The Australian paediatric multi-instrument comparison (P-MIC) study: data quality, feasibility, acceptability, and construct validity of the EQ-5D-Y-3L, EQ-5D-Y-5L, TANDI and PedsQL.

Renee Jones[1,2], Brendan Mulhern[3], Nancy Devlin[1], Rachel O'Loughlin[1], Xiuqin Xiong[1], Mina Bahrampour[3], Kristy McGregor[2], Shilana Yip[2], Harriet Hiscock[2,4,5], Kim Dalziel[1,2] on behalf of QUOKKA (Quality of Life in Kids: Key Evidence for Decision Makers in Australia)

1. Health Economics Unit, The University of Melbourne, Melbourne, Australia
2. Health Services and Economics, Murdoch Children's Research Institute, Melbourne, Australia
3. Centre for Health Economics Research and Evaluation, University of Technology Sydney, Australia
4. Health Services Research Unit, The Royal Children's Hospital, Melbourne, Australia
5. Department of Paediatric, The University of Melbourne, Melbourne, Australia

Funding: This research was funded by the Australian Government Medical Research Futures Fund, grant number 1200816 and EuroQol Research Foundation grant 361-RA. MCRI is supported by the Victorian Government's Operational Infrastructure Support Program. H.H. is supported by an NHMRC Practitioner Fellowship 1136222. R.J. is supported by the EuroQol Research Foundation grant 330-PHD. R.J., R.O. are supported by a Research Training Program Scholarship provided by the Australian Commonwealth Government and the University of Melbourne.

Conflicts of interest: K.D., N.D., B.M., H.H., R.J. have all received previous or current funding from the EuroQol Foundation who is the developer and copyright holder of some instruments included in this study. N.D. and B.M. are members of the EuroQol Group. The EuroQol Foundation are providing some direct funding for this research

Keywords: Health related quality of life, HRQoL, Patient reported outcome measure, PROM, Psychometric, EQ-5D-Y, PedsQL, paediatric.

This work will form part of two future publications titled:

- '*Comparative feasibility, acceptability, and quality of collecting paediatric health-related quality of life from the Australian Paediatric Multi-Instrument Comparison (P-MIC) Study*' and
- '*Comparative Psychometric Performance of common paediatric HRQoL instruments by child age, report type, and child health status: results from an Australian Paediatric Multi-Instrument Comparison (P-MIC) Study*'.

Abstract

Objectives: There is a lack of evidence on the performance of paediatric Health-related quality of life (HRQoL) instruments. Feasibility, acceptability, and construct validity are core aspects of the overall performance of a HRQoL instrument. Additionally, understanding the quality of paediatric HRQoL data collected using different administration modes is crucial. The paediatric multi-instrument comparison (P-MIC) study is the first of its kind: it uses concurrent measurement of common generic and condition-specific paediatric HRQoL instruments to facilitate comparisons of instrument performance. This paper aims to explore the quality of data collected in the P-MIC study, as well as present initial results on the feasibility, acceptability, and psychometric performance of the EQ-5D-Y-3L, EQ-5D-Y-5L, TANDI and PedsQL.

Methods: Data from the P-MIC study data cut 1, dated 6th May 2022, were used. The P-MIC study is collecting survey data on Australian children aged 2 to 18 years via The Royal Children's Hospital Melbourne Australia and online Australian survey panels, including a general population and health condition samples. Participants complete two surveys, an initial survey and follow-up survey, both involving the concurrent collection of PedsQL, EQ-5D-Y-3L, EQ-5D-Y-5L and TANDI, alongside other generic and condition-specific instruments. As PedsQL is a commonly used instrument validated in children aged 2 to 18 years it was chosen as the comparator instrument. To assess data quality, participant demographics, dropout rates, and quality checks were analysed descriptively. To assess feasibility and acceptability, self-reported difficulty completing each instrument, time to complete each instrument, and instrument response patterns were analysed descriptively. To assess construct validity, known group and convergent validity were assessed. Where appropriate, sub-group analysis was undertaken by child age, report type (proxy vs self-report), child health status, and recruitment method.

Results: More participants from the online panel sample were removed for not meeting minimum quality criteria (33.9%) compared to those recruited via hospital (2%). After their removal, the quality of the data, based on frequency of inconsistent responses to similar items, was similarly good across samples. A total of the 6,247 participants who completed the initial survey remained in the dataset for analysis, of these 2,142 (34.3%) completed the follow-up survey. Higher follow-up rates were found in the hospital sample (79.8%) compared to online panel samples (25.3%). Of the children recruited via the hospital, 76.4% had a chronic condition lasting at least 6 months, compared to 46.5% of the online panel condition groups sample. EuroQol instruments were quicker and participants reported they were easier to complete, however, had greater ceiling effect issues compared to PedsQL. All instruments demonstrated known group validity and correlations were generally in the directions expected.

Conclusion: PedsQL is widely used and accepted by clinicians in paediatrics. However, EuroQol's paediatric instruments have the advantage of being quicker and easier to complete. A disadvantage of EuroQol's paediatric instruments is their ceiling effects. The P-MIC study has shown that data collected in hospital settings is of better quality and includes more chronically unwell children than online panel settings. This can help to guide future multi-instrument comparison studies.

1. Introduction

Health-related quality of life (HRQoL) instruments aim to capture an individuals' perception of their own HRQoL in a reliable, valid, acceptable, and feasible way.(1) HRQoL instruments have many applications, including their use in clinical trials as an outcome measure to support economic evaluation, utilising HRQoL information accompanied by preference weights (values) to generate Quality Adjusted Life Year (QALY) estimates.(2) HRQoL instruments are now increasingly being used for other applications, such as routine clinical care and population health research.(2-4) In routine clinical care HRQoL instruments are used as patient reported outcome measures (PROMs) to assist in the identification of health problems, clinical decision making about treatments, and clinician-patient communication.(5-7) For HRQoL instruments to benefit society across this wide array of applications, it is vital they are able to accurately measure HRQoL to ensure the resulting evidence is robust.

Compared to adults, there are additional challenges when measuring HRQoL in children.(8) HRQoL instruments can be completed directly by the individual (self-report) or by a proxy (such as a parent). Proxy report is often required in younger children who cannot self-report and there are known differences between child self-report and parent proxy report.(8) Additionally, as children have a broad developmental range, particularly in very young children, different approaches to measuring HRQoL across different domains is required.(8) Given these challenges, few paediatric HRQoL instruments have been validated across a wide range of child ages, with most instruments validated from age five.(9) The Paediatric Quality of Life Inventory (PedsQL) Generic Core 4.0 is currently the most commonly used paediatric HRQoL instrument that has been validated in children younger than five.(10) The EuroQol group provides a suite of instruments that have the potential to allow for the measurement of HRQoL across the life course, from young children with the Toddler and Infant Questionnaire (TANDI),(11) to childhood and adolescence with the EQ-5D Youth (EQ-5D-Y) 3L and 5L,(12, 13) and into adulthood with the EQ-5D-5L.(14) However, the EQ-5D-Y-5L and TANDI are experimental measures and require further validation work.(11, 13, 15) Furthermore, there is a paucity of literature regarding the comparative performance of EuroQol instruments to commonly used and validated instruments such as the PedsQL.(16, 17)

Acceptability, feasibility, and construct validity are core aspects of understanding the overall performance of a HRQoL instrument.(7) The acceptability and feasibility of the TANDI, EQ-5D-Y-3L, EQ-5D-Y-5L, and PedsQL has previously been captured based largely on data

missingness.(16, 18-21) However, other aspects of acceptability, such as users own perspectives on difficulty or ease completing instruments is lacking. Feasibility of completing paediatric HRQoL instruments is often evaluated by assessing time to complete, however, few studies have presented data on time to complete for the TANDI, EQ-5D-Y-3L, EQ-5D-Y-5L, and PedsQL. Where this has previously been explored for the EQ-5D-Y-3L only an upper threshold (3 minutes) was reported and not an average time.(22) Instrument acceptability is also assessed by evaluating the floor and ceiling effects of instruments. These have been previously reported for PedsQL scales,(23, 24) EQ-5D-Y-3L,(11, 12, 20, 25-27) EQ-5D-Y-5L,(21, 28) and TANDI items,(11) highlighting issues with ceiling effects in the EQ-5D-Y-3L and EQ-5D-Y-5L.(21) However, previous studies have not compared both the floor and ceiling effects of these EuroQol instruments (EQ-5D-Y-3L, EQ-5D-Y-5L and TANDI) to the PedsQL in the same sample of children.

The construct validity of HRQoL instruments is commonly measured by assessing the known group validity and convergent validity.(23, 24) The known group validity of the TANDI, EQ-5D-Y-3L, EQ-5D-Y-5L and PedsQL has previously been reported, with each instrument demonstrating preliminary evidence of known group validity.(11, 12, 20, 25-27) However, previous evidence of known group validity has been completed using different methodology and in different samples, precluding comparisons across instruments. The convergent validity of both the EQ-5D-Y-3L and EQ-5D-Y-5L with the PedsQL was previously explored in a study of 286 patients aged 8 to 16 years recruited from five hospitals in Indonesia.(21) The study noted that the convergent validity of the EQ-5D-Y-3L and EQ-5D-Y-5L with the PedsQL spread from weak to strong.(21) However, evidence on the convergent validity of the EQ-5D-Y-3L and EQ-5D-Y-5L with the PedsQL in a general population sample and the convergent validity of the TANDI with the PedsQL is missing.

Furthermore, existing evidence for the acceptability, feasibility, and construct validity of the TANDI, EQ-5D-Y-3L, EQ-5D-Y-5L, and PedsQL is focused on single populations (e.g., specific conditions and age groups) and is often limited by smaller sample sizes. This precludes head-to-head comparisons of instrument acceptability, feasibility, and construct validity across a range of child ages and conditions in a large sample.(16, 17)

The Paediatric Multi-Instrument Comparison Study (P-MIC) was undertaken to address current gaps in this evidence.(29) Consensus-based Standards for the Selection of Health

Measurement Instruments (COSMIN) guidelines outline that understanding the quality of methods used to obtain data about instrument performance is vital to ensure the appropriate and transparent interpretation of conclusions.(30) As the P-MIC study will be used to generate large amounts of evidence regarding paediatric HRQoL instrument measurement accuracy, it is essential to understand the quality of the methods used in the study. Furthermore, HRQoL evidence can be collected via different pathways such as online panels or clinical samples. Understanding the differences in quality when collecting evidence via these different pathways is of international relevance in informing future similar research, and further developing the evidence base for paediatric HRQoL instruments. No previous study has compared the quality of collecting HRQoL data from a set of instruments using the same survey via different pathways.

The P-MIC study is the largest of its kind to compare common paediatric HRQoL instruments head-to-head in a range of child health contexts and settings. Using available Australian P-MIC study data, this study aims to understand the:

1) *Quality* of data collected in the P-MIC study in terms of participant and sample characteristics, minimum quality criteria, and inconsistent survey responses.
2) *Acceptability and feasibility* of the TANDI v2, EQ-5D-Y-3L, EQ-5D-Y-5L, and PedsQL Generic Core 4.0 in terms of self-reported difficulty completing, time to complete the instrument, and distribution of responses.
3) *Construct validity* of the TANDI v2, EQ-5D-Y-3L, EQ-5D-Y-5L, and PedsQL Generic Core 4.0 in terms of known group and convergent validity.
4) How aims 1-3 compare by child age (5-12 years versus 13-18 years), report type (proxy versus self-report), data collection type (hospital versus online panel), and online panel sample type (general population versus condition groups).

2. Methods

2.1. Study design and participants

Data from the P-MIC study data cut 1, dated 6th May 2022, were used.(29, 31) Although the majority of P-MIC recruitment has occurred, recruitment and data collection is still ongoing and hence interim data were used.(31) The P-MIC study prospectively collected demographic information alongside multiple generic paediatric HRQoL instruments concurrently from Australian children aged 2 to 18 years and their caregivers. The study includes three key samples of children: Sample 1) children with or without a health condition recruited via a large

tertiary paediatric hospital based in Victoria, Australia, Sample 2) the general population recruited via an online panel (available to all States and Territories across Australia), and Sample 3) nine condition-specific groups (Attention-Deficit/Hyperactivity Disorder (ADHD), anxiety and/or depression, Autism Spectrum Disorder (ASD), asthma, eating disorder, epilepsy, recurrent abdominal pain, sleep problems, and tooth problems) recruited via the same online panel or – for rarer conditions – via targeted recruitment through peak bodies associated with the disorder or tertiary hospitals, based within Victoria.

2.2. Data collection

Participants completed two surveys online via REDCap, an initial survey and follow-up survey at four weeks. A small subset of participants from the online panel general population sample received the follow-up survey at two days. HRQoL instruments were completed by the child (child self-report) if they were aged seven years or older and their caregiver reported that they were currently able to complete questions about their health and wellbeing. Where this was not possible, or the child was less than seven, the measures were completed by the caregiver (caregiver proxy report). For further information on P-MIC study methodology, including details of participant recruitment, please see the technical methods paper.(31)

2.3. Instruments and measures

The PedsQL Generic Core 4.0 versions for children aged 5 to 18 years include 23-items and the version for children aged 2 to 4 years includes 21-items, both cover 4 domains: physical functioning, emotional functioning, social functioning, and school functioning.(20) Respondents are asked to rate the frequency of each item over the past month on a 5-point scale from 0 (Never) to 4 (Almost always). To score PedsQL, raw item responses are reverse scored and linearly transformed (0=100, 1=75, 2=50, 3=25, 4=0).(10) PedsQL total score is calculated as the sum each item score divided by the number of items answered.(10) The EQ-5D-Y includes 5-items and cover 5 domains: mobility, looking after self, usual activities, pain/discomfort, and worried/sad.(32) Respondents are asked to rate the severity of each item on the day they are completing the instrument on a 3-point scale for the EQ-5D-Y-3L and a 5-point scale for the EQ-5D-Y-5L. The TANDI v2 includes 6-items and covers 6 domains: movement, play, pain, social interaction, communication, and eating.(33) Respondents are asked to rate the severity of each item on the day they are completing the instrument on a 3-point scale.(33)

2.4. Analysis

Analyses were completed in Stata Version 17 using available initial and follow-up P-MIC survey data for children aged 2 to 18 years.(34) The EuroQol instruments (EQ-5D-Y-3L, EQ-5D-Y-5L, and TANDI) were primarily compared to the PedsQL as it is a commonly used instrument validated across children aged 2 to 18 years. The EQ-5D-Y-3L and EQ-5D-Y-5L were assessed in children aged 5 to18 years and the TANDI were assessed in children aged 2 to 3 years. Where appropriate, sub-group analysis were completed by child age band (5 to 12 versus 13 to 18 years), report type (self-report versus proxy), data collection type (hospital versus online panel), and online panel sample type (general population versus condition groups). Sub-group analyses were completed with initial survey data only.

Aim 1, *quality* of P-MIC data.

Sample characteristics, including recruitment method and follow-up survey completion rates were reported descriptively. *Sociodemographic factors* were reported by caregivers during the initial survey, this included child, family, and caregiver characteristics and were reported descriptively. Where available, sociodemographic factors were compared to Australian population norms from the Longitudinal Study of Australian Children (LSAC).(35) *Minimum sample quality criteria* were set during the recruitment phase of the study. Any respondent not meeting minimum sample quality criteria were removed. *Inconsistent responses for similar items* were assessed by evaluating the proportion of participants who respond extremely inconsistently (+/- three levels) and very inconsistently (+/- two levels) to similar items. The most similar items were chosen from all available instruments in the P-MIC for assessment, these included the EQ-5D-Y-5L item 5 (pain)/CHU9D item 3 (pain) and EQ-5D-Y-5L item 2 (looking after self)/CHU9D item 8 (daily routine). Although this study does not include the CHU9D in instrument performance analysis, some items from the CHU9D were included in the quality assessment due to their similarity to EQ-5D-Y-5L items.

Aim 2, *acceptability and feasibility* of PedsQL, EQ-5D-Y-3L, EQ-5D-Y-5L, and TANDI.

Self-reported difficulty of each instrument was measured after completing each instrument on a 5-point scale from 1 'very easy' to 5 'very difficult'. *Time to complete each instrument* was automated via the online system, based on the number of seconds the instrument was open, before the participant clicked onto the next instrument. Participants were not able to click back. If participants left the instrument open on their electronic device, this 'break-time' would be included, hence, times were truncated at 600 seconds (10 minutes) as times above this were

not considered to be reasonable. *Distribution of responses* were evaluated by assessing the number of instrument items with *floor effects, ceiling effects and low response in the highest level*. An item was considered to have a large ceiling effect if more than 70% of participants selected the lowest severity or frequency outcome level,(11) an item was considered to have a large floor effect if more than 70% of participants selected the highest severity or frequency outcome level, and an item was considered to have a low response in highest level if less than 5% of participants selected the highest severity or frequency outcome level. Additionally, the total instrument ceiling effects for EuroQol instruments were assessed as the proportion who reported "no problems" across all items. Instruments were considered to have a ceiling effect at the total instrument level if more than 15% of respondents reported "no problems" across all items.(36)

Aim 3, *construct validity* of PedsQL, EQ-5D-Y-3L, EQ-5D-Y-5L, and TANDI.

Known group validity were assessed by comparing groups with expected differences in HRQoL. Group differences were assessed by comparing the means for each group and effect sizes were estimated using Cohen's D.(37) Effect sizes of 0.2-0.49 were considered small, 0.5-0.79 moderate, and ≥0.8 large.(37, 38) Children with a chronic health condition were considered a known group who were hypothesised to have differences in HRQoL compared to children without a chronic health condition. Additionally, sensitivity analysis was conducted on other known groups: children with an EQ VAS score ≤80, PedsQL total score ≤69.7, and PedsQL score ≤74.2. PedsQL cut points were derived from the self-report total score one standard deviation below the population mean in a large population sample (69.7), and the mean self-report total score from a large sample of children with a chronic condition (74.2).(10) PedsQL known group cut points were only applied to EuroQol instruments and not the PedsQL.
Convergent validity was assessed by estimating item correlations between PedsQL and EQ-5D-Y-3L, EQ-5D-Y-5L, and TANDI. Correlations were assessed using Spearman's correlation as the response data were not normally distributed. Correlations of 0.1-0.29 were considered weak, 0.3-0.49 moderate, and ≥0.5 strong.(37) Hypothesised correlations were set a priori where at least a moderate correlation between the EuroQol instrument item and PedsQL instrument item was expected and were based on item wording similarity.

3. Results

Of the 11,697 caregivers of children aged 2 to 18 years who consented to take part in the study, 9,103 (77.8%) completed the initial survey. Of this, 18 respondents were removed because they were a duplicate survey response and 2,838 respondents were removed because they did not meet minimum quality criteria, leaving 6,247 participants available for analysis.

3.1. Aim 1, Understand the quality of data collected in the P-MIC study.

The sample characteristics of the 6,247 children and their caregivers included in analysis are presented in Table 1. A total of 1,004 were children recruited via hospital (Sample 1), 1,875 were general population children recruited via an online panel (Sample 2), and 3,368 were children from one of nine condition-specific groups recruited primarily via an online panel (Sample 3). An even spread of ages and genders were achieved for both online panels and the hospital setting. A representative spread of incomes was also achieved across samples.(35) Children recruited via the hospital were not required to be a patient of the hospital, however, 785 (78.2%) reported they were a patient. Of the children recruited via the hospital (Sample 1) 76.4% had a chronic condition that had lasted or was expected to last more than six months, compared to 46.5% of the online panel condition group sample (Sample 3). The online panel samples (Sample 2 and 3) had a higher proportion of single parent households compared to participants recruited via the hospital (Sample 1) (28% versus 20.7%) and caregivers without a Bachelor degree or above (64.7% versus 51.4%).

Of the 6,247 participants who completed the initial survey, 2,142 (34.3%) completed the follow-up survey. Follow-up survey completion rates were higher in participants recruited via hospital (79.8%), compared to participants recruited via online panels (25.3%). The high follow-up survey completion rate in children aged 2 to 4 years old is due to that population being largely made up of children recruited via hospital (Sample 1). Additionally, follow-up survey completion rates were higher in participants from the online panel general population sample 2-day follow-up sample (72.2%) compared to participants from the online panel general population 4-week follow-up sample (35.5%) and online panel condition groups sample (20.1%).

Table 1: Sample and participant characteristics by child age, report type, data collection type, and online panel sample type.

Participant Characteristic	N (% of those who completed initial survey) or mean (sd)										
	Total sample	Child age			Report type (initial survey)		Data collection type		Online panel sample type		Australian Population Norm
		2-4 years	5-12 years	13-18 years	Self-report	Proxy report	Hospital	Online panel	General population	Condition groups	
Sample characteristics											
Completed initial survey, n(%)	6,247 (100)	800 (100)	3,491 (100)	1,956 (100)	3,528 (100)	2,719 (100)	1,004 (100)	5,243 (100)	1,875 (100)	3,368 (100)	n/a
Completed follow-up survey, n(%)	2,142 (34.3)	454 (56.8)	1,095 (31.4)	593 (30.3)	1,067 (30.2)	1,075 (39.5)	801 (79.8)	1,341 (25.6)	665 (35.5)	676 (20.1)	n/a
Gen pop 2-day follow-up	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	251 (72.2)	n/a	n/a
Gen pop 4-week follow-up	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	414 (27.1)	n/a	n/a
Sample, n(%)											
Recruited via hospital (sample 1)	1,004 (16.1)	274 (34.3)	485 (13.9)	245 (12.5)	443 (12.6)	561 (20.6)	n/a	n/a	n/a	n/a	n/a
Online panel general population (sample 2)	1,875 (30.0)	377 (47.1)	942 (26.9)	556 (28.4)	916 (33.7)	916 (33.7)	n/a	n/a	n/a	n/a	n/a
Online panel condition groups (sample 3)	3,368 (53.9)	149 (18.6)	2,064 (59.1)	1,155 (59.1)	1,242 (45.7)	1,1242 (45.7)	n/a	n/a	n/a	n/a	n/a
Asthma	370 (5.9)	0	278 (8.0)	91 (4.7)	237 (6.7)	133 (4.9)	n/a	n/a	n/a	n/a	n/a
Attention Deficit Hyperactivity Disorder (ADHD)	517 (8.3)	45 (5.3)	335 (9.6)	140 (7.2)	286 (8.1)	231 (8.5)	n/a	n/a	n/a	n/a	n/a
Autism Spectrum Disorder (ASD)	521 (8.3)	0	348 (10.0)	173 (8.8)	327 (9.3)	194 (7.1)	n/a	n/a	n/a	n/a	n/a
Anxiety or depression	470 (7.5)	0	181 (5.2)	289 (14.8)	382 (10.8)	88 (3.2)	n/a	n/a	n/a	n/a	n/a
Eating disorder	140 (2.2)	0	0	140 (7.2)	104 (3.0)	36 (1.3)	n/a	n/a	n/a	n/a	n/a
Epilepsy	196 (3.1)	11 (1.4)	132 (3.8)	53 (2.7)	120 (3.4)	76 (2.8)	n/a	n/a	n/a	n/a	n/a
Tooth problems	408 (6.5)	1 (0.1)	312 (8.9)	95 (4.7)	260 (7.4)	148 (5.4)	n/a	n/a	n/a	n/a	n/a
Sleep problems	376 (6.0)	94 (11.8)	215 (6.2)	67 (3.4)	161 (4.6)	215 (7.9)	n/a	n/a	n/a	n/a	n/a
Recurrent abdominal pain	370 (5.9)	0	263 (7.5)	107 (5.5)	249 (27.2)	121 (4.5)	n/a	n/a	n/a	n/a	n/a

Participant Characteristic	N (% of those who completed initial survey) or mean (sd)										
	Total sample	Child age			Report type (initial survey)		Data collection type		Online panel sample type		Australian Population Norm
		2-4 years	5-12 years	13-18 years	Self-report	Proxy report	Hospital	Online panel	General population	Condition groups	
Initial survey- proxy report, n(%)	2,719 (43.5)	800 (100)	1,445 (41.4)	474 (24.2)	n/a	n/a	561 (55.9)	2,158 (41.2)	916 (48.9)	1,242 (36.9)	n/a
Follow-up survey- proxy report, n(% completed follow-up)	1,159 (54.1)	456 (100)	523 (47.7)	180 (30.5)	n/a	n/a	496 (61.9)	663 (49.4)	400 (60.2)	263 (38.9)	n/a
Study Child characteristics											
Child age, mean (sd)	9.8 (4.5)	3.1 (0.8)	8.3 (2.2)	15.3 (1.7)	11.9 (3.4)	7.2 (4.4)	8.5 (4.8)	10.1 (4.4)	9.4 (4.8)	10.5 (4.1)	n/a
Child gender- Female, n(%)	2,854 (45.7)	362 (45.3)	1,556 (44.6)	936 (47.9)	1,624 (46.0)	1,230 (45.2)	444 (44.2)	2,410 (46.0)	894 (47.7)	1,516 (45.0)	48.8%
Child of Aboriginal and/or Torres Strait Islander origin- Yes, n(%)	392 (6.3)	52 (6.5)	221 (6.3)	119 (6.1)	224 (6.4)	168 (6.2)	32 (3.2)	360 (6.9)	70 (3.7)	290 (8.6)	3.8%
Child speaks language other than English at home- Yes, n(%)	596 (9.5)	124 (15.5)	316 (9.1)	156 (8.0)	306 (8.7)	290 (10.7)	144 (14.3)	452 (8.6)	247 (13.2)	205 (6.1)	13.1%
Child receiving care at The Royal Children's Hospital, n(%)	785 (12.6)	185 (23.1)	390 (11.2)	210 (10.7)	373 (10.6)	412 (15.2)	785 (78.2)	n/a	n/a	n/a	n/a
Child has chronic health condition or disability (lasting at least 6 months), n(%)	2,529 (40.5)	290 (36.3)	1,316 (37.7)	923 (47.2)	1,457 (41.3)	1,072 (39.4)	767 (76.4)	1,762 (33.6)	195 (10.4)	1,567 (46.5)	n/a
Special healthcare needs screener – Yes, n(%)	2,557 (40.9)	255 (31.9)	1,374 (39.4)	928 (47.4)	1,485 (42.1)	1,072 (39.4)	707 (70.4)	1,850 (35.3)	153 (8.2)	1,697 (50.4)	16.2%
PedsQL total score, mean (sd)	69.8 (19.2)	77.2 (16.9)	70.2 (18.3)	66.1 (20.6)	69.4 (19.3)	70.2 (19.0)	66.7 (20.2)	70.4 (18.9)	81.3 (15.0)	64.3 (18.1)	n/a
Caregiver and family characteristics											
Caregiver age, mean (sd)	39.9 (8.7)	34.5 (6.8)	37.9 (7.4)	45.7 (8.4)	41.5 (8.3)	37.8 (8.6)	40.4 (7.8)	39.8 (8.8)	40.6 (9.5)	39.4 (8.4)	39.7
Caregiver gender- Female, n(%)	5,142 (82.3)	693 (86.6)	2,895 (82.9)	1,554 (79.5)	2,873 (81.4)	2,269 (83.5)	881 (87.8)	4,261 (81.3)	1,474 (78.6)	2,787 (82.8)	n/a

Participant Characteristic	N (% of those who completed initial survey) or mean (sd)										
	Total sample	Child age			Report type (initial survey)		Data collection type		Online panel sample type		Australian Population Norm
		2-4 years	5-12 years	13-18 years	Self-report	Proxy report	Hospital	Online panel	General population	Condition groups	
Caregiver highest education level- bachelor degree or above, n(%)	2,336 (37.4)	374 (46.8)	1,322 (37.9)	640 (32.7)	1,239 (35.1)	1,097 (40.4)	488 (48.6)	1,848 (35.3)	758 (40.4)	1,090 (32.4)	28.6
Single parent household, n(%)	1,649 (26.8)	129 (16.4)	872 (25.2)	648 (33.9)	999 (28.7)	650 (24.4)	205 (20.7)	1,444 (28.0)	389 (21.0)	1,055 (31.8)	17.1
Remoteness (based on postcode)- major cities, n(%)	4,505 (72.1)	590 (73.8)	2,532 (72.5)	1,383 (70.7)	2,541 (72.0)	1,964 (72.2)	757 (75.4)	3,748 (71.5)	1,405 (74.9)	2,343 (69.6)	66.7

Abbreviations: n/a- not applicable. Note: Australian normative data obtained from the Longitudinal Study of Australian Children (LSAC) study.(35)

Table 2 summarises the reasons 2,838 (31.2%) participants were removed from the dataset for failing to meet minimum quality criteria. Few participants failed more than one minimum quality criteria (1.5%). The sample recruited via hospital had a lower proportion of participants removed for failing to meet minimum quality criteria compared to those recruited via online panel (2% versus 34.7%). Furthermore, the proportion of participants removed for failing to meet minimum quality criteria was higher in the online panel condition groups compared to the online panel general population sample (41.7% versus 17.0%). The most common reasons a participant was removed was because they had reported an inconsistent child age at initial and follow-up surveys (i.e. more than one year difference) or where the participant had screened into one of the online panel condition groups (Sample 3) and the caregiver did not report the child having this condition in the survey.

Table 2: Participants who completed the initial survey that did not meet minimum quality criteria by data collection type and online panel sample type.

Minimum quality criteria	Total n (%)	Data collection type		Online panel sample type	
		Hospital n (%)	Online panel n (%)	General population n (%)	Condition groups n (%)
Failed one or more minimum quality indicator criteria, n (% of all those who completed initial survey)	2,838 (31.2)	20 (2.0)	2,818 (34.7)	386 (17.0)	2,432 (41.7)
Failed two or more minimum quality indicator criteria, n (% of all those who completed initial survey)	137 (1.5)	5 (0.5)	132 (1.6)	17 (0.8)	115 (2.0)
Minimum quality criteria initial survey, n=9,103					
Child age outside of eligibility (2-18 years)	89 (1.0)	6 (0.6)	83 (1.0)	20 (0.9)	63 (1.1)
Caregiver age less than 18 years	134 (1.5)	9 (0.9)	125 (1.5)	41 (1.8)	84 (1.4)
Less than one third of the median time to complete the initial survey	152 (1.7)	5 (0.5)	147 (1.8)	49 (2.2)	98 (1.7)
Caregiver not reporting child disease condition for relevant disease group, n (% of all those who completed initial survey)	1,187 (13.0)	n/a	n/a	n/a	1,187 (20.3)
Minimum quality criteria follow-up survey, n=3,568					
Child age reported at follow-up not consistent with initial survey (i.e. child age reported as more than 1 year different)	1,415 (39.7)	n/a	n/a	292 (30.2)	1,123 (62.7)
Less than one third of the median time to complete the follow-up survey	11 (0.3)	5 (0.6)	6 (0.2)	3 (0.3)	3 (0.2)

Abbreviations: n/a- not applicable. Note: Participants may qualify for multiple criteria.

Table 3 summarises the quality of the data remaining in the P-MIC study after excluding respondents based on the minimum quality criteria. A total of 31 (0.5%) participants and 89 (1.4%) participants reported an extremely inconsistent response (+/- 3 levels) on the EQ-5D-Y-5L (pain)/CHU9D (pain) items and EQ-5D-Y-5L (looking after self)/CHU9D (daily routine) items, respectively in the initial survey. A similarly low proportion reported extremely inconsistent responses (+/- 3 levels) in the follow-up survey.

Table 3: Inconsistent responses to similar items and internal consistency of similar item combinations by child age, report type, data collection type, and online panel sample type.

Item combination	Survey		Child age			Report Type		Data collection type		Online panel sample type	
	Initial Survey	Follow-up Survey	2-4 years	5-12 years	13-18 years	Self	Proxy	Hospital	Online	General population	Condition groups
Extremely inconsistent response (+/- three levels) for items expected to be highly related, n(%)											
EQ-5D-Y-5L (pain)/CHU9D (pain)*	31 (0.5)	7 (0.3)	1 (0.1)	20 (0.6)	10 (0.5)	23 (0.7)	8 (0.3)	6 (0.6)	25 (0.5)	7 (0.4)	18 (0.5)
EQ-5D-Y-5L (looking after self)/CHU9D (daily routine)*	89 (1.4)	40 (1.9)	13 (1.6)	36 (1.0)	40 (2.0)	45 (1.3)	44 (1.6)	27 (2.7)	62 (1.2)	8 (0.4)	54 (1.6)
Very inconsistent response (+/-two levels) for items expected to be highly related, n(%)											
EQ-5D-Y-5L (pain)/CHU9D (pain)*	176 (2.8)	43 (2.0)	13 (1.6)	101 (2.9)	62 (3.2)	115 (3.3)	61 (2.2)	45 (4.5)	131 (2.5)	23 (1.2)	108 (3.2)
EQ-5D-Y-5L (looking after self)/CHU9D (daily routine)*	407 (6.5)	111 (5.2)	57 (7.1)	198 (5.7)	152 (7.8)	218 (6.2)	189 (7.0)	94 (9.4)	313 (6.0)	43 (2.3)	270 (8.0)

Abbreviations: n/a- not applicable. Note: *Children aged 2 to 4 years in the online panel samples (Sample 2 and 3) were allocated to receive either the EQ-5D-Y-5L adapted or EQ-5D-Y-3L adapted in the online panel groups and hence there are some missing data points (n=528) for children aged 2 to 4 years.

3.2. Aim 2, understand the acceptability and feasibility of the TANDI, EQ-5D-Y-3L, EQ-5D-Y-5L, and PedsQL.

Table 4 summarises the time to complete each instrument. For children aged 5 to 18 years, the median time to complete the initial survey varied by instrument with the shortest being the EQ-5D-Y-5L (28.8 seconds) and the longest being the PedsQL (96.1 seconds). For children aged 2 to 3 years, the TANDI had a shorter time to complete (26.4 seconds) compared to the PedsQL (85.5 seconds). Time to complete decreased for all instruments in the follow-up survey.

Figures 1a-e summarise the self-reported difficulty of completing each instrument. For children aged 5 to 18 years old, a large proportion of participants reported each instrument was very easy to complete, however the EQ-5D-Y-5L had the highest proportion of participants reporting it

very easy to complete (52.7%), compared to the EQ-5D-Y-3L (inc. VAS) (46.9%) and PedsQL (43.6%). For children aged 2 to 3 years old, a larger proportion of participants reported the TANDI was very easy to complete (66.9%) compared to the PedsQL (56.1%).

Instrument acceptability and feasibility varied by sub-groups, with faster completion times and greater ease of completion reported by younger (versus older) children, proxies (versus self-report), online panels (versus hospital sample), and general population children (versus condition groups).

Table 4: Time to complete each instrument in seconds by child age, report type, data collection type, and online panel sample type.

Instrument	Survey		Child age		Report Type		Data Collection Type		Online Panel Sample Type	
	Initial	Follow-up	5-12 yr olds	13-18 yr olds	Self-report	Proxy-report	Hospital	Online panel	General population	Condition groups
5 to 18 year olds (n=5,444), median time to complete in seconds (IQR)										
PedsQL	96.1 (73.4, 133.5)	90.1 (68.6, 127)	96.4 (72.7, 137.1)	95.7 (74.4, 127.5)	95.8 (72.8, 139.8)	96.6 (74.6, 126.4)	126.5 (95.2, 187.1)	92.5 (71.4, 126.1)	89.2 (67.5, 126.4)	93.7 (73.2, 125.8)
EQ-5D-Y-3L & VAS	46.2 (34.1, 64.7)	42.8 (30.3, 61.4)	46.4 (33.9, 66.0)	45.9 (34.4, 63.4)	45.7 (33.6, 65.4)	46.8 (35.0, 63.2)	59.8 (43.1, 89.7)	44.8 (33.1, 61.1)	41.2 (30.4, 57.3)	46.0 (34.3, 62.7)
EQ-5D-Y-5L	28.8 (20.8, 41.7)	27.3 (19.3, 40.7)	29.1 (20.7, 42.5)	28.4 (20.9, 40.4)	28.9 (20.8, 43)	28.7 (20.8, 39.6)	36.7 (25.7, 55.4)	27.9 (20.4, 39.7)	25.0 (18.1, 36.1)	29.3 (21.5, 41.2)
2 to 3 year olds (n=486) , median time to complete in seconds (IQR)										
PedsQL	85.5 (63.9, 117.6)	77.5 (58.4, 104.4)	n/a	n/a	n/a	n/a	97.6 (74.5, 133.1)	76.8 (59.0, 102)	73.8 (58.6, 100.9)	93.3 (63.7, 145.1)
TANDI	26.4 (19.2, 37.8)	24.6 (17.7, 37.2)	n/a	n/a	n/a	n/a	28.9 (20.0, 42.1)	25.3 (18.6, 34.5)	24.9 (18.5, 33.6)	29.3 (23.6, 45.4)

Abbreviations: n/a- not applicable.

Self-reported difficult for 5 to 18 year olds (n=5,444)

Figure 1a. PedsQL self-reported difficulty for initial survey by child age, report type, data collection type, and online panel sample type.

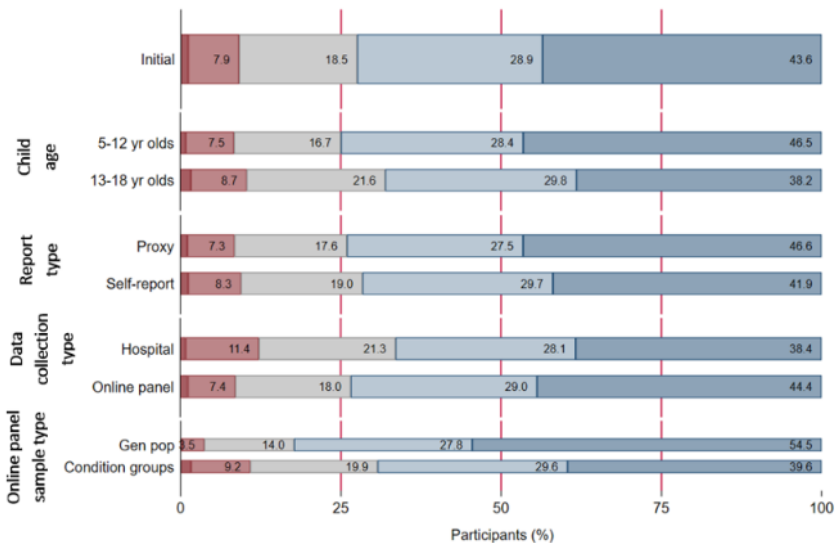


Figure 1b. EQ-5D-Y-3L (inc VAS) self-reported difficulty for initial survey by child age, report type, data collection type, and online panel sample type.

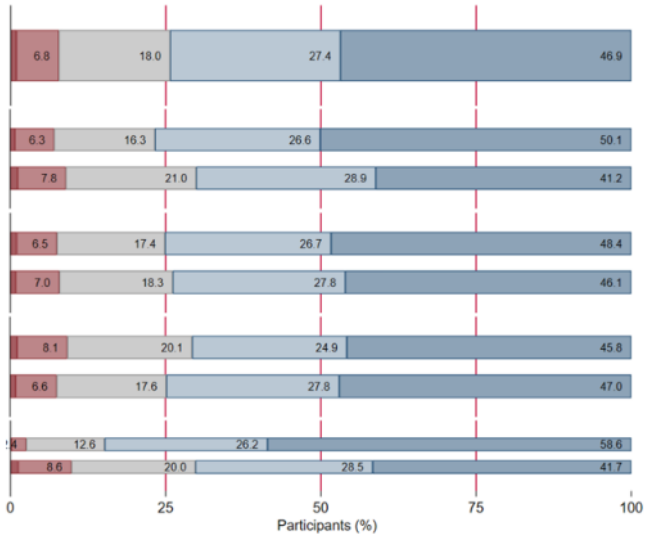
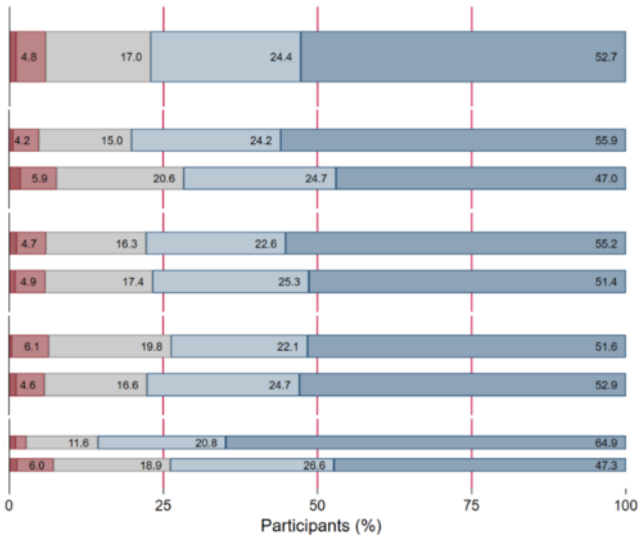


Figure 1c. EQ-5D-Y-5L self-reported difficulty for initial survey by child age, report type, data collection type, and online panel sample type.



Self-reported difficult for 2 to 3 year olds (n=486)

Figure 1d. PedsQL self-reported difficulty for initial survey by data collection type, and online panel sample type.

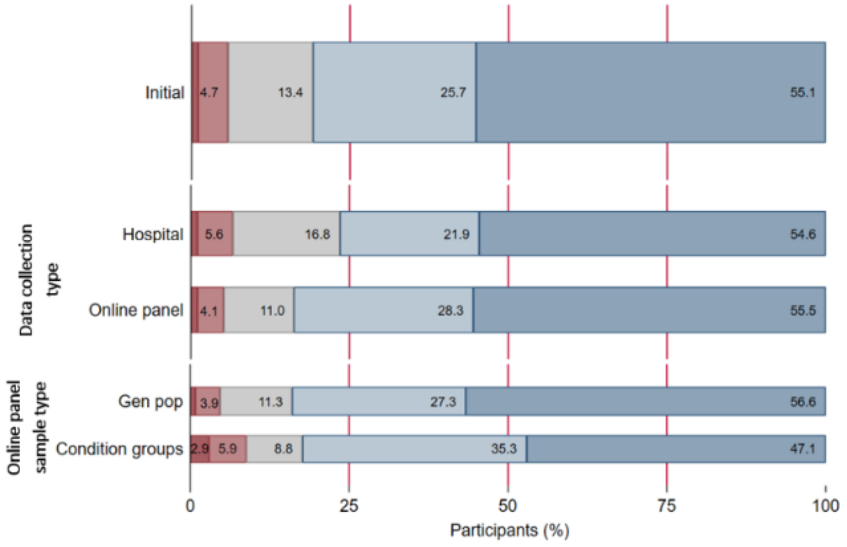
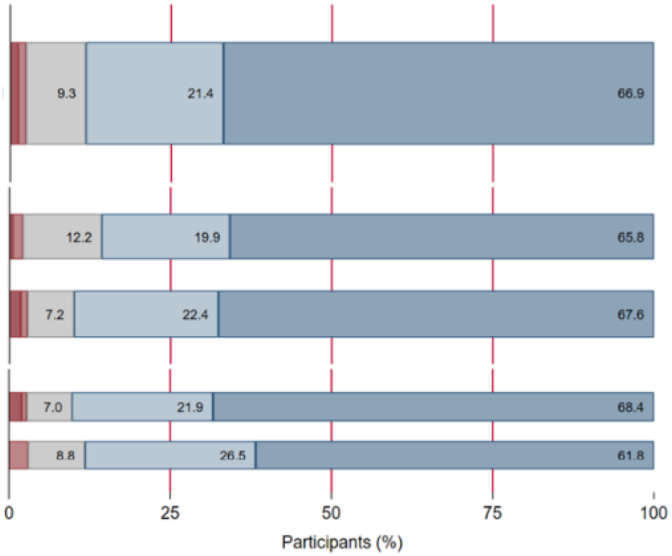


Figure 1e. TANDI self-reported difficulty for initial survey by data collection type, and online panel sample type.



Note: Where less than 2.5% of participants selected the category, the numeric value is not displayed.

Table 5 summarises the large ceiling effects and low response in worst level for each instrument. For children aged 5 to 18 years in the initial survey, no items in the PedsQL had large ceiling effects compared to 60% of items in both the EQ-5D-Y-3L and EQ-5D-Y-5L. Additionally, 52.2% of PedsQL items had low response in the worst level, compared to 80% of EQ-5D-Y-3L and 100% of EQ-5D-Y-5L items. For children aged 2 to 3 years, 14.3% of PedsQL items had large ceiling effects compared to 83.3% of TANDI items. Additionally, 76.2% of the PedsQL items had low response in the worst level, compared to 50% of TANDI items. A drop off occurred in large ceiling effects and low response in the worst levels for all instruments in the hospital sample (versus online panel) and condition groups (versus general population). No large floor effects (more than 70% selecting the highest outcome level) were found for any instrument item.

Among children aged 5 to 18 years, 36.3% (n=1,977) reported "no problems" on all five EQ-5D-Y-3L dimensions and 30.2% (n=1,647) on EQ-5D-Y-5L. Among children aged 2 to 3 years, 44% (n=216) reported "no problems" on all six TANDI dimensions. Participants from the online panel general population reported fewer problems across all dimensions compared to those from the hospital and condition groups samples. Overall ceiling effects were lower in the EQ-5D-Y-5L (18.8%, 21.0%) compared to the 3L (23.6%, 26.7%) in the hospital and condition groups samples, respectively.

Table 5: Number and percent of instrument items with large ceiling effects and low response in worst level by child age, report type, data collection type, and online panel sample type.

Instrument	Survey		Child age		Report type		Recruitment type		Online panel sample type	
	Initial	Follow-up	5-12 yr olds	13-18 yr olds	Self-report	Proxy-report	Hospital	Online panel	General population	Condition groups
5 to 18 year olds (n=5,444), n (%) of items										
PedsQL (23 items)										
Large ceiling effect	0	0	0	1 (4.3)	1 (4.3)	0	0	0	2 (8.7)	0
Low response in worst level	12 (52.2)	13 (56.5)	14 (60.9)	8 (34.5)	17 (73.9)	10 (43.5)	5 (21.4)	15 (65.2)	23 (100)	8 (2.7)
EQ-5D-Y-3L (5 items)										
Large ceiling effect	3 (60)	3 (60)	4 (80)	3 (60)	3 (60)	4 (80)	1 (20)	3 (60)	5 (100)	2 (40)

Instrument	Survey		Child age		Report type		Recruitment type		Online panel sample type	
	Initial	Follow-up	5-12 yr olds	13-18 yr olds	Self-report	Proxy-report	Hospital	Online panel	General population	Condition groups
Low response in highest level EQ-5D-Y-5L (5 items)	4 (80)	4 (80)	5 (100)	3 (60)	4 (80)	2 (40)	2 (40)	4 (80)	5 (100)	4 (80)
Large ceiling effect	3 (60)	3 (60)	3 (60)	2 (40)	3 (60)	2 (40)	1 (20)	3 (60)	4 (80)	2 (40)
Low response in worst level	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)	4 (80)
2 to 3 year olds (n=486) , n (%) of items										
PedsQL (21 items)										
Large ceiling effect	3 (14.3)	3 (14.3)	n/a	n/a	n/a	n/a	2 (9.5)	4 (19.0)	4 (19.0)	2 (9.5)
Low response in worst level	16 (76.2)	15 (71.4)	n/a	n/a	n/a	n/a	12 (57.1)	19 (90.5)	18 (85.7)	13 (61.9)
TANDI (6 items)										
Large ceiling effect	5 (83.3)	5 (83.3)	n/a	n/a	n/a	n/a	3 (50)	5 (83.3)	5 (83.3)	2 (33.3)
Low response in worst level	3 (50)	4 (66.7)	n/a	n/a	n/a	n/a	3 (50)	4 (66.7)	5 (83.3)	5 (83.3)

Abbreviations: n/a- not applicable. Note: An item was considered to have a large ceiling effect if more than 70% of participants selected the lowest severity or frequency outcome level and an item was considered to have a low response in worst level if less than 5% of participants selected the highest severity or frequency outcome level.

3.3. Aim 3, construct validity of the TANDI, EQ-5D-Y-3L, EQ-5D-Y-5L, and PedsQL.

Table 6 summarises the known group validity for groups of children with better or worse health using the PedsQL total score and EuroQol instrument sum scores. All instruments demonstrated large effects sizes across all known groups with the exception of the PedsQL in children aged 2 to 3 years for the known group of children with a chronic health condition which showed a medium effect size.

Table 6: Known group differences for instrument total score.

Instrument	Known group															
	Chronic health condition (condition lasting at least 6 months)				VAS score =<80 (higher score indicates better health)				PedsQL total score =<69.7 (higher score indicates better health)				PedsQL total score =<74.2 (higher score indicates better health)			
	M-Yes	M-No	P	ES	M-Yes	M-No	P	ES	M-Yes	M-No	P	ES	M-Yes	M-No	P	ES
5 to 18 year olds (n=5,444)																
PedsQL	58.9	75.6	<0.001	0.96	58.6	76.5	<0.001	1.1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
EQ-5D-Y-3L	7.4	6.0	<0.001	-0.89	7.4	5.9	<0.001	-0.99	7.5	5.6	<0.001	-1.3	7.3	5.5	<0.001	-1.2
EQ-5D-Y-5L	8.8	6.4	<0.001	-0.91	8.8	6.3	<0.001	-0.92	9.0	5.8	<0.001	-1.2	8.6	5.7	<0.001	-1.1
2 to 3 year olds (n=486)																
PedsQL	71.2	83.0	<0.001	0.73	65.3	83.5	<0.001	1.22	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Instrument	Known group															
	Chronic health condition (condition lasting at least 6 months)				VAS score =<80 (higher score indicates better health)				PedsQL total score =<69.7 (higher score indicates better health)				PedsQL total score =<74.2 (higher score indicates better health)			
	M-Yes	M-No	P	ES	M-Yes	M-No	P	ES	M-Yes	M-No	P	ES	M-Yes	M-No	P	ES
TANDI	8.6	7.0	<0.001	-0.82	9.2	7.0	<0.001	-1.10	10.0	6.8	<0.001	-1.8	9.4	6.7	<0.001	-1.5

Abbreviations: n/a- not applicable. Note: Green cells indicates large effect size (≥0.8), yellow indicates medium effect size (0.5-0.79), & white cells indicates small effect size (0.2-0.49).

Table 7 summarises the correlations between each EuroQol instrument (EQ-5D-Y-3L, EQ-5D-Y-5L and TANDI) and PedsQL at the item level. The PedsQL had largely weak to moderate correlations with most EuroQol instrument items. Where correlations were hypothesised is generally where the highest correlations were observed, except for the TANDI social interaction and PedsQL "getting teased by other children" correlation which was weak. Additionally, the PedsQL item "not being able to do things that other children can do" was highly correlated with the TANDI social interaction and communication items which were not hypothesised correlations. Of the hypothesised correlations with the PedsQL, 100% of the EQ-5D-Y-3L, 100% of the EQ-5D-Y-5L, and 85.7% of the TANDI correlations were at least moderate. Of the correlations with the PedsQL where a correlation was not expected, 50.9% of the EQ-5D-Y-3L, 57.7% of the EQ-5D-Y-5L, and 24.3% of the TANDI correlations were weak correlations.

Table 7: Initial survey correlations between the PedsQL and EuroQol instruments.

	EQ-5D-Y-3L items (n=5,444)					EQ-5D-Y-5L items (n=5,444)						TANDI items (n=486)				
PedsQL items	Mobility	Looking after self	Usual activities	Pain/ discomfort	Sad/ worried	Mobility	Looking after self	Usual activities	Pain/ discomfort	Sad/ worried	Movement	Play	Pain	Social Interaction	Communication	Eating
Walking more than one block	**0.40**	0.35	0.40	0.30	0.23	**0.42**	0.37	0.39	0.31	0.24	**0.64**	0.44	0.26	0.36	0.37	0.21
Running	**0.38**	0.28	0.37	0.33	0.26	**0.40**	0.29	0.38	0.36	0.29	**0.62**	0.43	0.20	0.37	0.39	0.22
Participating in sports activity or exercise	**0.37**	0.34	0.44	0.33	0.29	**0.39**	0.35	0.45	0.35	0.31	0.50	**0.46**	0.24	0.43	0.43	0.26
Lifting something heavy	0.30	0.32	0.34	0.32	0.24	0.30	0.32	0.34	0.28	0.24	0.42	0.31	0.16	0.24	0.30	0.22

	EQ-5D-Y-3L items (n=5,444)					EQ-5D-Y-5L items (n=5,444)						TANDI items (n=486)				
PedsQL items	**Mobility**	**Looking after self**	**Usual activities**	**Pain/ discomfort**	**Sad/ worried**	**Mobility**	**Looking after self**	**Usual activities**	**Pain/ discomfort**	**Sad/ worried**	**Movement**	**Play**	**Pain**	**Social Interaction**	**Communication**	**Eating**
Taking a bath or shower by him or herself	0.27	**0.59**	0.41	0.27	0.19	0.30	**0.62**	0.43	0.18	0.21	0.33	0.35	0.23	0.33	0.34	0.26
Doing chores around the house	0.24	**0.43**	0.43	0.16	0.29	0.25	**0.46**	0.45	0.23	0.31	0.32	0.33	0.24	0.27	0.34	0.27
Getting aches or aches	0.32	0.25	0.35	**0.57**	0.33	0.32	0.27	0.37	**0.58**	0.34	0.25	0.26	**0.39**	0.16	0.16	0.27
Low energy level	0.28	0.21	0.37	0.40	0.41	0.29	0.22	0.38	0.40	0.42	0.26	0.29	0.28	0.26	0.25	0.26
Feeling afraid or scared	0.14	0.26	0.32	0.26	**0.50**	0.16	0.29	0.33	0.28	**0.53**	0.13	0.17	0.21	0.28	0.18	0.25
Feeling sad or blue	0.16	0.21	0.35	0.32	**0.59**	0.18	0.22	0.38	0.33	**0.62**	0.14	0.26	0.31	0.29	0.25	0.29
Feeling angry	0.12	0.27	0.32	0.22	0.43	0.13	0.29	0.35	0.23	0.45	0.20	0.25	0.21	0.26	0.29	0.31
Trouble sleeping	0.19	0.25	0.37	0.32	0.42	0.21	0.26	0.38	0.34	0.45	0.23	0.29	0.33	0.24	0.29	0.35
Worrying about what will happen to him or her	0.15	0.18	0.32	0.31	**0.55**	0.17	0.19	0.33	0.34	**0.59**	0.21	0.26	0.25	0.33	0.28	0.28
Getting along with other children	0.14	0.28	0.33	0.19	0.32	0.16	0.29	0.35	0.19	0.34	0.25	**0.34**	0.20	**0.48**	0.39	0.22
Other kids not wanting to be his or her friend	0.16	0.29	0.35	0.20	0.36	0.18	0.30	0.36	0.21	0.37	0.23	0.27	0.14	**0.36**	0.34	0.22
Getting teased by other children	0.14	0.24	0.31	0.21	0.36	0.16	0.25	0.32	0.23	0.39	0.10	0.17	0.10	**0.22**	0.23	0.15
Not being able to do things that other children his or her age can do	0.31	0.43	**0.48**	0.28	0.33	0.33	0.44	**0.50**	0.30	0.37	0.40	0.41	0.24	0.53	0.57	0.33
Keeping up when playing with other children	0.29	0.37	**0.43**	0.25	0.30	0.30	0.39	**0.45**	0.28	0.33	0.33	**0.38**	0.25	0.40	0.42	0.26
Paying attention in class	0.17	0.31	0.36	0.21	0.31	0.17	0.31	0.39	0.23	0.33	n/a	n/a	n/a	n/a	n/a	n/a
Forgetting things	0.18	0.30	0.38	0.26	0.35	0.20	0.31	0.40	0.28	0.37	n/a	n/a	n/a	n/a	n/a	n/a
Keeping up with schoolwork	0.16	0.31	0.39	0.22	0.33	0.18	0.31	0.40	0.24	0.35	0.38	0.37	0.23	0.35	0.45	0.23
Missing school because of not feeling well	0.24	0.21	0.38	0.36	0.34	0.25	0.22	0.39	0.38	0.35	0.22	0.24	0.20	0.22	0.24	0.21
Missing school to go to doctor or hospital	0.29	0.28	0.39	0.35	0.30	0.30	0.29	0.41	0.37	0.31	0.37	0.35	0.21	0.33	0.32	0.23

Note: **Bold** indicates a higher correlation was hypothesised. Green cells indicate strong correlation (≥0.5); yellow cells indicate moderate correlation (0.3 to 0.49); and white cells indicate weak correlation (0.1-0.29).

4. Discussion

We have outlined 1) the quality of data collected in the P-MIC study, 2) the acceptability and feasibility of three EuroQol instruments (EQ-5D-Y-3L, EQ-5D-Y-5L, and TANDI) alongside the most commonly used generic paediatric measure (PedsQL core generic 4.0) to parents and children, and 3) the construct validity of instruments, by child age, report type, data collection type, and online panel sample type. The results demonstrate that the sample of children recruited via the hospital were more chronically unwell, had a higher follow-up survey completion rate, and had fewer quality issues when compared to the online panel samples. This has implications for both data collection modes, and the assessment of data collected online that are described below. In terms of feasibility and acceptability, the EuroQol instruments were quicker and participants reported they were easier to complete compared to the PedsQL. However, EuroQol instruments had larger ceiling effects compared to the PedsQL. All instruments demonstrated known group validity. Correlations of EuroQol instrument items with PedsQL items were generally in the directions hypothesised. Where expected, there was also evidence of divergence between PedsQL and EuroQol instruments.

The P-MIC study builds on previous work done to compare adult HRQoL instruments in a MIC study and is the first of its kind worldwide to collect common paediatric HRQoL instruments head-to-head concurrently across a wide range of child ages, conditions, and settings.(39) It allows for a direct comparison of both generic and condition specific instruments within the same dataset which is currently missing from the literature.(16, 17) The P-MIC study also has a large sample that includes children and families from a range of socioeconomic groups, geographic locations in Australia, and cultural groups, and provides results that are both generalisable to the Australian population and allow for exploration of important subgroups. The P-MIC study was designed with strong sample quality assessment procedures, resulting in high-quality data from both hospital and online samples to enable robust conclusions.(30) Due to survey logistics, it was required that participants answer all questions, and hence we were not able to determine which instrument items resulted in missing data (a possible proxy for relevance or acceptability to children and caregivers. However, this has been explored in previous research. (10, 11, 16-21)

Compared to the LSAC Australian population norm, the general population sample recruited online (Sample 2) had very similar sociodemographic composition, however, did have a lower proportion of children with special healthcare needs (16.2% vs 8.2%).(35) This difference is likely because of the hierarchical nature of the online panel recruitment, where children were first screened for one

of the nine common condition groups (Sample 3) and only if they did not meet the screening for the condition groups were they recruited to the general population sample (Sample 2).(31) It is acknowledged that other recruitment approaches to data collection could have been employed and resulted in slightly different characteristics, however, a hierarchical approach was used to ensure that the more difficult to reach condition group sample targets were met.

Compared to time to complete estimates provided by instrument developers for the PedsQL and EQ-5D-Y, participants in the P-MIC study completed instruments more quickly. The PedsQL administration guidelines note that the estimated time to complete is 5 minutes,(40) whilst participants in the P-MIC study completed the PedsQL in a median time of 1.6 minutes and 1.4 minutes for children aged 5 to 18 years and 2 to 3 years, respectively. The user guide for the EQ-5D-Y states that the instrument takes a few minutes to complete and a previous study noted the EQ-5D-Y-3L was completed in less than 3 minutes, whilst participants in the P-MIC completed the EQ-5D-Y-3L (inc VAS) in a median time of 46.2 seconds and the EQ-5D-Y-5L in a median time of 28.8 seconds.(22, 41) Estimates provided by instrument developers do not distinguish between administration mode and setting, hence the quicker completion times noted in this study may be due to the online data collection mode. However, instrument developers may also be conservatively estimating time to complete and future instrument users should consider that these time estimates may be an over estimation if administering online. Faster completion times and greater ease of completion were reported for younger children, proxies, online panels, and general population children compared to their counterparts. The faster completion and greater ease completing for younger children may be because proxies are completing on behalf of younger children, additionally, proxy adults may find instruments quicker and easier to complete compared to children. The faster completion and greater ease completing in the general population compared to the condition groups may be that the general population sample do not need to so carefully consider which level to choose as they are more likely selecting the lowest severity/frequency outcome level for each item. Additionally, although the EQ-5D-Y-5L was faster and easier to complete compared to the EQ-5D-Y-3L, it is likely this is because the 3L also included the VAS whereas the 5L did not.

This study identified large ceiling effect issues in EuroQol instruments at both the item and instrument level, which is consistent with several previous study findings.(11, 19, 21, 42) The differences in instrument recall period (today versus one month) and response level wording (severity versus frequency) may contribute to why we see fewer items with large ceiling effects in

PedsQL compared to EuroQol instruments. In adults, the EQ-5D-5L is preferred over the 3L given the reduction in overall ceiling effects.(43) Although a previous study noted no difference in ceiling effects between the EQ-5D-Y-3L and 5L,(21) this study demonstrated the EQ-5D-Y-5L had slightly lower overall ceiling effects compared to the 3L (30.2% versus 36.3%). These overall ceiling effects, even in the EQ-5D-Y-5L, are higher than the 15% threshold outlined by guidelines.(36)

The results suggest that the PedsQL, EQ-5D-Y-3L, EQ-5D-Y-5L, and TANDI have benefits in the assessment of paediatric HRQoL. Divergence between EuroQol instruments and the PedsQL was found, indicating the instruments may measure different dimensions of HRQoL which has been explored in more detail in a recent study.(44) The lack of statistically strong correlations between EuroQol items and PedsQL items where relationships could be due to differences in the item used to assess the broader underlying construct. For example, there are differences in the response options (severity versus frequency), the way in which supposedly overlapping constructs of HRQoL are measured, the item format (statements versus likert scale items) and the recall period (today vs one month). The instruments all indicate the tendency to be sensitive to known group differences. However, these results need to interpreted in light of the limited indicators of known groups used, and also the use of total sum scores on the HRQoL instruments to indicate differences. The total score on the PedsQL is an accepted way to score responses.(20) Using the total sum score of EuroQol instruments is more challenging given the aggregation of multiple single item dimensions into a single indicator, but it is notable that the effect sizes ae of similar magnitude. This was done due to the lack of established non preference and preference-based scoring for all three paediatric EuroQol instruments. Future work on the P-MIC data will explore known group validity in more detail, for example by assessing differences based on condition specific indicators, and using both preference and non-preference based scores.

The results presented here have implications for the use and choice of instruments in the design of studies, and use of HRQoL data in decision making. This is because they suggest that including only one instrument may result in important aspects of HRQoL potentially not being measured. The choice of instrument to use in studies may be guided by the aspects of HRQoL that are impacted by a condition or treatment. If possible, those designing studies may consider including multiple instruments to ensure a more holistic assessment of HRQoL and to enable sensitivity analysis of HRQoL outputs such as economic evaluation where different instruments are used. However, it is recognised that this is not always possible, and therefore instrument users should consider the intended purpose of the study, as well as the key HRQoL constructs that require measurement, and

prioritise outcomes based on this. For example, the PedsQL currently provides a detailed profile assessment of HRQoL, and can be mapped to utilities for input into health technology assessment (HTA).(45) The aim of the EQ-5D instruments is not to provide a detailed profile assessment of HRQoL, but to provide a brief profile, and elicit values, and work is ongoing to develop value sets. These different features, and how they impact the HRQoL data collected, should also be considered in the choice of instrument. They also mean that the comparison is limited and should be interpreted with caution. Our analysis has focused on comparing scores at the item and instrument level. Therefore, comparing an instrument with 23 items to an instrument with five is likely to result in divergence, and we have not generated results to fully inform the use of each instrument in HTA. When available, further work should also compare EQ-5D-Y-3L and Y-5L value sets with the PedsQL mapped utilities to understand the benefits of each in more detail.

The results about sample quality also highlight broader concerns with the quality of online panel data collection that are important considerations in the future collection of online self report and valuation data. For example, the P-MIC study included a follow-up survey that enabled additional validation questions to be included that were successful at identifying illegitimate responses. If this were a single time point study, these illegitimate responses would not be able to be identified. The quality procedures in the P-MIC emphasise the importance of quality checking samples, particularly when collecting data via online panels. The data quality procedures established in this study were transparently reported and can be used as a reference for future studies collecting HRQoL data via online panels.

5. Conclusion

This study outlines the quality of the data collected in the P-MIC study, highlighting the importance of quality checking procedures particularly when recruiting via online panels, to ensure high quality data collection. This study identified that all EuroQol's paediatric instruments for children (EQ-5D-Y-3L and EQ-5D-Y-5L for age 5 to 18 years and TANDI age 2 to 3 years) are acceptable, feasible and demonstrate known group validity when compared to the PedsQL across a broad range of child ages and conditions. Compared to the PedsQL, EuroQol's paediatric instruments have the advantage of being quicker and easier to complete. A disadvantage of EuroQol's paediatric instruments when compared to the PedsQL is their ceiling effects. This study is just one piece of the puzzle; future research is planned on how these instruments compare

across a wider array of psychometric properties, as well as how these instruments compare to other commonly used generic and condition-specific instruments.

6. Next steps for the P-MIC

The P-MIC study is a rich data set that includes other generic HRQoL instruments, condition-specific instrument and child health condition sub-groups not included in this analysis. The P-MIC study team have future analysis planned to further develop this evidence base, including further psychometrics analysis of all generic and condition specific HRQoL instruments included in the P-MIC (at the self-report and utility level), examination of instrument performance across child ages and health conditions, examination of instrument responsiveness, dimensionality analysis,(44) Item Response Theory (IRT) analysis, and EQ-HWB validity testing. The P-MIC team are interested in further collaborations with EuroQol members, in particular, extending the data collection into other countries.

7. References

1. Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health services insights. 2013;6:61-8.
2. Kromm SK, Bethell J, Kraglund F, Edwards SA, Laporte A, Coyte PC, et al. Characteristics and quality of pediatric cost-utility analyses. Quality of Life Research. 2012;21(8):1315-25.
3. Black N, Jenkinson C. Measuring patients' experiences and outcomes. Bmj. 2009;339:b2495.
4. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. Bmj. 2010;340:c186.
5. Black N. Patient reported outcome measures could help transform healthcare. BMJ : British Medical Journal. 2013;346:f167.
6. Higginson IJ, Carr AJ. Using quality of life measures in the clinical setting. BMJ. 2001;322(7297):1297-300.
7. Al Sayah F, Jin X, Johnson JA. Selection of patient-reported outcome measures (PROMs) for use in health systems. Journal of patient-reported outcomes. 2021;5(Suppl 2):99.
8. Economic Evaluation in Child Health: Oxford University Press; 2009.
9. 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):325-40.
10. 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.
11. Verstraete J, Ramma L, Jelsma J. Validity and reliability testing of the Toddler and Infant (TANDI) Health Related Quality of Life instrument for very young children. Journal of Patient-Reported Outcomes. 2020;4(1):94.
12. 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):887-97.
13. 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 : an

international journal of quality of life aspects of treatment, care and rehabilitation. 2019;28(7):1951-61.
14. Feng Y-S, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Quality of Life Research. 2021;30(3):647-73.
15. Astrom M, Krig S, Ryding S, Cleland N, Rolfson O, Burstrom K. EQ-5D-Y-5L as a patient-reported outcome measure in psychiatric inpatient care for children and adolescents - a cross-sectional study. Health and quality of life outcomes. 2020;18(1):164.
16. Rowen D, Keetharuth AD, Poku E, Wong R, Pennington B, Wailoo A. A Review of the Psychometric Performance of Selected Child and Adolescent Preference-Based Measures Used to Produce Utilities for Child and Adolescent Health. Value in Health. 2021;24(3):443-60.
17. 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.
18. Verstraete J, Marthinus Z, Dix-Peek S, Scott D. Measurement properties and responsiveness of the EQ-5D-Y-5L compared to the EQ-5D-Y-3L in children and adolescents receiving acute orthopaedic care. Health and Quality of Life Outcomes. 2022;20(1):28.
19. Verstraete J, Lloyd AJ, Jelsma J. Performance of the Toddler and Infant (TANDI) Health-Related Quality of Life Instrument in 3–4-Year-Old Children. Children. 2021;8(10):920.
20. Varni JW, Seid M, Kurtin PS. PedsQL™ 4.0: Reliability and Validity of the Pediatric Quality of Life Inventory™ Version 4.0 Generic Core Scales in Healthy and Patient Populations. Medical Care. 2001;39(8).
21. Fitriana TS, Purba FD, Rahmatika R, Muhaimin R, Sari NM, Bonsel G, et al. Comparing measurement properties of EQ-5D-Y-3L and EQ-5D-Y-5L in paediatric patients. Health and Quality of Life Outcomes. 2021;19(1):256.
22. 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. Quality of Life Research. 2013;22(1):173-83.
23. Lamping DL, Schroter S, Marquis P, Marrel A, Duprat-Lomon I, Sagnier PP. The community-acquired pneumonia symptom questionnaire: a new, patient-based outcome measure to evaluate symptoms in patients with community-acquired pneumonia. Chest. 2002;122(3):920-9.
24. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.
25. Åström M, Krig S, Ryding S, Cleland N, Rolfson O, Burström K. EQ-5D-Y-5L as a patient-reported outcome measure in psychiatric inpatient care for children and adolescents – a cross-sectional study. Health and Quality of Life Outcomes. 2020;18(1):164.
26. Desai AD, Zhou C, Stanford S, Haaland W, Varni JW, Mangione-Smith RM. Validity and Responsiveness of the Pediatric Quality of Life Inventory (PedsQL) 4.0 Generic Core Scales in the Pediatric Inpatient Setting. JAMA Pediatrics. 2014;168(12):1114-21.
27. 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;30:140-7.
28. Scott D, Ferguson GD, Jelsma J. The use of the EQ-5D-Y health related quality of life outcome measure in children in the Western Cape, South Africa: psychometric properties, feasibility and usefulness - a longitudinal, analytical study. Health and quality of life outcomes. 2017;15(1):12-.
29. Jones R, Mulhern B, McGregor K, Yip S, Loughlin R, Devlin N, et al. Psychometric Performance of HRQoL Measures: An Australian Paediatric Multi-Instrument Comparison Study Protocol (P-MIC). Children. 2021;8(8):714.
30. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Quality of life research : an

international journal of quality of life aspects of treatment, care and rehabilitation. 2021;30(8):2197-218.
31. Jones R, Mulhern B, Devlin N, Hiscock H, O'Loughlin R, McGregor K, et al. Australian Paediatric Multi-Instrument Comparison (P-MIC) Study: Technical Methods Paper [Online]. Melbourne, Australia; 2022. Available from: https://www.quokkaresearchprogram.org/_files/ugd/ea8836_9b1a5d92b8ce4a089159522e5d3e8903.pdf.
32. 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.
33. Verstraete J, Ramma L, Jelsma J. Validity and reliability testing of the Toddler and Infant (TANDI) Health Related Quality of Life instrument for very young children. Journal of patient-reported outcomes. 2020;4(1):94-.
34. StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC; 2019.
35. Edwards B. Growing Up in Australia: the Longitudinal Study of Australian Children: entering adolescence and becoming a young adult. Family Matters. 2014(95):5-14.
36. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology. 2007;60(1):34-42.
37. Cohen J. A power primer. Psychol Bull. 1992;112(1):155-9.
38. Sullivan GM, Feinn R. Using Effect Size—or Why the P Value Is Not Enough. Journal of Graduate Medical Education. 2012;4(3):279-82.
39. Richardson J, Iezzi A, Maxwell A. Cross-national comparison of twelve quality of life instruments: MIC Paper 1 Background, questions, instruments.; 2012.
40. Varni JW. PedsQL Administration Guidelines. 1998. Available from: www.pedsql.org/pedsqladmin.html.
41. EuroQol Research Foundation. EQ-5D-Y User Guide. 2020. Available from: https://euroqol.org/publications/user-guides/.
42. 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):887-97.
43. Janssen MF, Bonsel GJ, 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):675-97.
44. Bahrampour M, Jones R, Dalziel K, Devlin N, Mulhern B. What Does the EQ-5D-Y-5L Measure in Comparison to other Generic Pediatric Health Related Quality of Life Instruments? A Dimensionality Assessment Using Factor Analysis. Paper presented at the Scientific Plenary Meeting of the EuroQol Group; Chicago2022.
45. Khan KA, Petrou S, Rivero-Arias O, Walters SJ, Boyle SE. Mapping EQ-5D utility scores from the PedsQL™ generic core scales. Pharmacoeconomics. 2014;32(7):693-706.