev. 2013;39(6):782–8.
13. Shoukri MM. Measurement of agreement. Wiley StatsRef: Statistics Reference Online. © 2014, John Wiley & Sons, Ltd
14. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002.
15. Babineau J. Product review: covidence (systematic review software). J Can Health Libr Assoc Journal de l'Association des Bibliothèques de la Santé du Canada. 2014;35(2):68–71. https://doi.org/10.5596/c14-016.
16. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev. 2021;10(1):89. https://doi.org/10.1186/s13643-021-01626-4.
17. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted

with personality disorder samples. BMC Med Res Methodol. 2013;13:61. https://doi.org/10.1186/1471-2288-13-61.
18. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: measures of agreement. Perspect Clin Res. 2017;8(4):187–91. https://doi.org/10.4103/picr.PICR_123_17.
19. Liu J, Tang W, Chen G, Lu Y, Feng C, Tu XM. Correlation and agreement: overview and clarification of competing concepts and measures. Shanghai Arch Psychiatry. 2016;28(2):115–20. https://doi.org/10.11919/j.issn.1002-0829.216045.
20. Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Pract Assess Res Eval. 2004;9:4.
21. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37–46. https://doi.org/10.1177/001316446002000104.
22. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63. https://doi.org/10.1016/j.jcm.2016.02.012.
23. Juniper EF, Guyatt GH, Jaeschke R. How to develop and validate a new health-related quality of life instrument. In: Spilker B, editor. Quality of life and pharamacoeconomics in clinical trials. 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1996. p. 49–56.
24. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New York: Academic Press; 1988. p. 82.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
26. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. 2020;368: l6890. https://doi.org/10.1136/bmj.l6890.
27. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme. Version. 2006;1:b92. https://doi.org/10.13140/2.1.1018.4643.
28. Raudenbush SW. Analyzing effect sizes: random-effects models. The handbook of research synthesis and meta-analysis. 2nd ed. New York: Russell Sage Foundation; 2009. p. 295–315.
29. Sánchez-Meca J, López-López JA, López-Pina JA. Some recommended statistical analytic practices when reliability generalization studies are conducted. Br J Math Stat Psychol. 2013;66(3):402–25. https://doi.org/10.1111/j.2044-8317.2012.02057.x.
30. Sun S. Meta-analysis of Cohen's kappa. Health Serv Outcomes Res Method. 2011;11(3):145–63. https://doi.org/10.1007/s10742-011-0077-3.
31. Tufanaru C, Munn Z, Stephenson M, Aromataris E. Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness. Int J Evid Based Healthc. 2015;13(3):196–207. https://doi.org/10.1097/xeb.0000000000000065.
32. Kmet LM, Cook LS, Lee RC. Standard quality assessment criteria for evaluating primary research papers from a variety of fields. ERA. 2004. https://doi.org/10.7939/R37M04F16.
33. Papaioannou D, Brazier J, Paisley S. Systematic searching and selection of health state utility values from the literature. Value Health. 2013;16(4):686–95. https://doi.org/10.1016/j.jval.2013.02.017.
34. 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.
35. 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.
36. 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.
37. Glaser AW, Furlong W, Walker DA, Fielding K, Davies K, Feeny DH, et al. Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the UK. Eur J Cancer. 1999;35(2):256–61. https://doi.org/10.1016/s0959-8049(98)00367-0.
38. 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.
39. 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.
40. 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.
41. Wolke D, Chernova J, Eryigit-Madzwamuse S, Samara M, Zwierzynska K, Petrou S. Self and parent perspectives on health-related quality of life of adolescents born very preterm. J Pediatr. 2013;163(4):1020-6.e2. https://doi.org/10.1016/j.jpeds.2013.04.030.
42. Fu L, Talsma D, Baez F, Bonilla M, Moreno B, Ah-Chu M, et al. Measurement of health-related quality of life in survivors of cancer in childhood in Central America: feasibility, reliability, and validity. J Pediatr Hematol Oncol. 2006;28(6):331–41. https://doi.org/10.1097/00043426-200606000-00003.
43. 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.
44. Kulpeng W, Sornsrivichai V, Chongsuvivatwong V, Rattanavipapong W, Leelahavarong P, Cairns J, et al. Variation of health-related quality of life assessed by caregivers and patients affected by severe childhood infections. BMC Pediatr. 2013;13:122. https://doi.org/10.1186/1471-2431-13-122.
45. 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.
46. Perez Sousa M, Olivares Sánchez-Toledo PR, Gusi FN. Parent-child discrepancy in the assessment of health- related quality of life using the EQ-5D-Y questionnaire. Arch Argent Pediatr. 2017;115(6):541–6. https://doi.org/10.5546/aap.2017.eng.541.
47. Shiroiwa 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.
48. Banks BA, Barrowman NJ, Klaassen R. Health-related quality of life: changes in children undergoing chemotherapy. J Pediatr Hematol Oncol. 2008;30(4):292–7. https://doi.org/10.1097/MPH.0b013e3181647bda.
49. Ungar WJ, Boydell K, Dell S, Feldman BM, Marshall D, Willan A, et al. A parent-child dyad approach to the assessment of health

status and health-related quality of life in children with asthma. Pharmacoeconomics. 2012;30(8):697–712. https://doi.org/10.2165/11597890-000000000-00000.

50. Bharij A, Neighbors K, Alonso EM, Mohammad S. Health utility and quality of life in pediatric liver transplant recipients. Pediatr Transpl. 2020;24(4): e13720. https://doi.org/10.1111/petr.13720.
51. Perez-Sousa MA, Olivares PR, Garcia-Hermoso A, Gusi N. Does anthropometric and fitness parameters mediate the effect of exercise on the HRQoL of overweight and obese children/adolescents? Qual Life Res. 2018;27(9):2305–12. https://doi.org/10.1007/s11136-018-1893-5.
52. van Summeren J, Klunder JW, Holtman GA, Kollen BJ, Berger MY, Dekker JH. Parent-child agreement on health-related quality of life in children with functional constipation in primary care. J Pediatr Gastroenterol Nutr. 2018;67(6):726–31. https://doi.org/10.1097/mpg.0000000000002124.
53. 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-e807. https://doi.org/10.1097/brs.0000000000003431.
54. 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.
55. Brunner HI, Maker D, Grundland B, Young NL, Blanchette V, Stain AM, et al. Preference-based measurement of health-related quality of life (HRQL) in children with chronic musculoskeletal disorders (MSKDs). Med Decis Mak. 2003;23(4):314–22. https://doi.org/10.1177/0272989x03256008.
56. Sung L, Young NL, Greenberg ML, McLimont M, Samanta T, Wong J, et al. Health-related quality of life (HRQL) scores reported from parents and their children with chronic illness differed depending on utility elicitation method. J Clin Epidemiol. 2004;57(11):1161–6. https://doi.org/10.1016/j.jclinepi.2004.05.003.
57. Belfort MB, Zupancic JA, Riera KM, Turner JH, Prosser LA. Health state preferences associated with weight status in children and adolescents. BMC Pediatr. 2011;11:12. https://doi.org/10.1186/1471-2431-11-12.
58. Lee JM, Rhee K, O'Grady MJ, Basu A, Winn A, John P, et al. Health utilities for children and adults with type 1 diabetes. Med Care. 2011;49(10):924–31. https://doi.org/10.1097/MLR.0b013e318216592c.
59. Penn A, Lowis SP, Stevens MC, Shortman RI, Hunt LP, McCarter RJ, et al. A detailed prospective longitudinal assessment of health status in children with brain tumors in the first year after diagnosis. J Pediatr Hematol Oncol. 2011;33(8):592–9. https://doi.org/10.1097/MPH.0b013e31821388c0.
60. Rhodes ET, Prosser LA, Lieu TA, Songer TJ, Ludwig DS, Laffel LM. Preferences for type 2 diabetes health states among adolescents with or at risk of type 2 diabetes mellitus. Pediatr Diabetes. 2011;12(8):724–32. https://doi.org/10.1111/j.1399-5448.2011.00772.x.
61. Sinlapamongkolkul P, Surapolchai P. Health-related quality of life in Thai children with thalassemia as evaluated by PedsQL and EQ-5D-Y: a single-center experience. Mediterr J Hematol Infect Dis. 2020;12(1): e2020036. https://doi.org/10.4084/mjhid.2020.036.
62. Czyzewski DI, Mariotto MJ, Bartholomew LK, LeCompte SH, Sockrider MM. Measurement of quality of well being in a child and adolescent cystic fibrosis population. Med Care. 1994;32(9):965–72. https://doi.org/10.1097/00005650-199409000-00007.
63. Sims-Williams HJ, Sims-Williams HP, Mbabazi Kabachelor E, Warf BC. Quality of life among children with spina bifida in Uganda. Arch Dis Child. 2017;102(11):1057–61. https://doi.org/10.1136/archdischild-2016-312307.
64. Rogers HJ, Vermaire JH, Gilchrist F, Schuller AA. The relationship between caries-specific quality of life and generic wellbeing in a Dutch pediatric population. Dent J (Basel). 2019. https://doi.org/10.3390/dj7030067.
65. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual Life Outcomes. 2003;1:54. https://doi.org/10.1186/1477-7525-1-54.
66. 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.
67. 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.
68. Sprangers MAG, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. J Clin Epidemiol. 1992;45(7):743–60. https://doi.org/10.1016/0895-4356(92)90052-O.
69. Weinfurt KP, Trucco SM, Willke RJ, Schulman KA. Measuring agreement between patient and proxy responses to multidimensional health-related quality-of-life measures in clinical trials. An application of psychometric profile analysis. J Clin Epidemiol. 2002;55(6):608–18. https://doi.org/10.1016/s0895-4356(02)00392-x.
70. Sneeuw KC, Aaronson NK, Sprangers MA, Detmar SB, Wever LD, Schornagel JH. Value of caregiver ratings in evaluating the quality of life of patients with cancer. J Clin Oncol. 1997;15(3):1206–17. https://doi.org/10.1200/JCO.1997.15.3.1206.
71. Raley S, Bianchi SM, Wang W. When do fathers care? Mothers' economic contribution and fathers' involvement in child care. AJS Am J Sociol. 2012;117(5):1422–59. https://doi.org/10.1086/663354.
72. Creswell C, Violato M, Fairbanks H, White E, Parkinson M, Abitabile G, et al. Clinical outcomes and cost-effectiveness of brief guided parent-delivered cognitive behavioural therapy and solution-focused brief therapy for treatment of childhood anxiety disorders: a randomised controlled trial. Lancet Psychiatry. 2017;4(7):529–39. https://doi.org/10.1016/s2215-0366(17)30149-9.
73. Vermeulen KM, Jansen DE, Knorth EJ, Buskens E, Reijneveld SA. Cost-effectiveness of multisystemic therapy versus usual treatment for young people with antisocial problems. Crim Behav Ment Health. 2017;27(1):89–102. https://doi.org/10.1002/cbm.1988.
74. Rajmil L, López AR, López-Aguilà 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.
75. Abraham S, Edginton E, Cottrell D, Tubeuf S. Measuring health-related quality of life measures in children: lessons from a pilot study. Res Psychother. 2022. https://doi.org/10.4081/ripppo.2022.581.

Authors and Affiliations

Diana Khanna[1] · Jyoti Khadka[1,2] · Christine Mpundu-Kaambwa[1] · Kiri Lay[1] · Remo Russo[3,4] · Julie Ratcliffe[1] on behalf of The Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team

Jyoti Khadka
jyoti.khadka@flinders.edu.au

Christine Mpundu-Kaambwa
christine.mpundu-kaambwa@flinders.edu.au

Julie Ratcliffe
julie.ratcliffe@flinders.edu.au

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

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

[3] Department of Paediatric Rehabilitation, Women's and Children's Hospital, Adelaide, SA, Australia

[4] Faculty of Health Sciences, School of Medicine, Flinders University, Adelaide, SA, Australia

PharmacoEconomics
https://doi.org/10.1007/s40273-024-01356-0

ORIGINAL RESEARCH ARTICLE

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 Khanna[1] · **Jyoti Khadka**[1,2] · **Christine Mpundu-Kaambwa**[1] · **Gang Chen**[5] · **Kim Dalziel**[3,4] · **Nancy Devlin**[3] · **Julie Ratcliffe**[1] · **in collaboration with the Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team**

Accepted: 9 November 2023

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-ranked 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 [$N = 85$]	Parent [$N = 85$]
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 AC_1 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)
Overall						
N	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 (*p* value)			0.32	0.19	1	0.01[a]
Age group						
6–7 years						
N (%)	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 (*p* value)			0.78	0.49	0.79	0.48
8–10 years						
N (%)	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 (*p* value)			0.78	0.49	0.79	0.48
11–12 years						
N (%)	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 (*p* value)			0.09	0.15	0.43	0.01[a]
Long-term health condition/s						
No						
N (%)	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 (*p* value)			0.84	0.29	0.48	0.03[a]
Yes						
N (%)	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 (*p* 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 *p* 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)	
Age group					
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 p 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 ($n = 32$)					
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				
Health condition					
No ($n = 59$)					
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 ($n = 26$)					
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 p value[a]
	Some pain/discomfort	4 (0.15)	7 (0.27)	9 (0.35)	
	A lot of pain/discomfort	1 (0.04)		1 (0.04)	
Feeling worried, sad or unhappy	Not worried, sad or unhappy	20 (0.77)	13 (0.5)	14 (0.54)	0.19
	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 PedsQL™, 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 the 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)
Age group			
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)
Health condition			
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, AC_1 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

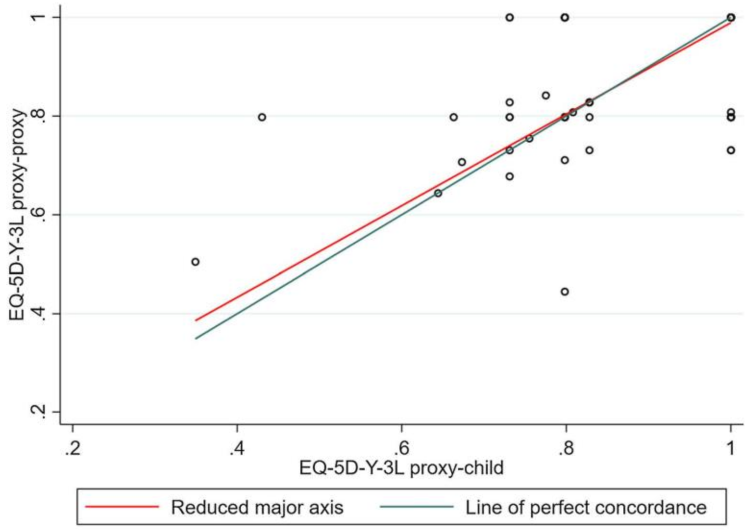


Fig. 2 Intra-proxy gap in agreement between the proxy-proxy and proxy-child versions of the EQ-5D-Y-3L. EQ-5D-Y-3L proxy-proxy = proxy version 1 report; EQ-5D-Y-3L proxy-child = proxy version 2 report

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.

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. Shiroiwa 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. Moum 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. 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 interrater 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. 2071.0—census of population and housing: reflecting Australia: stories from the Census, 2016. Canberrra: 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.
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.
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.
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-Aguilà 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.0000000000003431.
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, Hatziagorou 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 Khanna[1] · Jyoti Khadka[1,2] · Christine Mpundu-Kaambwa[1] · Gang Chen[5] · Kim Dalziel[3,4] · Nancy Devlin[3] · Julie Ratcliffe[1] · 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

Applied Health Economics and Health Policy (2023) 21:937–947
https://doi.org/10.1007/s40258-023-00831-7

ORIGINAL RESEARCH ARTICLE

Child–Parent Agreement in the Assessment of Health-Related Quality of Life Using the CHU9D and the PedsQL^{TM}

Diana Khanna[1] · Jyoti Khadka[1,2] · Christine Mpundu-Kaambwa[1] · Julie Ratcliffe[1] · in Collaboration with the Quality of Life in Kids: Key Evidence to Strengthen Decisions in 6 Australia (QUOKKA) Project Team

Accepted: 3 September 2023 / Published online: 29 September 2023

Abstract

Objective This study examined the inter-rater agreement between child-self and parental proxy health-related quality of life (HRQoL) ratings (overall and domain level) using two different generic child-specific measures, the Child Health Utility 9D (CHU9D) and the Pediatric Quality of Life Inventory (PedsQL^{TM}), in a community-based sample of Australian children. A secondary objective was to investigate the impact of age on child–parent agreement across the dimensions of the two measures.

Methods A total of 85 child–parent dyads (children aged 6–12 years) recruited from the community completed the self and proxy versions of the CHU9D and the PedsQL^{TM}, respectively. The inter-rater agreement was estimated using Concordance Correlation Coefficients (CCC) and Gwet's Agreement Coefficient (AC_1) for the overall sample and across age-groups.

Results Agreement was low for overall HRQoL for both the CHU9D (CCC = 0.28) and the PedsQL^{TM} (CCC = 0.39). Across the CHU9D dimensions, agreement was the highest for 'sad' (AC_1 = 0.83) and lowest for 'tired' (AC_1 = 0.31). The PedsQL^{TM} demonstrated stronger agreement (AC_1 = 0.41–0.6) for the physical health dimension but weaker for the psychosocial dimensions ($AC_1 < 0.4$). Except for the 'tired' dimension, agreement was consistent across age-groups with the CHU9D, whilst the PedsQL^{TM} showed poor agreement for most of the psychosocial health items among the older age-groups only (8–10 and 11–12 years).

Conclusion This study highlights that the agreement between child and parent proxy reported HRQoL may be influenced by both the measure used and the age of the child. These findings may have implications for the economic evaluation of healthcare interventions and services in child populations when both child and proxy perspectives are considered in the assessment of child HRQoL.

Key Points for Decision Makers

Child–parent agreement at the individual dimension level was higher for CHU9D than for PedsQL^{TM}. In contrast, agreement for overall HRQoL was lower for CHU9D relative to the PedsQL^{TM}.

In general, younger children (6–7 years) reported comparable agreement with parental proxies to their older counterparts providing some evidence to indicate that they may be able to meaningfully self-report.

✉ Diana Khanna
khan0420@flinders.edu.au

Jyoti Khadka
jyoti.khadka@flinders.edu.au

Christine Mpundu-Kaambwa
christine.mpundu-kaambwa@flinders.edu.au

Julie Ratcliffe
julie.ratcliffe@flinders.edu.au

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

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

1 Introduction

Health-related quality of life (HRQoL) is a key component of evaluating health outcomes to determine the value of health technologies, and a crucial indicator for appraising their quality [1, 2]. HRQoL measures can be broadly categorised into two main types: preference-based and non-preference-based. The primary distinction between the two is that the former measures generate health state utilities [3]. This allows for the calculation of Quality-Adjusted Life Years (QALYs), a key outcome measure in cost-utility analysis (CUA), widely used by healthcare decision-makers globally to inform reimbursement decisions for healthcare interventions and services [4].

The Child Health Utility 9D (CHU9D) is a preference-based HRQoL measure for application with children and young people and has been validated for the age group of 7–17 years. It is the only measure amongst the nine generic preference-based HRQoL measures that was designed exclusively for this population [5, 6]. The CHU9D has an adolescent specific value set available, facilitating the calculation of utilities based on the HRQoL preferences of young people themselves [7].

The Paediatric Quality of Life Inventory 4.0 (PedsQL^{TM} 4.0) Generic Core Scales is a generic HRQoL measure developed for use in both healthy and patient populations of children and adolescents [8, 9]. The PedsQL^{TM} has demonstrated reliability and validity as a self-report measure in children as young as 5–7 years old [8]. Since it is currently a non-preference-based measure, it is not possible to calculate utilities for the purposes of applying PedsQL^{TM} in economic evaluation. However, the instrument has been widely applied and recognised as a valuable tool for measuring HRQoL in a variety of paediatric populations in both clinical and research settings [10].

Self-reporting of a person's HRQoL from their own perspective is preferable wherever possible. However self-report is often challenging in child health research, especially for children with severe health conditions, very young children and for children with intellectual impairments/developmental delays [11, 12]. Hence, it is common for parents to serve as proxy respondents when assessing the HRQoL of children [13, 14]. While parents can provide valuable information about their child's HRQoL, it is important to note that they may not always have the same perception as the child [15]. Previous research has reported discrepancies between the child's self-reported HRQoL and that reported by their parents [16–18]. However, it is crucial to evaluate how closely the report provided by the parents aligns with the child's self-report to determine the extent to which the parental-report is representative of the child's own HRQoL.

In their review of parent–child reports of HRQoL, predominantly using the PedsQL^{TM}, Eiser and Varni [15] reported that the level of agreement between parents and children may be influenced by several variables. Potential factors identified as contributing to limited parent–child agreement included the type of dimension assessed [15]. Similar to the findings in the studies assessing self and proxy concordance in the reporting of HRQoL within the adult population [19, 20], dimensions associated with objective aspects of health typically showed higher agreement as compared with the more subjective (emotional or social) dimensions [15, 16]. A recent systematic review of self and proxy reporting of generic preference-based paediatric HRQoL measures by our team identified 17 studies reporting dimension-level agreement in children with and without health conditions. In contrast with more observable HRQoL dimensions relating to physical health and functioning, the agreement was observed to be lower for psychosocial-related dimensions (e.g. 'emotion' and 'pain' attributes of the Health Utilities Index Mark 2/3 or the 'having pain or discomfort' and 'feeling worried, sad, or unhappy' dimensions of the EQ-5D-Y) [17].

The age of the child is another important factor that may impact the child–parent agreement in the assessment of child HRQoL. However, the role of age is not yet clearly understood with inconsistent results reported for different age groups [15, 17]. A study by Cremeens and colleagues suggested that the age of the child may influence the level of agreement for the PedsQL^{TM} and may interact with the specific dimension being assessed [21]. In a sample of healthy children aged 5.5–8.5 years, they reported a significant agreement between older children (7.5–8.5 years) and parents for overall HRQoL. However, at the dimension level, a significant agreement was observed for the younger children (5.5–6.5 years) within the physical health dimension and for the older children within the psychosocial dimensions (7.5–8.5 years) [21]. To date, the differential effect of age on agreement remains largely unexplored, particularly using preference-based measures.

The main objective of this study was to examine the level of parent-child agreement in reporting of children's HRQoL (aged 6–12 years) using the CHU9D (a preference-based measure of children's HRQOL) and the PedsQL^{TM} (a non-preference-based measure of children's HRQOL) in a community-based sample of Australian children. A secondary objective was to explore the impact of age on child–parent agreement across the dimensions of the two measures.

2 Methods

2.1 Participants and Study Design

Participants for the study were recruited through a partnership with an independent research company, Stable

Research Australia. Parents who had previously expressed interest in participating in research studies were sent an invitation letter with details about the study. Children aged 6–12 years, along with their parents, were eligible to participate in this cross-sectional study provided the child was able to read and understand written English and did not have reading disorders or any other condition that would impact their ability to self-complete the measures.

Participants were selected using a proportional stratified random sampling method to ensure a representative sample of the general population in terms of socio-demographic characteristics and common health conditions affecting children, including asthma, anxiety disorders, conduct disorders, depressive disorders, autism spectrum disorders (ASD) and dental caries [22, 23]. To estimate Gwet's Agreement Coefficient (AC) between two raters with an acceptable error margin of 20%, a minimal sample size of $N = 25$ is necessary [23]. Therefore, the study aimed to achieve a sufficiently large ($N > 25$) and representative sample for a robust statistical analysis of child-parent agreement.

Parents provided information about the child's age, gender and presence of any long-term health condition/s. Additionally, the parents were also asked about their own socio-demographic characteristics including age, gender and postcode. Written informed consent to participate in the study was sought from the parent on behalf of the child prior to commencing the interview.

The study was approved by the Flinders University's Human Research Ethics Committee (Project ID 4178).

2.2 Procedure

Semi-structured, face-to-face interviews were conducted in April 2021, at Flinders University in South Australia. Child–parent dyads from the community consenting to participate in the study were invited. During the interview, the child was asked to complete the CHU9D and PedsQL^{TM}, and a single-item self-rated general health question, administered via an online platform (REDCap).

The parent completed hard copy (paper and pen) proxy versions of the CHU9D and the PedsQL^{TM} in the same interview room as their child while wearing noise cancelling headphones, to prevent their responses from being influenced by any conversations between the interviewer and the child. Additionally, the parent was also invited to complete an assessment of their own HRQoL using the EQ-5D-3L measure and the single-item self-rated general health question. Both online and paper-pen administrations are equivalent [24] as long as they are consistent for each rater [25]. The respective method for each rater was chosen as a matter of convenience and resource availability.

2.3 Measures

2.3.1 CHU9D

The CHU9D, a validated generic preference-based measure of children's HRQoL, includes nine dimensions: "Worried", "Sad", "Pain", "Tired", "Annoyed", "Schoolwork/homework", "Sleep", "Daily routine" and "Activities" and each dimension has five response levels. A scoring algorithm can be used to generate individual level utilities for all possible response combinations to the CHU9D. These utilities required for the calculation of quality adjusted life years (QALYs) for economic evaluation. The utilities range from 1 (full health) to (− 0.1059) for the most severe (PITS) state [7]. An Australian adolescent-specific preference-based scoring algorithm, derived from Australian adolescents aged 11–17 years, was applied in this study to calculate the CHU9D generated utilities [7].

2.3.2 PedsQL™ 4.0 generic core scales

The PedsQL^{TM} 4.0 Generic Core Scales include 23 items that are grouped into four scales (dimensions): physical functioning (8 items), emotional functioning (5 items), social functioning (5 items) and school functioning (5 items). The psychosocial dimensions represent the emotional, social and school functioning subscales of the PedsQL^{TM} whilst the physical dimension represent the physical functioning scale. Since the PedsQL^{TM} does not take into account preferences, equal weights are assigned for each of its 23 items when calculating the total score. Items were scored in reverse and transformed into a 0–100 continuous scale ($0 = 100$, $1 = 75$, $2 = 50$, $3 = 25$, $4 = 0$), such that higher scores denoted better HRQoL. To calculate the mean for individual Scale scores, the items were summed across and divided by the number of items answered. The average individual Scale scores were used to compute a total summary score [8].

2.3.3 EQ-5D-3L

The EQ-5D-3L measures HRQoL across five dimensions: "mobility", "self-care", "usual activities", "pain/discomfort" and "anxiety/depression". Each dimension has three different response options, ranging from no problems to severe problems [26]. An Australian adult scoring algorithm was applied to calculate the adult utilities.

2.4 Statistical Analysis

Data were analysed using Stata (16.1, Stata Corp LLC, College Station, TX) [27]. Differences in self-reported and proxy-reported CHU9D utilities and PedsQL^{TM} scores and inter-rater agreement were examined both for the overall sample and by age group (6–7 years, 8–10 years and 11–12 years). Additionally, the inter-rater differences and overall concordance were also examined for the subgroups categorised by the presence or absence of health condition (yes/no) and parent gender (female/male). Wilcoxon matched-pairs signed-rank test was used to compare the differences in child and proxy-reported overall HRQoL. Child–parent agreement was estimated using CCC for continuous data, e.g. CHU9D utilities, due to the non-normal distribution of the data [28]. Gwet's AC_1 was used to analyse agreement for categorical data e.g. CHU9D dimension-level HRQoL [29]. Agreement was compared between CHU9D dimensions and overlapping PedsQL^{TM} item/s representing the corresponding CHU9D dimensions [30]. The statistical significance level was set at 0.05.

Both CCC and AC_1 take values between − 1 and 1, with higher values indicating better agreement. The agreement results were interpreted using Altman's scale, which categorises agreement less than or equal to 0.2, 0.4, 0.6, 0.8, and 1 as poor, fair, moderate, good and very good [29]. A weighted version of Gwet's agreement coefficient (AC_2) accounts for partial agreement in adjacent categories allowing the measure to capture the varying degrees of agreement between the child–parent dyad [23]. The results for the weighted AC_2 using linear weights have been provided in the Supplementary Information (Table S1).

The Socio-Economic Indexes for Areas-Index of Relative Socioeconomic Disadvantage (SEIFA-IRSD) was used to estimate the socio-economic status of the participants based on information provided from the 2011 Australian Census using the residential post codes. The SEIFA-IRSD deciles measures the relative disadvantage of an area [31]. The SEIFA-IRSD deciles were grouped into quintiles, with the first six deciles categorised as disadvantaged areas (quintiles 1–3) and the last four as advantaged areas (quintiles 4–5).

3 Results

3.1 Child–Parent Participant Characteristics

A total of 89 child–parent dyads were identified as eligible and invited to participate in the study. Of those, four dyads were unable to attend the interview at the scheduled time, resulting in a response rate of 96% ($N = 85$). The children in the sample had an average age of 9 years (range 6–12 years) and the majority (56%) were female. The parents in the sample had an average age of 41 years (range 29–53 years) with the vast majority (81%) being child–mother dyads (Table 1). Most parents and children rated their own health as good to excellent on the self-rated general health question. This was further supported by the EQ-5D-3L measure, where parents reported a mean utility of 0.87 [standard deviation (SD) = 0.01]. Just under one-third (31%) of the children in the sample were identified by their parents as living with one or more health condition (Table 1). A proportion of the study participants (37%) resided in areas with relative socioeconomic disadvantage.

3.2 Child–Parent Difference in Reported HRQoL and Overall Concordance

Table 2 describes the child and parent reported HRQoL scores and the dyad agreement using the CHU9D and the PedsQL^{TM}, respectively. Overall, parents underreported children's HRQoL with the CHU9D but overreported with the PedsQL^{TM}. The difference in medians across the age groups was the largest for ages 11–12 years with the CHU9D and ages 6–7 years with the PedsQL^{TM}. However, these differences were not found to be statistically significant.

The overall agreement between child–parent dyads for both measures was fair with a lower agreement for CHU9D (0.28) (Fig. 1a) than for the PedsQL^{TM} (0.39) (Fig. 1b). The agreement between parents and 8–10-year-olds was good for both measures. For overall HRQoL, this was the only age group that demonstrated a statistically significant level of agreement across both measures.

Descriptive analysis indicated that the largest difference in medians in HRQoL ratings between children and proxies across the subgroups was observed in children without any reported health conditions using the CHU9D, while the PedsQL^{TM} also demonstrated a notable inter-rater gap within this subgroup. However, these differences were not statistically significant. Within the same subgroup, a lower agreement between child–parent dyads was also observed with both measures. Additionally, in comparison to the mother–child dyads, father–child dyads exhibited a lower agreement with the CHU9D but higher agreement with the PedsQL^{TM} (Supplementary Information Table S2).

3.3 Comparison of Agreement for CHU9D Dimensions and PedsQL^{TM} Items

Table 3 presents the agreement coefficients (AC_1), for the CHU9D dimensions and the corresponding representative PedsQL^{TM} items, for the overall sample and by age group. Child–parent agreement ranged from 0.65 to 0.95 for the CHU9D dimensions and 0.45 to 0.75 for the relevant PedsQL^{TM} items. The agreement was higher for CHU9D dimensions than for the corresponding PedsQL^{TM} items. Among the dimensions related to subjective (internal) experiences, agreement was the highest for 'sad' (CHU9D = 0.83) and 'feeling sad' (PedsQL^{TM} = 0.37) within the respective measures. The agreement was high for 'pain' (0.73) with the CHU9D, whereas its equivalent dimension in the PedsQL^{TM} showed the lowest (poor) agreement (0.15)

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

	Child ($N = 85$) N (%)	Parent ($N = 85$) N (%)
Age		
Mean (standard deviation)	9.13 (2)	41.7 (5.6)
Median (IQR)	9 (4)	41 (9)
Gender		
Female	47 (56)	68 (81)
Male	37 (44)	16 (19)
Self-rated general health		
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		
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)	–
Parent's HRQoL (EQ-5D-3L utility)	–	
Mean (standard deviation)	–	0.87 (0.01)
Socio-economic status 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)

SEIFA, Socio-Economic Indexes for Australia; IQR, interquartile range; HRQoL, health-related quality of life

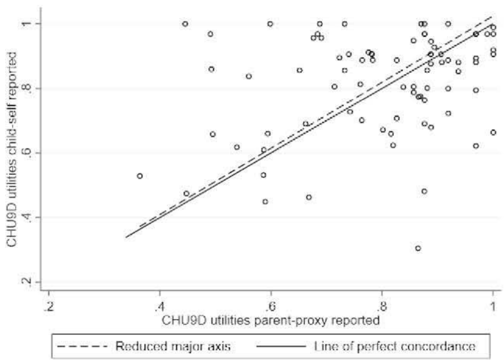
Table 2. Description of child and proxy reported HRQoL values and agreement using CHU9D and the PedsQL™ 4.0 generic core scales

	Child			Parent			Agreement	
	N	Mean (SD)	Median (IQR)	N	Mean (SD)	Median (IQR)	N	CCC (95% CI)
CHU9D								
Overall	81	0.82 (0.16)	0.86 (0.22)	84	0.79 (0.16)	0.83 (0.21)	80	0.28 (0.07, 0.47)
Age group								
6–7 years	22	0.81 (0.18)	0.86 (0.2)	22	0.82 (0.15)	0.87 (0.2)	21	− 0.18 (− 0.55, 0.26)
8–10 years	29	0.79 (0.16)	0.81 (0.22)	30	0.77 (0.16)	0.8 (0.21)	29	0.69 (0.43, 0.84)
11–12 years	30	0.84 (0.14)	0.89 (0.17)	32	0.79 (0.17)	0.84 (0.23)	30	0.19 (− 0.16, 0.5)
PedsQL™								
Overall	83	76.84 (13.49)	78.41 (15.46)	85	78.76 (12.53)	81.82 (15.76)	83	0.39 (0.2, 0.56)
Age group								
6–7 years	22	79.41 (13.6)	82.61 (12.55)	23	86.07 (9.01)	86.96 (13.04)	22	0.02 (− 0.31, 0.34)
8–10 years	30	73.11 (13.42)	75.57 (18.86)	30	75.69 (12.37)	78.98 (15.85)	30	0.67 (0.43, 0.83)
11–12 years	31	78.61 (13.11)	81.82 (13.64)	32	76.38 (13.04)	79.55 (18.18)	31	0.29 (− 0.06, 0.58)

Altman's scale interpretation: Less than or equal to 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

CCC, concordance correlation coefficient; SD, standard deviation; IQR, interquartile range

a ***Concordance between child and parent reported HRQoL utilities using the CHU9D***



b ***Concordance between child and parent reported HRQoL scores using the PedsQL™***

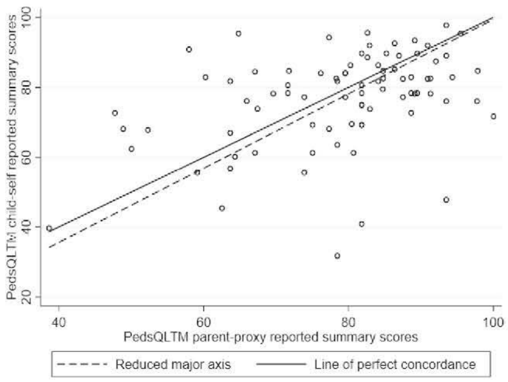


Fig. 1 **a** Concordance between child and parent reported HRQoL utilities using the CHU9D. **b** Concordance between child and parent reported HRQoL scores using the PedsQL™

compared with all other dimensions within the measure. The weakest agreement across the CHU9D dimensions was observed for 'tired' (0.31) followed by 'worried' (0.45). In addition to the items related to the psychosocial health mentioned above, a poor agreement was also observed for the PedsQL^{TM} item 'having trouble sleeping' (0.16). For the physical functioning related dimensions, agreement ranging between moderate to good was observed with both the CHU9D and the PedsQL^{TM}.

Across the age groups, for the CHU9D dimensions, the only statistically non-significant agreement was observed between parents and children aged 6–7 years for 'tired' (0.19). Moreover, for most dimensions, the agreement was lower for the 6–7-year-olds. In contrast, agreement across the majority of the relevant PedsQL^{TM} items was higher for the youngest age group (6–7 years) relative to the older age groups (8–10 and 11–12 years). Furthermore, an insignificant agreement was observed for several non-physical health-related items such as 'having low energy level', 'feeling angry' and 'having trouble sleeping' with both the older age-groups. They also demonstrated a poor agreement for the 'getting aches and pain' item. Additionally, an insignificant agreement was also seen between parents and 11–12-year-olds for the 'worrying what will happen to them' item.

4 Discussion

This study is the first, to our knowledge, to investigate child-parent agreement of child overall and dimension level HRQoL in a community-based sample of children using two generic HRQoL measures, the CHU9D and PedsQL™ 4.0. This study showed contrasting agreement for overall and dimension-level HRQoL using the two measures. The agreement between parents and children for HRQoL scores was stronger for the PedsQL^{TM}, but weaker for the CHU9D. Conversely, agreement for the individual dimensions was stronger for the CHU9D compared with the PedsQL^{TM} items.

The discrepancy in the consistency of agreement may be attributed, at least in part, to the statistical method used to measure the agreement. This study used two different methods to estimate agreement between the child and parent ratings: CCC for overall HRQoL and Gwet's AC_1 for dimension level HRQoL. Utilities or summary scores combine responses from different dimensions to estimate the overall HRQoL of the child. However, when analysing inter-rater agreement, the dimension/item level responses can offer a more direct measurement of agreement as it provides the disaggregated evaluations of the two raters, i.e. the child and the parent. This may be more informative regarding the specific areas of agreement or disagreement between the child and the parent and, therefore, provides a better understanding of the concordance in evaluations of each aspect of HRQoL. Furthermore, the estimation of CCC in this study may have been affected by an increased level of variation in ratings resulting from the high number of rater pairs, which could have potentially led to an underestimation of the true magnitude of the CCC [23].

The inter-rater differences in HRQoL scores across age groups using both measures did not correspond with the trends in agreement observed at the individual dimension level. For instance, in comparison with the other age groups, the 11–12 years age group had the greatest inter-rater gap with the CHU9D utilities. However, the dimension level agreement was similar across age groups. Additionally, while the same age group had the smallest inter-rater

Table 3 Comparison of child–parent agreement in CHU9D dimensions with relevant PedsQL™ items by age group

CHU9D dimensions	Overall (N = 85)	6–7 years (N = 23)	8–10 years (N = 30)	11–12 years (N = 32)	Relevant PedsQL items	Overall (N = 85)	6–7 years (N = 23)	8–10 years (N = 30)	11–12 years (N = 32)
Worried	0.45	0.43	0.42	0.49	2.1 Feeling afraid or scared	0.32	0.45	0.36	0.23
					2.5 Worrying about what will happen to him or her	0.28	0.48	0.26	0.18[#]
Sad	0.83	0.86	0.89	0.76	2.2 Feeling sad	0.37	0.6	0.4	0.22
Pain	0.73	0.75	0.74	0.7	1.7 Getting aches and pains	0.15	0.27	0.1[#]	0.15[#]
Tired	0.31	0.19[#]	0.45	0.27	1.8 Having a low energy level	0.26	0.53	0.2[#]	0.13[#]
Annoyed	0.56	0.5	0.63	0.55	2.3 Feeling angry	0.2	0.53	0.1[#]	0.07[#]
School work/homework	0.49	0.4	0.58	0.48	4.3 Keeping up with schoolwork	0.3	0.54	0.23	0.22
Sleep	0.54	0.6	0.46	0.58	2.4 Having trouble sleeping	0.16	0.37	0.13[#]	0.07[#]
Daily routine (eating, having a bath/shower, getting dressed)	0.52	0.44	0.55	0.55	1.5 Taking a bath or shower by him or herself (N = 22)	0.52	0.52		
Able to join in activities (playing out with friends, doing sports, joining things)	0.63	0.66	0.58	0.66	1.3 Participating in sports activity or exercise	0.52	0.64	0.42	0.54
					3.1 Getting along with other children	0.43	0.65	0.44	0.27
					3.5 Keeping up when playing with other children	0.5	0.61	0.39	0.52

Agreement statistics estimated using AC_1

Altman's scale interpretation: less than or equal to 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

[#] Agreement coefficient not statistically significant at 95% confidence level

difference with the PedsQL™ summary scores, they demonstrated lower agreement levels across most of its items compared with the youngest age group. Hence, it is important to acknowledge that the differences in the aggregated child and proxy reported HRQoL scores do not provide a measure of agreement [17].

Towards the opposite end of the age spectrum, a recent systematic review investigated the level of agreement between adult proxies and older adults with cognitive impairment [32]. Their findings indicated that there was some evidence suggesting higher levels of agreement in more observable HRQoL dimensions, such as physical health and mobility, compared with less observable dimensions like emotional well-being [32]. Typically, the available evidence indicates that parents also tend to be more concordant at reporting HRQoL dimensions related to the more easily observable attributes compared with those that are more subjective (internal) to the child [15, 17]. However, in this study, we found that with the CHU9D, a high level of agreement was obtained for the psychosocial health dimension 'sad'. The agreement for physical health-related dimensions ('daily routine' and 'able to join in activities') was low but moderate. These findings contrasted with the agreement observed for similar PedsQL™ items. For example, agreement was higher for PedsQL™ physical health items, i.e., 'participating in sports activity or exercise' and 'taking a bath or shower by him or herself' as compared with the 'feeling sad' item. Previous studies have reported a low agreement for pain using preference-based [33–37] and non-preference-based measures [38, 39]. In this study, a substantially higher level of agreement was observed for the 'pain' dimension with the CHU9D as compared with the 'getting aches and pains' item of the PedsQL™. Therefore, these findings suggest a possible interaction between the measure used and the dimension under consideration in determining the degree of agreement.

The findings in this study indicated a higher agreement for the CHU9D dimensions compared with the corresponding

PedsQL^{TM} items. Whilst both the measures were developed for use in children and adolescents in the development and validation of the instrument, the CHU9D followed a bottom-up approach that directly involved children in the development and validation of the instrument [5], whereas the PedsQL^{TM} adopted a top-down approach and was developed on the basis of a broader study of HRQoL in children with cancer [40]. The difference in agreement may also be attributed to the timeframe of assessment for each measure. In the CHU9D, respondents are asked about the (child's) health 'today' whereas the PedsQL^{TM} asks the respondent to report on their health over the 'past one month'. Thus, one possible explanation for the higher agreement found within the CHU9D dimensions may be its shorter time frame, which may reduce recall bias and result in less variability in perceived HRQoL [4]. Another contributing factor may be the difference in what the CHU9D and PedsQL^{TM} measures assess. The CHU9D measures the severity of impairment whereas the PedsQL^{TM} which measures frequency. For example, in the CHU9D dimension 'sad', the response levels range from 'don't feel sad' to 'feel very sad', whilst the PedsQL^{TM} response levels for the corresponding item 'feeling sad' range from 'never' to 'almost always' [4].

Studies reporting the level of child–parent agreement predominantly focus on samples including children aged 8 years and above [41–45]. The evidence for agreement in younger age groups, e.g. 6–7 years old and capable of self-reporting their HRQoL using the PedsQL^{TM} or the CHU9D is limited [21, 46]. In this study, dyads comprising the youngest age-group (6–7 years) reported relatively lower agreement with the CHU9D. This may be owing to children in this age group differing in their understanding of HRQoL as compared with their parents [47]. Younger children under 10 years of age have been reported to have difficulties with comprehension and recall of health-related events, as well as the associated frequency and severity [47]. However, except for the 'tired' dimension, there was no clear association between age and agreement across any other CHU9D dimensions. In contrast to the CHU9D findings, the older age groups, particularly the 11–12-year-olds, showed worse agreement for the PedsQL^{TM} items compared with the youngest age group comprising 6–7-year-olds. The evidence in the literature examining the relationship between age of the child and agreement using both preference and non-preference-based measures is inconsistent [17, 21, 48, 49]. This study found conflicting results in the same population for the two measures. The reasons for these discrepancies are unclear. Further research including mixed methods studies, which combine quantitative investigations with in-depth qualitative research using cognitive interviewing techniques, for example 'think-aloud', may be helpful in providing a more detailed understanding of the reasons for these discrepancies in reporting child HRQoL [50].

The existing literature on the influence of health status of the child on agreement is inconsistent for both preference and non-preference-based measures [17, 18]. Some studies suggest that in chronic illnesses, greater severity of the disease [51] or a higher frequency of exacerbations [52] may be associated with higher levels of child-parent agreement. However, for chronic conditions like cancer, there is a lack of consensus regarding the degree of agreement [15, 17]. Conversely, acute illnesses have been associated with lower inter-rater agreement [53]. Notably, Catchpool et al. reported a low agreement (Pearson's correlation coefficient = 0.13) in a sample of Australian children aged 11–12 years and their parents with the CHU9D [54]. Similarly, in this study, a lower agreement was observed for the overall HRQoL across both the measures for children without any reported health condition than those with reported health conditions. Additionally, a higher maternal than paternal involvement in childcare has been linked to the higher mother–child agreement levels evident in literature [17, 55, 56]. In this study, a similar trend was observed with the CHU9D, but this was not consistently reflected with the PedsQL^{TM}. Other studies have indicated that parental gender might not significantly confound parent proxy reports of child HRQoL [57, 58]. Considering that the literature is inconclusive, and the limited sample size of this study, further research with a larger sample size is warranted to substantiate our findings.

This study has limitations that are important to highlight. The study was conducted in a community-based sample of South Australian children who were relatively healthy. Hence, the findings may not be generalisable to more diverse samples including children with regular contact with health services and children with disabilities. Whilst the study sample was relatively small, good representation was achieved across age groups and approximately one-third of children were living with health conditions and/or living in areas of relative disadvantage. However, the main findings, particularly in relation to age group analyses, need to be interpreted with caution and further research needs to be conducted to substantiate these findings in larger community-based and patient samples. The CHU9D utility weights employed in this study were established using adolescents aged 11–17 years and then applied to a sample that included a younger age group. It is recognised that the value sets derived from children/adolescents may differ from those derived from adults adopting a child's perspective [59]. Nevertheless, additional research is required to determine the youngest age at which children can provide valuations, taking into account ethical considerations, and to explore the potential impact of this on valuing child HRQoL across different age groups. Moreover, as the utility weights were used to estimate the CHU9D scores, an additional preference-weighted step not currently available for the PedsQL™, score comparisons between the two measures were difficult. Finally, the study investigated agreement between child–parent dyads using the CHU9D and PedsQL^{TM} only, and hence the findings may not necessarily be generalisable to other measures used in the assessment of HRQoL in child populations.

5 Conclusion

This study found a low child–parent agreement for overall HRQoL across both measures, with CHU9D exhibiting a lower agreement relative to the PedsQL^{TM}. In contrast, at the individual dimension level, inter-rater agreement was higher for CHU9D than for PedsQL^{TM}. CHU9D showed the highest agreement with the dimensions of 'sad' and 'pain', whereas for the PedsQL^{TM}, agreement was the highest for the physical health items. There was no clear interaction between age and CHU9D dimensions. However, for the relevant PedsQL^{TM} items, the dimension level agreement was stronger for the youngest children (6–7 years) in the sample and weaker for older children (8–10 and 11–12 years), particularly for the psychosocial health items. Further research in larger and more diverse study samples and across age groups is needed to substantiate these findings. The introduction of a preference-based scoring algorithm for the PedsQL^{TM} will also facilitate empirical comparisons of child parental agreement at overall utility level and enable the impact of child and parent perspectives on HRQoL benefits for economic evaluations of interventions targeted at paediatric populations to be assessed.

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.

Declarations

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. D.K. is supported by a PhD 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 All authors declare no conflicts of interest.

Ethics approval The study was approved by the Flinders University's Human Research Ethics Committee (Project ID 4178).

Consent to participate All participants provided informed consent before participating in the study.

Consent for publication Not applicable.

Data availability statement The datasets generated during 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.

References

1. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press; 2017.
2. Varni JW, Burwinkle TM, Lane MM. Health-related quality of life measurement in pediatric clinical practice: an appraisal and precept for future research and application. Health Qual Life Outcomes. 2005;3(1):34. https://doi.org/10.1186/1477-7525-3-34.
3. Bulamu NB, Kaambwa B, Ratcliffe J. A systematic review of instruments for measuring outcomes in economic evaluation within aged care. Health Qual Life Outcomes. 2015;13:179. https://doi.org/10.1186/s12955-015-0372-8.
4. 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 PedsQLTM 4.0 Short Form 15. Patient Patient Centered Outcomes Res. 2018;11(1):29–37. https://doi.org/10.1007/s40271-017-0265-5.
5. 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.
6. 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.
7. 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.
8. Varni JW, Burwinkle TM, Seid M. The PedsQL as a pediatric patient-reported outcome: reliability and validity of the PedsQL Measurement Model in 25,000 children. Expert Rev Pharmacoecon Outcomes Res. 2005;5(6):705–19. https://doi.org/10.1586/14737167.5.6.705.
9. 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<0329:TPAAPP>2.0.CO;2.

10. Varni JW, Limbers CA. The pediatric quality of life inventory: measuring pediatric health-related quality of life from the perspective of children and their parents. Pediatr Clin N Am. 2009;56(4):843–63. https://doi.org/10.1016/j.pcl.2009.05.016.
11. Seid M, Varni JW, Jacobs JR. Pediatric health-related quality-of-life measurement technology: intersections between science, managed care, and clinical care. J Clin Psychol Med Settings. 2000;7(1):17–27. https://doi.org/10.1023/A:1009541218764.
12. Fischer KI, Barthel D, Otto C, Ravens-Sieberer U, Thyen U, Klein M, et al. Minimal associations between clinical data and children's self-reported health-related quality of life in children with chronic conditions—a cross-sectional study. Front Pediatr. 2019. https://doi.org/10.3389/fped.2019.00017.
13. 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.
14. Germain N, Aballéa S, Toumi M. Measuring the health-related quality of life in young children: how far have we come? J Mark Access Health Policy. 2019;7(1):1618661. https://doi.org/10.1080/20016689.2019.1618661.
15. Eiser C, Varni JW. Health-related quality of life and symptom reporting: similarities and differences between children and their parents. Eur J Pediatr. 2013;172(10):1299–304. https://doi.org/10.1007/s00431-013-2049-9.
16. 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.
17. 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.
18. Upton P, Lawford J, Eiser C. Parent–child agreement across child health-related quality of life instruments: a review of the literature. Qual Life Res. 2008;17(6):895. https://doi.org/10.1007/s11136-008-9350-5.
19. Roydhouse JK, Cohen ML, Eshoj HR, Corsini N, Yucel E, Rutherford C, et al. The use of proxies and proxy-reported measures: a report of the international society for quality of life research (ISOQOL) proxy task force. Qual Life Res. 2022;31(2):317–27. https://doi.org/10.1007/s11136-021-02937-8.
20. Hernández JD, Spir MA, Payares K, Posada AM, Salinas FA, Garcia HI, et al. Assessment by proxy of the SF-36 and WHO-DAS 2.0. A systematic review. J Rehabil Med. 2023;55:4493. https://doi.org/10.2340/jrm.v55.4493.
21. Cremeens J, Eiser C, Blades M. Factors influencing agreement between child self-report and parent proxy-reports on the Pediatric Quality of Life Inventory 4.0 (PedsQL) generic core scales. Health Qual Life Outcomes. 2006;4:58. https://doi.org/10.1186/1477-7525-4-58.
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. 4 ed: Advanced Analytics, LLC; 2014.
24. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes. 2015;13:167. https://doi.org/10.1186/s12955-015-0362-x.
25. Young NL, Varni JW, Snider L, McCormick A, Sawatzky B, Scott M, et al. The Internet is valid and reliable for child-report: an example using the Activities Scale for Kids (ASK) and the Pediatric Quality of Life Inventory (PedsQL). J Clin Epidemiol. 2009;62(3):314–20. https://doi.org/10.1016/j.jclinepi.2008.06.011.
26. 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.
27. StataCorp. Stata Statistical Software: Release 16. College Station: StataCorp LLC; 2019.
28. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989;45(1):255–68. https://doi.org/10.2307/2532051.
29. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol. 2013;13:61. https://doi.org/10.1186/1471-2288-13-61.
30. Xiong X, Dalziel K, Huang L, Mulhern B, Carvalho N. How do common conditions impact health-related quality of life for children? Providing guidance for validating pediatric preference-based measures. Health Qual Life Outcomes. 2023;21(1):8. https://doi.org/10.1186/s12955-023-02091-4.
31. Australian Bureau Of Statistics. Socio-Economic Indexes for Areas (SEIFA) 2033.0.55.001 2018. https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/756EE3DBEFA869EFCA258259000BA746/$File/SEIFA%202016%20Technical%20Paper.pdf. Accessed 10 Aug 2023.
32. Hutchinson C, Worley A, Khadka J, Milte R, Cleland J, Ratcliffe J. Do we agree or disagree? A systematic review of the application of preference-based instruments in self and proxy reporting of quality of life in older people. Soc Sci Med 2022;305:115046. https://doi.org/10.1016/j.socscimed.2022.115046.
33. Janse AJ, Sinnema G, Uiterwaal CSPM, Kimpen JLL, Gemke RJBJ. Quality of life in chronic illness: children, parents and paediatricians have different, but stable perceptions. Acta Paediatr. 2008;97(8):1118–24. https://doi.org/10.1111/j.1651-2227.2008.00847.x.
34. Shiroiwa 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.
35. Perez Sousa M, Olivares Sánchez-Toledo PR, Gusi Fuerte N. Parent-child discrepancy in the assessment of health-related quality of life using the EQ-5D-Y questionnaire. Arch Argent Pediatr. 2017;115(6):541–6. https://doi.org/10.5546/aap.2017.eng.541.
36. Perez-Sousa MA, Olivares PR, Garcia-Hermoso A, Gusi N. Does anthropometric and fitness parameters mediate the effect of exercise on the HRQoL of overweight and obese children/adolescents? Qual Life Res. 2018;27(9):2305–12. https://doi.org/10.1007/s11136-018-1893-5.
37. Fu L, Talsma D, Baez F, Bonilla M, Moreno B, Ah-Chu M, et al. Measurement of health-related quality of life in survivors of cancer in childhood in Central America: feasibility, reliability, and validity. J Pediatr Hematol Oncol. 2006;28(6):331–41. https://doi.org/10.1097/00043426-200606000-00003.

38. Hemmingsson H, Ólafsdóttir LB, Egilson ST. Agreements and disagreements between children and their parents in health-related assessments. Disabil Rehabil. 2017;39(11):1059–72. https://doi.org/10.1080/09638288.2016.1189603.
39. 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.
40. Varni JW, Seid M, Rode CA. The PedsQL™: Measurement model for the pediatric quality of life inventory. Med Care. 1999;37(2).
41. Petersen S, Hägglöf B, Stenlund H, Bergström E. Psychometric properties of the Swedish PedsQL, Pediatric Quality of Life Inventory 4.0 generic core scales. Acta Paediatr. 2009;98(9):1504–12. https://doi.org/10.1111/j.1651-2227.2009.01360.x.
42. Gkoltsiou K, Dimitrakaki C, Tzavara C, Papaevangelou V, Varni JW, Tountas Y. Measuring health-related quality of life in Greek children: psychometric properties of the Greek version of the Pediatric Quality of Life Inventory(TM) 4.0 Generic Core Scales. Qual Life Res. 2008;17(2):299–305. https://doi.org/10.1007/s11136-007-9294-1.
43. Kook SH, Varni JW. Validation of the Korean version of the pediatric quality of life inventory 4.0 (PedsQL) generic core scales in school children and adolescents using the Rasch model. Health Qual Life Outcomes. 2008;6:41. https://doi.org/10.1186/1477-7525-6-41.
44. Hijkoop A, Ten Kate CA, Madderom MJ, I. Jsselstijn H, Reuser JA, Koopman H, et al. Sex differences in children's health status as measured by the Pediatric Quality of Life Inventory (PedsQL)™: cross-sectional findings from a large school-based sample in the Netherlands. BMC Pediatr. 2021;21(1):580. https://doi.org/10.1186/s12887-021-03059-3.
45. Pakpour AH. Psychometric properties of the Iranian version of the Pediatric Quality of Life Inventory™ Short Form 15 Generic Core Scales. Singap Med J. 2013;54(6):309–14. https://doi.org/10.11622/smedj.2013123.
46. Lopez-Espejo MA, Nuñez AC, Ruz M, Saez V, Moscoso OC, Escobar RG. Caregiver-child agreement in health-related quality of life of autistic children and adolescents. J Autism Dev Dis. 2022. https://doi.org/10.1007/s10803-022-05689-7.
47. Ungar WJ, Mirabelli C, Cousins M, Boydell KM. A qualitative analysis of a dyad approach to health-related quality of life measurement in children with asthma. Soc Sci Med. 2006;63(9):2354–66. https://doi.org/10.1016/j.socscimed.2006.06.016.
48. Annett RD, Bender BG, DuHamel TR, Lapidus J. Factors influencing parent reports on quality of life for children with asthma. J Asthma. 2003;40(5):577–87. https://doi.org/10.1081/jas-120019030.
49. Rajmil L, López AR, López-Aguilà 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.
50. Collins D. Pretesting survey instruments: an overview of cognitive methods. Qual Life Res. 2003;12(3):229–38. https://doi.org/10.1023/A:1023254226592.
51. April KT, Feldman DE, Platt RW, Duffy CM. Comparison between children with juvenile idiopathic arthritis (JIA) and their parents concerning perceived quality of life. Qual Life Res. 2006;15(4):655–61. https://doi.org/10.1007/s11136-005-3690-1.
52. Britto MT, Kotagal UR, Chenier T, Tsevat J, Atherton HD, Wilmott RW. Differences between adolescents' and parents' reports of health-related quality of life in cystic fibrosis. Pediatr Pulmonol. 2004;37(2):165–71. https://doi.org/10.1002/ppul.10436.
53. 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.
54. 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.
55. Raley S, Bianchi SM, Wang W. When do fathers care? Mothers' economic contribution and fathers' involvement in child care. Am J Sociol. 2012;117(5):1422–59. https://doi.org/10.1086/663354.
56. Meryk A, Kropshofer G, Hetzer B, Riedl D, Lehmann J, Rumpold G, et al. Disagreement between mothers' and fathers' rating of health-related quality of life in children with cancer. Qual Life Res. 2023;32(6):1683–91. https://doi.org/10.1007/s11136-023-03341-0.
57. Doostfatemeh M, Ayatollahi SMT, Jafari P. Assessing the effect of child's gender on their father–mother perception of the PedsQL™ 4.0 questionnaire: an iterative hybrid ordinal logistic regression/item response theory approach with Monte Carlo simulation. Health Qual Life Outcomes. 2020;18(1):348. https://doi.org/10.1186/s12955-020-01601-y.
58. Rensen N, Steur LMH, Schepers SA, Merks JHM, Moll AC, Kaspers GJL, et al. Determinants of health-related quality of life proxy rating disagreement between caregivers of children with cancer. Qual Life Res. 2020;29(4):901–12. https://doi.org/10.1007/s11136-019-02365-9.
59. Bailey C, Howell M, Raghunandan R, Salisbury A, Chen G, Coast J, et al. Preference elicitation techniques used in valuing children's health-related quality-of-life: a systematic review. Pharmacoeconomics. 2022;40(7):663–98. https://doi.org/10.1007/s40273-022-01149-3.