



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Jones, R;Mulhern, B;McGregor, K;Yip, S;O'Loughlin, R;Devlin, N;Hiscock, H;Dalziel, K

Title:

Psychometric performance of HRQoL measures: An Australian Paediatric multi-instrument comparison study protocol (P-MIC)

Date:

2021-08-01

Citation:

Jones, R., Mulhern, B., McGregor, K., Yip, S., O'Loughlin, R., Devlin, N., Hiscock, H. & Dalziel, K. (2021). Psychometric performance of HRQoL measures: An Australian Paediatric multi-instrument comparison study protocol (P-MIC). *Children*, 8 (8), <https://doi.org/10.3390/children8080714>.

Persistent Link:



<https://hdl.handle.net/11343/287808>

License:

[CC BY](#)

Article

Psychometric Performance of HRQoL Measures: An Australian Paediatric Multi-Instrument Comparison Study Protocol (P-MIC)

Renee Jones ¹, Brendan Mulhern ², Kristy McGregor ¹, Shilana Yip ¹, Rachel O'Loughlin ³, Nancy Devlin ³, Harriet Hiscock ^{1,4,5}, Kim Dalziel ^{1,3,*} and on behalf of the Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team [†]

- ¹ Health Services, Murdoch Children's Research Institute, Melbourne 3052, Australia; renee.jones@mcri.edu.au (R.J.); kristy.mcgregor@mcri.edu.au (K.M.); shilana.yip@mcri.edu.au (S.Y.); Harriet.Hiscock@rch.org.au (H.H.)
- ² Centre for Health Economics Research and Evaluation, University of Technology Sydney, Ultimo 2007, Australia; Brendan.Mulhern@chere.uts.edu.au
- ³ Centre for Health Policy, Health Economics Unit, The University of Melbourne, Melbourne 3053, Australia; rneely@student.unimelb.edu.au (R.O.); nancy.devlin@unimelb.edu.au (N.D.)
- ⁴ Health Services Research Unit, Royal Children's Hospital, Melbourne 3052, Australia
- ⁵ Department of Paediatrics, The University of Melbourne, Melbourne 3010, Australia
- * Correspondence: kim.dalziel@unimelb.edu.au; Tel.: +61-4015-91310
- † Membership of the "Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team" is provided in the Acknowledgements.



check for updates

Citation: Jones, R.; Mulhern, B.; McGregor, K.; Yip, S.; O'Loughlin, R.; Devlin, N.; Hiscock, H.; Dalziel, K.; on behalf of the Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team. Psychometric Performance of HRQoL Measures: An Australian Paediatric Multi-Instrument Comparison Study Protocol (P-MIC). *Children* **2021**, *8*, 714. <https://doi.org/10.3390/children8080714>

Academic Editor: Wendy J. Ungar

Received: 23 July 2021

Accepted: 17 August 2021

Published: 20 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: Background: There is a lack of psychometric evidence about pediatric health-related quality of life (HRQoL) instruments. Evidence on cost effectiveness, involving the measurement of HRQoL, is used in many countries to make decisions about pharmaceuticals, technologies, and health services for children. Additionally, valid instruments are required to facilitate accurate outcome measurement and clinical decision making. A pediatric multi instrument comparison (P-MIC) study is planned to compare the psychometric performance and measurement characteristics of pediatric HRQoL instruments. Methods: The planned P-MIC study will collect data on approximately 6100 Australian children and adolescents aged 2–18 years via The Royal Children's Hospital Melbourne and online survey panels. Participants will complete an initial survey, involving the concurrent collection of a range of pediatric HRQoL instruments, followed by a shorter survey 2–8 weeks later, involving the collection of a subset of instruments from the initial survey. Children aged ≥ 7 years will be asked to self-report HRQoL. Psychometric performance will be assessed at the instrument, domain, and item level. Conclusions: This paper describes the methodology of the planned P-MIC study, including benefits, limitations, and likely challenges. Evidence from this study will guide the choice of HRQoL measures used in clinical trials, economic evaluation, and other applications.

Keywords: pediatrics; quality of life; cost-benefit analysis; psychometrics; chronic disease; mental health; patient reported outcome measures



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The increasing recognition of the importance of including the child's perspective on their health and wellbeing in decision making has resulted in a large increase in the availability and use of pediatric health-related quality of life (HRQoL) instruments [1,2]. Accurate measurement and valuation of HRQoL among children plays a vital role in the development of robust evidence for economic evaluation and population health studies, as well as routine outcomes measurement, clinical decision making and patient choice [1,3].

Economic evaluation frequently incorporates preference-based HRQoL information to produce estimates of Quality Adjusted Life Years (QALYs). QALYs combine HRQoL impacts and survival impacts in a single metric, allowing comparisons across health problems,

interventions, and populations. QALYs are based on patients' HRQoL which is usually captured with validated survey instruments. These instruments comprise a descriptive system covering different HRQoL aspects, which is weighted and scored using QALY weights (or utility values) derived from the stated preferences, usually of the general population for each health state. QALYs inform cost-effectiveness evidence which is formally used in many countries to make decisions about public funding for pharmaceuticals, technologies, and health services. In Australia, bodies such as the Pharmaceutical Benefits Advisory Committee (PBAC) and Medical Services Advisory Committee (MSAC) use evidence about cost effectiveness to inform decisions about the best use of health care resources [4]. These decisions have large cost implications, with approximately \$12.6 and \$25 billion spent through the Pharmaceutical Benefits Scheme (PBS) and Medicare Benefits Schedule (MBS), respectively, in Australia in 2019–2020 [5,6]. These are federally funded reimbursement schemes ensuring low or no cost to consumers, and similar reimbursement processes are seen around the world.

There is increasing interest in the use of HRQoL information for patient reported outcome measures (PROMs) in clinical practice. In the clinical setting, HRQoL instruments can be used to aid identification of health problems, including unexpected or hidden problems; decision making, including prioritization of health problems or treatments; monitoring of changes in health over time; communication; and patient decision making, including decision making about treatments [7]. PROMs information has been shown to improve HRQoL outcomes and satisfaction with pediatric clinical care [3,8,9]. In this context it is important that the HRQoL instrument is sensitive to change, informative to clinicians and patients, consistent with a patient's clinical history and allowing for a short collection time. However, there is a paucity of literature on the psychometric properties, feasibility and validity of potential PROM instruments. A 2015 systematic review identified 35 generic paediatric PROMs with none having all their psychometric properties assessed [9]. Eight of the 35 measures identified were preference-based measures [9].

Although there are established methods for measuring and valuing HRQoL in adults [10], considerable challenges arise in applying these methods to children, in particular very young children [11]. For instance, the developmental range of children precludes a 'one size fits all' approach and younger children are unable to self-report their HRQoL, requiring proxy completion (usually by parents) [11].

A 2015 systematic review examining the application of preference-based HRQoL instruments in pediatric populations identified nine HRQoL instruments that have been used in pediatric populations (Adolescent Health Utility Measure, Assessment of Quality of Life (AQoL-6D), Child Health Utility (CHU9D), EQ-5D-Y-3L, Health Utilities Index Mark 2 (HUI2) and Mark 3 (HUI3), 16D, 17D, and the Quality of Wellbeing Scale (QWB)) and concluded that more evidence on the performance of these instruments is required [12]. Additionally, a 2021 systematic review examining the psychometric performance of key generic pediatric preference-based instruments (CHU9D, EQ-5D-Y-3L, HUI2, and HUI3) also identified gaps in evidence on the psychometric performance of pediatric HRQoL instruments. Key limitations in the currently available literature include the small sample size of some studies, wide variation in methods, a focus on single instruments, limited coverage and heterogeneity of disease conditions or age groups, meaning that broad synthesis of evidence is challenging [13]. These limitations have resulted in ambiguity regarding which pediatric HRQoL instruments are most appropriate for which context, inhibiting clinicians in instrument use for clinical care, researchers in evidence collection and decision makers in evidence appraisal. No study has previously compared a wide set of generic HRQoL measures accompanied by preference weights concurrently in the same cohort of children (also called a pediatric multi instrument comparison (P-MIC) study).

There is a clear evidence deficit regarding the comparative performance of pediatric HRQoL instruments. Considering the large implications on resource allocation and clinical care, addressing this evidence deficit provides the opportunity to improve value for money across the health care system, resulting in improved population health outcomes for

children. This study aims to generate new evidence on the comparative performance of available pediatric HRQoL instruments. This will be done by assessing the validity and reliability of instruments across a range of age and disease groups in a single study. This evidence will improve and inform users' choice of pediatric HRQoL instruments in Australia and around the world. This paper aims to describe the study methodology of the proposed study.

2. Materials and Methods

2.1. Study Aims

To compare the performance of a range of pediatric HRQoL instruments in terms of validity, reliability, responsiveness, acceptability, feasibility, and consistency across age and disease groups.

2.2. Study Design

This P-MIC study will involve the prospective collection of multiple generic and disease-specific paediatric HRQoL instruments concurrently via two online surveys. The first is a 15–30-min survey, and the second is a five-minute follow-up survey two to eight weeks later. Collecting these paediatric HRQoL instruments concurrently across a range of age and disease groups will allow the relative performance and measurement characteristics of these instruments in children to be assessed. This study was approved by The Royal Children's Hospital (RCH) Human Research Ethics Committee (HREC/71872/RCHM-2021) on the 20th May 2021 and is prospectively registered with the Australia New Zealand Clinical Trials Registry (ANZCTR) (ACTRN12621000657820).

2.3. Timeline

Recruitment for this study began on the 4th of June 2021 and is expected to be completed by January 2022. Data analysis and publication is expected to be ongoing, with main results completed by December 2022.

2.4. Participants

Any parent, caregiver, or guardian of a child(ren) aged 2–18 years (inclusive) at study enrolment is eligible to take part. Participants unable to communicate in written English, unable to answer or comprehend the questions or who do not reside in Australia will be excluded from participating. Additional eligibility criteria apply to the disease-specific samples (see Section 2.6.1 for further details).

The study aims to collect data on 6100 Australian children via the recruitment of parents/caregivers. The sample will be split into three groups:

- Sample one, recruited via hospital: 1000 parents/caregivers will be recruited via The RCH, Melbourne, Australia, allowing data to be collected from children with a wide range of health conditions and severities.
- Sample two, online panel population sample: 1500 parents/caregivers of children from a general population sample will be recruited via an online survey company, Pureprofile Australia, allowing data to be collected from a large population sample. This sample will be further divided into two groups:
 - Sample 2a, online panel population sample, normal follow-up: 1300 parent/caregivers will be followed-up in line with the rest of the cohort, two to eight weeks after initial survey.
 - Sample 2b, online panel population sample, short follow-up: to assess reliability, 200 parents/caregiver will be followed up at a shorter interval, two days after initial survey.
- Sample three, online panel disease-specific groups sample: 3600 parents/caregivers of children with one of nine health conditions will be recruited via an online survey company, Pureprofile Australia. Disease groups were chosen that had sufficient prevalence to allow for recruitment of groups and with a focus on conditions with

larger expected HRQoL decrements [14]. Currently six of the disease groups have been confirmed and include: attention-deficit/hyperactivity Disorder (ADHD), anxiety and/or depression, autism spectrum disorder (ASD), asthma, dental decay, and sleep problems. An additional three disease groups will be selected from the following: diabetes, epilepsy, eating disorder, recurrent abdominal pain, frequent headaches. We aim to collect data on 400 unique children from each of the nine disease groups.

2.5. Recruitment

2.5.1. Sample One, Sample Recruited via Hospital ($n = 1000$)

Research Assistants (RAs) will approach parents/caregivers for recruitment from a range of RCH departments, including outpatient clinic waiting rooms, surgical department waiting rooms and the Emergency Department Short Stay Unit (SSU) and provide them with a link to the initial survey. The SSU will provide a list of approved families for RAs to approach on the day when recruiting from this department to ensure appropriate families are approached and to minimize the disruption to care and avoidable distress. Poster adverts with QR codes to the initial survey will be placed in high traffic areas of the hospital. Online advertisements with a link to the initial survey will be placed on hospital telehealth appointments (this advert will appear for any family attending a hospital appointment via telehealth), newsletters, and social media. Recruitment resources will also be utilised to target hospital departments where children likely to be experiencing higher decrements in quality of life, such as children with rare genetic conditions, children with serious sequelae arising across childhood following birth at an extremely low birthweight, and children admitted to the Paediatric Intensive Care Unit (PICU). Potential participants from PICU will be approached by PICU research staff at their pre-operative clinic visit to minimize the recruitment in inappropriate situations (e.g., where the child is unlikely to survive) and to minimize stress to families of children with an unplanned admission. If needed, we may also extend to other areas of the hospital or recruit via other methods such as sharing the advertisement with parent/caregivers from the onsite RCH childcare centre; with participants from a previous study who have indicated they would like to be contacted for future research; sending letters to potential participants from participating hospital departments; and approaching families in the local RCH playground.

2.5.2. Sample Two (Online Panel Population Sample, $n = 1500$) and Sample Three (Online Panel Disease Groups Sample, $n = 3600$)

The recruitment of the online panels will be managed by Pureprofile Australia. All existing members have completed a double opt-in process to join the panel and have agreed to complete online surveys over the course of their membership. They are provided with the opportunity to accept an offer to participate in new surveys. Participants will be randomly selected from this panel to take part in this study if they meet eligibility criteria and fulfil the health status/disease-specific quotas that have been prespecified. Participants will be selected based on quotas for age and sex. As children may have multiple conditions, participants for sample three will be invited to take part on a least fill basis, whereby, children with rarer conditions will be invited to take part for the rarer condition disease group where they will only receive the disease specific measure for this condition even if they report their child has another of the disease specific conditions.

2.6. Data Collection

Data will be collected via two online REDCap surveys. Table 1 provides a schedule of which instruments will be collected at which time points for each sample. Appendix A provides a justification for and description of instruments included in the study.

Table 1. Summary of instruments by study sample.

| Instrument | Sample 1, Sample Recruited via Hospital | | Sample 2, Online Panel Population Sample | | Sample 3, Online Disease Group Sample | |
|---|---|-----------|--|-----------|---------------------------------------|-----------|
| | Initial | Follow-Up | Initial | Follow-Up | Initial | Follow-Up |
| Demographic and non-HRQoL instruments | | | | | | |
| Sociodemographic questions | X | | X | | X | |
| SDQ | X | | X | | X | |
| EQ-HWB | X | | X | | X | |
| Core HRQoL instruments | | | | | | |
| PedsQL | X | X | X | X | X | X |
| TANDI (if <=3 years) | X | X | X | X | X | X |
| EQ-5D-Y 3L & 5L original (if >=5 years) | X | X | X | X | X | X |
| EQ-5D-Y 3L & 5L adapted (if <=4 years) | X | X | | | | |
| EQ-5D-Y 3L original and adapted or EQ-5D-Y 5L original and adapted (if <=4 years) | | | X * | X * | X * | X * |
| CHU9D | X | X | X | X | X | X |
| Global Health Measure (single item) | X | X | X | X | X | X |
| Additional HRQoL instrument blocks | | | | | | |
| HUI 2/3 and EQ-5D-5L (>11 years) | | | X * | X * | X * | X * |
| AQoL-6D (>4 years) | | | X * | X * | X * | X * |
| PROMIS-25 (>4 years) | | | X * | X * | X * | X * |
| Disease specific instruments | | | | | | |
| Disease specific instruments (as per the recommended age range of the instrument) | | | | | X * | |

X- indicates the instrument will be collected from the sample/time point. * Participant will only receive, if allocated, instrument based on disease group, and/or randomization to receive additional instrument, and/or randomization to receive EQ-5D-Y 3L original and adapted or EQ-5D-Y 5L original and adapted. Abbreviations: HRQoL health-related quality of life, SDQ Strengths and Difficulties Questionnaire, EQ-HWB EQ Health and Wellbeing Short Version, PedsQL Paediatric Quality of Life Inventory, TANDI Toddler and Infant Questionnaire, EQ-5D-Y EQ-5D Youth, CHU9D Child Health Utility, HUI2/3 Health Utilities Index Mark 2/3, EQ-5D-5L, AQoL-6D Assessment of Quality of Life, PROMIS-25 Patient-Reported Outcome Measurement Information System 25.

2.6.1. Initial Survey

At the start of the initial survey, parents/caregivers will be asked screening question(s) to confirm their eligibility. Parents/caregivers from every sample will be first asked to confirm they are the parent, caregiver or guardian of a child aged 2–18 years. Parents/caregivers from sample three, the online panel disease groups, will undergo additional screening to ensure they are the parent/caregiver of a child with one of the nine health conditions (Appendix B). Currently six of the nine disease groups and the corresponding screening questions for these groups have been confirmed. Once the additional three disease groups have been confirmed, the corresponding screening questions will be selected for these groups. Following screening, all participants will be required to provide informed consent. The initial survey includes sociodemographic questions, the Strengths and Difficulties Questionnaire (SDQ), EQ Health and Wellbeing Short Version for parents (EQ-HWB-S) followed by several pediatric HRQoL instruments. HRQoL instruments will be blocked into three groups: core, additional and disease-specific. Participants will be randomized to additional instruments to minimize responder burden. Core instruments, which all participants will receive, include: the Global Health Measure, CHU9D, Paediatric Quality of Life Inventory (PedsQL) Core 4.0, EQ-5D-Y, and Toddler and Infant Questionnaire (TANDI). Participants in all samples, where the child is aged five years or older, will receive both the EQ-5D-Y 3L and 5L original version. Participants in sample one, where the child is four years or younger, will receive both the EQ-5D-Y 3L and 5L adapted versions with guidance notes. Participants from samples two and three, where the child is four years or younger, will be randomized to receive either (1) the EQ-5D-Y-3L adapted

version with guidance notes and original version or (2) the EQ-5D-Y-5L adapted version with guidance notes and original version. Additionally, participants from samples two and three, the online panel samples, will be randomized to receive one of three additional block(s) of instruments: (1) EQ-5D-5L & HUI2/3, (2) AQoL-6D or, (3) Patient-Reported Outcome Measurement Information System 25 (PROMIS-25) Paediatric Profile. Participants from sample three, the online disease group sample, will receive a corresponding disease-specific HRQoL instrument. Currently six of the nine disease groups have been confirmed and the corresponding disease specific instruments for these groups are: PedsQL asthma model (asthma), KIDSCREEN-27 (ASD), Sleep Disturbance Scale for Children (SDSC) (sleep problems), Child Perceptions Questionnaire (CPQ 11-14) (dental decay), The revised Children’s Anxiety and Depression Scale Short form (RCADS-25) (anxiety and/or depression), and Strengths and Weaknesses of ADHD Symptoms and Normal Behavior Scale (SWAN) (ADHD). Once the additional three disease groups have been confirmed, the corresponding disease specific measure will be selected for these groups. Disease-specific instruments are selected based on the following hierarchical criteria: (1) instrument validated in children, (2) disease-specific QoL instrument, (3) symptom scale or measure, (4) generic QoL instrument commonly used in, or appropriate for use in, children with the condition of interest. Appendix A provides the rationale for choosing each disease specific instrument.

Children aged seven years or older and identified by their parent/caregiver as being able to complete the survey will be invited to complete the HRQoL questions (self-report). If the child is under seven years of age or is unable to complete the survey (e.g., not cognitively able to complete or is undergoing a procedure in hospital) the parent/caregiver will be asked to complete the HRQoL questions on their behalf as proxy. Participants will be given HRQoL instruments to complete that match the child’s age, person reporting (proxy report vs. self-report) and child disease group (if part of the online disease panel) (see Figure 1).

| | | Child Age | | | | | | | | | | | | | | | | |
|---------------------------|--|--|---|------------|---|--|---|---|---|----------|----|---|----|----|----|----|----|----|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| Demographic and non-HRQoL | | Sociodemographic questions (parent/caregiver report) | | | | | | | | | | | | | | | | |
| | | SDQ 2-4 (parent/caregiver proxy-report) | | | | SDQ 5-10 (parent/caregiver proxy-report) | | | | | | SDQ 11-18 (Child self-report if able, otherwise proxy-report) | | | | | | |
| | | Carer QoL- EQ-HWB (parent self-complete) | | | | | | | | | | | | | | | | |
| Core HRQoL | | Parent/caregiver proxy-report | | | | | | Child self-report if able, otherwise proxy-report | | | | | | | | | | |
| | | PedsQL 2-4 | | PedsQL 5-7 | | PedsQL 5-7 if proxy or 8-12 if self | | PedsQL 8-12 | | | | PedsQL 13-18 | | | | | | |
| | | TANDI | | n/a | | | | | | | | | | | | | | |
| | | EQ-5D-Y 3L and 5L (adapted with guidance notes) | | | | EQ-5D-Y 3L and 5L | | | | | | | | | | | | |
| | | CHU9D (with guidance notes) | | | | CHU9D | | | | | | | | | | | | |
| Additional HRQoL* | | Global Health Measure | | | | | | | | | | | | | | | | |
| | | HUI 2/3 | | | | | | | | | | | | | | | | |
| | | n/a | | | | | | | | EQ-5D-5L | | | | | | | | |
| | | n/a | | | | AQoL-6D | | | | | | | | | | | | |
| Disease Specific | | n/a | | | | PROMIS-25 | | | | | | | | | | | | |
| | | Disease-specific instruments (will be applied as per the recommended age range of the instrument)* | | | | | | | | | | | | | | | | |

Figure 1. Summary of instruments by age. * Participant will only receive if allocated instrument based on disease group and/or randomization to receive additional instrument. Abbreviations: HRQoL health-related quality of life, SDQ Strengths and Difficulties Questionnaire, EQ-HWB EQ Health and Wellbeing Short Version, PedsQL Paediatric Quality of Life Inventory, TANDI Toddler and Infant Questionnaire, EQ-5D-Y EQ-5D Youth, CHU9D Child Health Utility, HUI2/3 Health Utilities Index Mark 2/3, EQ-5D-5L, AQoL-6D Assessment of Quality of Life, PROMIS-25 Patient-Reported Outcome Measurement Information System 25.

The order in which the HRQoL instruments are presented for completion will be randomized to minimize order and survey fatigue effects. In addition, the EQ-5D-Y-3L, EQ-5D-Y-5L, and, if relevant the EQ-5D-5L, will be presented to participants with another HRQoL instrument separating them, given their similarities. For participants in samples two and three, where the child is four years or younger and is receiving both the EQ-5D-Y (3L or 5L) original and adapted version, the original version will be displayed directly before the adapted version. The order of instruments will be the same for the initial and follow-up survey for each participant. The consent and sociodemographic questions will always appear first in the initial survey.

2.6.2. Follow-Up Survey

A shorter version of the initial survey will be sent out to the majority of participants two to eight weeks after completion of the initial survey. A small proportion ($n = 200$) of participants from sample two, the online population panel sample, will be selected at random and asked to complete the follow-up survey two days post initial survey completion, and this group will form sample 2b. The follow-up survey will comprise the core set of measures described above. Additionally, participants from samples two and three, will receive the same instrument(s) they were randomized to receive in the initial survey. It is expected to take approximately five to ten minutes to complete. For consistency, if the parent/caregiver completed HRQoL questions (proxy report) in the initial survey, they will be asked to again complete the questions on behalf of their child. If a child previously completed the HRQoL questions (self-report) in the initial survey they will be asked to complete them in the follow-up survey. No sociodemographic questions, SDQ or EQ-HWB questions will be included in the follow-up survey. Questions about change in the child's health status since the first survey will be included in the follow-up survey (See Appendix C).

2.7. Participant Duration and Reimbursement

The initial survey will mark the beginning of a participants' involvement in the study. Participants can enter the study at any point and are not required to be on the same schedule as other participants. The completion of the follow-up survey will mark the end of a participants' involvement in the study. Participants from sample one, the sample recruited via hospital, will be emailed a \$15 online gift voucher once they have completed the follow-up survey to reimburse them for their time. Participants from samples two and three, the online panel samples, will also be reimbursed for their time by Pureprofile Australia, the reimbursement dependent on the questionnaire length, ranging between \$3–\$15 for a 15–30-min survey. These reimbursed amounts are to compensate participants for some of their time, but are not seen as high enough to unduly coerce.

2.8. Sample Size

There is no commonly accepted methodology for calculating sample sizes for studies assessing psychometric properties and hence sample sizes are not often reported [15]. We completed outcome specific sample size calculations where possible to inform recruitment targets for key outcomes, responsiveness, and known-group validity. For all sample size calculations, we focused on the PedsQL measure as it is widely used and has established minimal clinically important difference (MCID) [16]. To ensure enough children were recruited with a change in health status between initial and follow-up surveys, we first calculated the estimated sample size required to assess responsiveness. Using a two-sided paired test with type I error of 1%, it was determined that a sample of 139 would give statistical power of 0.9 to detect the accepted MCID of 4.36 in the total score of the PedsQL self-report [16]. A sensitivity calculation was completed using proxy-report scores and determined that a sample of 190 would give statistical power of 0.9 to detect the accepted MCID of 4.5 in the total score of the PedsQL proxy-report using a two-sided paired test with type I error of 1% [16]. Additionally, we calculated the estimated sample size required

to assess known-group validity. Using a one-sided test with type I error of 1%, it was determined that a sample of 186 (93 per group) would give statistical power of 0.9 to detect a difference between healthy children and children with a chronic condition on the total score of the PedsQL proxy report [16]. A sensitivity calculation was completed using the PedsQL self-report scores and it was determined that a sample of 128 (64 per group) would give statistical power of 0.9 to detect a difference between healthy children and children with a chronic condition on the total score of the PedsQL self-report using a one-sided test with type I error of 1% [16]. Additional information on sample size calculations is available in Appendix D. The total sample size of 6100 was estimated from a similar adult study and a review of previous pediatric HRQoL studies [13,17]. Although the overall sample size is greater than the outcome specific sample sizes estimated above, this larger total sample is required to complete further item level analysis such as Item Response Theory (IRT) and factor analysis in addition to subgroup analysis, such as by disease group and child age. Specific sample size calculations have not been performed to support each of these scenarios.

2.9. Statistical Analysis

The psychometric properties of the pediatric HRQoL instruments will be analyzed at the overall, domain, dimension, and item levels. They will be tested and compared across the following outcomes: validity, reliability, responsiveness, acceptability and feasibility, and consistency [18]. We will also use Factor Analysis and Modern Test Theory methods such as IRT to understand what the instruments are measuring and the characteristics of individual items. Additionally, we will investigate how the outcomes described above vary by certain factors such as child age, gender, and disease group (including acute versus chronic conditions), family socio-economic status (SES), and presence of anxiety/depression comorbidity. HRQoL instruments will be scored to produce item scores, domain scores and a total utility score where appropriate using published scoring algorithms or value sets to enable further analysis of validity. To minimize missing data we will: (1) work with the online survey panel company to set rules regarding minimum acceptable data, (2) follow up with participants if there is missing survey data at the time of survey completion, and (3) use the functionality of REDCap to require questions to be answered, so that if a question is left blank the survey will prompt the participant to go back and complete the question. Any missing data will be managed by following instrument guidelines.

2.9.1. Validity

Validity is defined as the degree to which the HRQoL instrument measures the construct(s) it purports to measure [18]. Validity will be measured using within-scale analysis (measured using factor analysis), known group differences (measured by descriptively comparing a priori assumptions regarding expected differences between groups), convergent validity (measured by analyzing the correlation of similar constructs from different instruments) and discriminant validity (measured by analyzing whether dimension responses are independent of known groups) [18,19].

2.9.2. Reliability

Reliability is defined as the stability of a participant's responses on the instrument. Reliability will be assessed using test-retest reliability, measured by agreement on dimension-level responses between the initial survey and the re-test survey two days later [18,19]. This outcome will only be assessed in the $n = 200$ participants from sample 2b.

2.9.3. Responsiveness

Responsiveness is defined as the ability of an instrument to detect change in the construct to be measured [18]. Responsiveness of instruments will be assessed using dimension level responses from children whose proxy respondents reported a change in

general health status from the initial survey to the follow-up survey in comparison to those not reporting a change (see Appendix C) [18,19]. This will be assessed to determine the extent to which instruments are responsive to change in general health status.

2.9.4. Acceptability and Feasibility

Acceptability and feasibility will be measured by assessing the time to complete instruments [18,19], and self-reported difficulty (measured by a single, study-designed survey question asking participants to report how difficult the instrument was to complete). Survey missingness or drop-out will also be reported as indicators of acceptability and feasibility, noting the points at which participants prematurely exit the survey.

2.9.5. Consistency

Consistency is defined as the degree to which the summary and dimension specific responses on each instrument are consistent with dimension and summary responses of other similar instruments [18]. Consistency of instruments will be assessed by comparing the consistency of summary and dimension specific responses on each instrument using item-total correlations and Cronbach alpha coefficients [18,19].

2.9.6. Factor Analysis

To assess measurement overlap between instruments we will also use exploratory and confirmatory factor analysis. These techniques identify domains of items measuring similar constructs, and can be used for individual instruments, or to identify domains from a pool of items taken from different instruments.

2.9.7. Item Response Theory

We will use IRT to examine item performance at the dimension and domain level. IRT is a latent scale technique that allows for the assessment of item and model fit, information provided by items within dimensions, and differential item functioning across demographic groups. IRT will be used as an exploratory supplementary tool to add to the understanding of instruments, and the results will be considered in light of the appropriateness of IRT for the instrument (for example considering instrument structure) and the results from other psychometric methods.

2.10. Data Security

For sample one, the sample recruited via hospital, only identifying information required for the purposes of follow up will be collected (email and phone number). This identifying information will only be accessed by members of the research team who are required to contact participants for follow up. Data will be captured electronically via REDCap and securely stored in the Murdoch Children's Research Institute's (MCRI's) REDCap database system, which is backed up nightly to a local backup server, with a monthly backup taken to tape and stored offsite. REDCap maintains an audit trail of data create/update/delete events that is accessible to project users who are granted permission to view it. Access to REDCap will be provided via an MCRI user account or (for external collaborators) via a REDCap user account created by the MCRI system administrator. The permissions granted to each user within each REDCap project will be tightly controlled and in accordance with ethics approval.

3. Discussion

This study is the first to concurrently collect primary evidence from the same cohort of participants on a broad set of pediatric HRQoL instruments across a diverse sample that is adequately powered to compare the psychometric performance of multiple HRQoL instruments for children.

Potential challenges to this study are the recruitment of a large sample recruited via hospital during the COVID-19 pandemic. Recruitment for this sample is primarily

intended to come from face-to-face recruitment which will likely be impacted by any changes to the COVID-19 situation in Victoria, Australia. Additional recruitment strategies outlined in Section 2.5.1 have been planned to help minimize this risk. A limitation of this study is that minimal child level clinical information will be obtained via the survey and thus we will be unable to assess how instruments perform by factors such as disease severity. Responsiveness of HRQoL has traditionally been a difficult area of psychometric performance to assess and our ability to assess responsiveness relies on having between 139 and 190 children who change health status during the period of the study. There are two potential difficulties, firstly in recruiting this number of children who change health status. We have targeted conditions known to have a higher risk of change, nonetheless our ability to recruit adequate numbers of children who change health status is unknown. Secondly, there is no one gold standard method for measuring change in health status via self-reported data. We have relied on several questions aimed at identifying health status change (see Appendix C); however, it is unknown to what extent these questions will adequately identify change. This study can test these factors. Due to the large numbers of children to be recruited and the desire to seek children with a variety of conditions, it is unknown how many will be recruited with each condition in the sample recruited via hospital. Specific strategies to target more seriously ill children will be employed, but it is unknown what this balance will be prior to understanding how recruitment success progresses. Another limitation of the study is the decision around the length of follow-up. For consistency, a single follow-up period will be selected for all samples, excepting the test–retest sample (sample 2b). A period that is too short will perhaps not give some groups of children enough time to change their health status and this will be different for each condition. A period that is too long will place pressure on completion rates. There is likely to be no one right answer across all patient groups so a consistent compromise will be sought. In addition, once a follow up period is specified, it is unlikely that all participants will complete in the exact same amount of time. Variation by a couple of days/weeks either side of the set follow up time is expected which creates some variation in the data. However, this variation will be recorded to allow for its impact to be tested in analyses.

With limited resources and increasingly costly medications and technologies potentially available for children, it is important that decisions about the allocation of public funding for medications and technologies are informed by best possible evidence, including cost-effectiveness evidence. The accurate measurement of HRQoL is key to ensuring that utility as an input to cost-utility analysis is accurate, but currently there is a lack of sound evidence on the psychometric performance of paediatric HRQoL instruments. This lack of evidence is impacting the ability of decision makers to make informed choices about paediatric interventions. The data collected from this study provides an additional opportunity to psychometrically assess potential instruments for use as a routine clinical PROM in paediatric clinic settings, thereby filling this—and other gaps in the literature—including the lack of comparative data across instruments and child age and conditions. Future work from this study could look at international extensions and the development of psychometric testing protocols. Our work will allow for clear comparison of evidence across instruments [13] and will be of national and international significance.

Author Contributions: Conceptualization, K.D., N.D., B.M. and H.H.; methodology, K.D., N.D., B.M., H.H., R.J., R.O., S.Y. and K.M.; writing—original draft preparation, R.J., K.D., S.Y. and K.M.; writing—review and editing, R.J., K.D., N.D., B.M., H.H., R.O., S.Y. and K.M. All authors have read and agreed to the published version of the manuscript.

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. Prof Harriet Hiscock is supported by an NHMRC Practitioner Fellowship 1136222. Rachel O’Loughlin is supported by a Research Training Program Scholarship provided by the Australian Commonwealth Government and the University of Melbourne.

Institutional Review Board Statement: This study was approved by The Royal Children’s Hospital (RCH) Human Research Ethics Committee (HREC/71872/RCHM-2021) on the 20th May 2021.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank the QUOKKA study investigators and research fellows: Richard Norman, Rosalie Viney, Julie Ratcliffe, Deborah Street, Gang Chen, Cate Bailey, Christine Mpundu-Kaabwa, and Tessa Peasgood for their guidance in developing this study protocol. Additionally, we would like to sincerely thank members of the consumer advisory group and others who helped pilot, test and refine different elements of this study protocol.

Conflicts of Interest: K.D., N.D., B.M., H.H. have all received previous or current funding from the EuroQol Foundation who is the developer 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 and are also the developers of some of the instruments included in this study.

Appendix A. Overview of Instruments

Table A1. Overview of instruments with justification for inclusion in study.

| Instrument | Description | Application in This Study | Justification for Inclusion in P-MIC |
|----------------------------|---|---------------------------|---|
| Core non-HRQoL Instruments | | | |
| EQ-HWB-S v1.2 | Part of the EuroQol family of instruments, the EQ-HWB-S is a new experimental instrument designed to assess the impact of being a care recipient or caregiver on health and wellbeing [20]. The short version currently has 9 items. | See Figure 1 and Table 1. | Included following advice from consumer advisory group who felt it was important to include a carer. QoL measure due to the strong relationship between child and carer. QoL in children who have chronic conditions. The EQ-HWB-S was chosen because it is a promising new instrument that requires further validation work and is an opportunity to add to the evidence base. |
| SDQ | The SDQ is a well-validated behavioral screening questionnaire used to assess a child’s emotional and behavioral wellbeing [21,22]. It has 25 items measuring 5 domains [23]. Cut points on each domain can be used to indicate borderline or clinically elevated symptoms. Versions of the SDQ exist for children aged 2–17 years, with self-report available for children aged 11–17 years. | See Figure 1 and Table 1. | Included to capture emotional wellbeing of children to allow the exploration of how HRQoL instruments perform for children with a mental health problem or mental health comorbidity. The SDQ was chosen because it is well validated across a wide age range with cut points to indicate clinically relevant symptoms. |
| Core HRQoL Instruments | | | |
| PedsQL Generic Core 4.0 | The PedsQL Generic Core 4.0 is a 23 item HRQoL instrument covering 4 domains [24]. Validated version of the PedsQL exist for children aged 2–18 years with self-report available for children ≥ 8 years [24]. | See Figure 1 and Table 1. | Identified as priority for inclusion as it is a generic instrument widely used in pediatric populations with some evidence for performance, but requiring more. |
| TANDI v2 | The TANDI is a new proxy HRQoL instrument developed specifically for children ≤ 3 years based on the structure of the EQ-5D-Y-3L [25]. It consists of 6 items covering 6 domains [25]. | See Figure 1 and Table 1. | Identified as priority for inclusion because it is a promising new instrument for young children requiring further validation work. |
| EQ-5D-Y (3L and 5L) | The EQ-5D-Y-3L and recently developed EQ-5D-Y-5L, are youth versions adapted from the EQ-5D adult version [26,27]. Both the EQ-5D-Y-3L and EQ-5D-Y-5L are HRQoL instruments consisting of 5 items covering 5 domains. The EQ-5D-Y-3L has 3 severity levels and the 5L has 5. The EQ-5D-Y also includes a Visual Analogue Scale (VAS). Validated versions of the EQ-5D-Y exist for children aged 4–18 years, with self-report from ≥ 8 years. | See Figure 1 and Table 1. | The EQ-5D-Y-3L was identified as priority for inclusion as it is a generic instrument widely used in pediatric populations with good evidence for performance, but requiring more [13]. The EQ-5D-Y-5L was identified as priority for inclusion because it is a promising new instrument that requires further validation work. |

Table A1. Cont.

| Instrument | Description | Application in This Study | Justification for Inclusion in P-MIC |
|-------------------------------|--|---------------------------|---|
| CHU9D | The CHU9D was developed for the purpose of measuring HRQoL in children and young people [28]. It consists of 9 items covering 9 domains. It was initially developed for children aged 7–11 years; however, has been validated in children up to 17 years [29,30]. Additionally, a proxy version of the instrument is being trialed in younger children. | See Figure 1 and Table 1. | Identified as priority for inclusion as it is a generic instrument widely used in pediatric populations with good evidence for performance, but requiring more [13]. |
| Additional HRQoL Instruments | | | |
| AQoL-6D Adolescent | The AQoL-6D Adolescent was adapted from the AQoL-6D adult version [31]. It is a HRQoL instrument consisting of 20 items covering 6 domains. It was developed with adolescents aged 12–18 years, however, has been used in children aged 11 years [32,33]. | See Figure 1 and Table 1. | Although not as widely used as the core instruments in pediatric populations [13], this instrument was identified for inclusion as an additional instrument because of the opportunity to add to the evidence base regarding how this ‘adolescent’ tool might perform in younger children. |
| HUI2/3 | The HUI3 was developed to address issues in the HUI2 and allow for the use of the HUI in both clinical and general populations. It consists of 15 items with between 4 and 6 severity levels. The HUI2 classification system has 7 domains [34] and the HUI3 has 8 [35]. The HUI is officially recommended for use in children ≥ 5 years, with self-report recommended for children ≥ 8 ; however, some studies have used the HUI in children as young as 1 year old [36]. | See Figure 1 and Table 1. | This instrument was identified for inclusion because it is a generic instrument widely used in pediatric populations [13]. It was chosen as an additional instrument because the evidence for the psychometric performance of this instrument is more mixed compared with that of core instruments [13]. |
| EQ-5D-5L | The EQ-5D-5L is an adult HRQoL instrument adapted from the EQ-5D-3L [37]. It consists of 5 items with 5 severity levels covering 5 domains. The EQ-5D-5L is recommended to be used in adults and children as young as 16 years. | See Figure 1 and Table 1. | Identified for inclusion as additional instrument because of the opportunity to add to the evidence base regarding how this ‘adult’ tool might perform in younger adolescents. |
| PROMIS-25 Paediatric Profile | The PROMIS-25 Paediatric Profile is a relatively new HRQoL instrument. It consists of 25 items covering 6 domains [38]. It has child-self report and proxy-report versions available, with child self-report recommended for ≥ 8 years age and proxy report from ≥ 5 years. | See Figure 1 and Table 1. | Identified for inclusion as additional instrument because it is a promising new instrument requiring further validation work. |
| Disease Specific Instruments | | | |
| PedsQL Asthma Module (Asthma) | The PedsQL Asthma Module is a validated asthma specific pediatric HRQoL instrument [39]. It consists of 28 items (26 items for children aged 2–4 years) covering 4 domains [39]. There are self and proxy report versions for children ≥ 2 years. | See Figure 1 and Table 1. | The PedsQL Asthma Module was chosen as the disease-specific instrument for the asthma group as it is an asthma specific HRQoL instrument that has previously been validated in children. |
| KIDSCREEN-27 (ASD) | The KIDSCREEN-27 is a validated generic HRQoL instrument developed for use in children and adolescents and was adapted from the KIDSCREEN-52 [40,41]. It consists of 27 items covering 5 domains. A self-report and proxy version of the KIDSCREEN-27 is available. | See Figure 1 and Table 1. | The KIDSCREEN-27 was chosen as the disease-specific instrument for the ASD group as, although this is a generic HRQoL instrument, it has previously been recommended as a robust HRQoL instrument in children with ASD and no appropriate ASD-specific HRQoL instrument or symptom scale was available [42]. Additionally, a recent literature review found the KIDSCREEN-27 was one of two instruments that aligned with the Convention on the Rights of Persons with a Disability [43]. |

Table A1. Cont.

| Instrument | Description | Application in This Study | Justification for Inclusion in P-MIC |
|---|--|--|--|
| SDSC (Sleep problems) | The Sleep Disturbance Scale for Children (SDSC) is a validated questionnaire designed to assess sleep-related behaviors in children and adolescents [44]. It consists of 26 items covering 6 domains. Only a proxy report version of the SDSC is available and it was initially validated in children aged 6–16 years; however, another study has successfully utilized the SDSC in children as young as 3 years [45]. | As only a proxy report version of the instrument is available, this study will only collect the SDSC via proxy report. | The SDSC was chosen as the disease-specific instrument for the sleep problems group as it is a validated symptom measure in children covering a range of domains and no appropriate sleep-specific HRQoL instrument was available. |
| CPQ 11-14 (Dental decay) | The Child Perceptions Questionnaire (CPQ) 11-14 is a validated child self-reported oral HRQoL instrument, and is the most commonly used oral HRQoL instrument in children [46,47]. The CPQ 11-14 consists of 16 core items and 2 global items. It was initially designed for children aged 11–14 years; however, recent evidence indicates it may be used in children as young as 5 [48]. | As only a self-report version of the 16-item instrument is available, this study will generate a proxy report version to ensure the ability to collect this instrument via proxy report in cases where the child is not able to self-report. | The CPQ 11-14 was chosen as the disease-specific instrument for the dental decay group as it is an oral health-specific HRQoL instrument that has previously been validated in children. |
| RCADS-25 (Mental health- anxiety and/or depression) | The Revised Children’s Anxiety and Depression Scale (RCADS-25) is a symptom-based scale measuring anxiety and depression in children [49]. It consists of 25 items, 15 related to anxiety and 10 related to depression, covering 6 domains (generalized anxiety disorder, major depressive disorder, obsessive compulsive disorder, panic disorder, separation anxiety disorder and social phobia). The RCADS-25 has been validated in children aged 7–18 years and has both child self-report and proxy report versions available [49]. | See Figure 1 and Table 1. | The RCADS-25 was chosen as the disease-specific instrument for the anxiety and/or depression group as it is a validated symptom measure in children covering a range of domains and no appropriate depression- or anxiety-specific HRQoL instrument was available. |
| SWAN-adapted (ADHD) | The Strengths and Weaknesses of Attention-Deficit/Hyperactivity Disorder Symptoms and Normal Behavior Scale (SWAN) is an ADHD symptom scale that was developed as an adaptation of the Swanson, Nolan and Pelham (SNAP) rating scale [50]. The SWAN is an 18-item proxy-reported measure with 7 severity levels covering 3 symptom areas. The instrument is validated in children aged 6–18 years, but has been used in children as young as 4 [51]. | With approval from developer, instrument was adapted slightly to increase readability for lay user. As only a proxy report version of the instrument is available, this study will only collect the SWAN via proxy report. | The SWAN was chosen as the disease-specific instrument for the ADHD group as it is a validated symptom measure in children covering a range of domains and no appropriate ADHD-specific HRQoL instrument was available. |

Appendix B. Additional Screening Questions for Sample Three, Online Disease Group Panel

Table A2. Disease group screening questions for sample three.

| Disease Group | Screening Question | Inclusion Criteria |
|----------------|---|---|
| Asthma | Has your child been diagnosed with Asthma by a doctor? Yes/No | If ‘Yes’ to screening question. |
| ASD | Has your child been diagnosed with autism spectrum disorder (ASD) by a doctor? Yes/No | If ‘Yes’ to screening question. |
| Sleep problems | How much is (study child)’s sleeping pattern or habits a problem for you? Not a problem at all/A small problem/A moderate problem/A large problem | If ‘A moderate problem’ or ‘A large problem’ to screening question. |
| Dental decay | Has your child had tooth decay in the last 2 years (excluding preventative care)? Yes/No | If ‘Yes’ to screening question. |

Table A2. Cont.

| Disease Group | Screening Question | Inclusion Criteria |
|---------------------------|--|--|
| Anxiety and/or depression | Assessed via the SDQ. | As per SDQ clinical cut off for internalizing (anxiety and depression) problems. |
| ADHD | Has your child been diagnosed with Attention-Deficit/Hyperactivity Disorder (ADHD) by a doctor? Yes/No | If 'Yes' to screening question. |

Appendix C. Items Used to Assist in the Measurement of Change in Child's Health Status between Initial and Follow-Up Survey

Table A3. Description of items used to measure change in child's health between initial and follow-up survey.

| Item Description | Purpose | Sample 1, Sample Recruited via Hospital | | Sample 2, Online Panel Population Sample | | Sample 3, Online Disease Group Sample | |
|--|--|---|-----------|--|-----------|---------------------------------------|-----------|
| | | Initial | Follow-Up | Initial | Follow-Up | Initial | Follow-Up |
| Hospital department the child is currently receiving care from (reported by parent/caregiver). | Help identify groups of children expected to change at follow-up. | X | | | | | |
| Two item Children with Special Health Care Needs (CSHCN) Screener [52]. | Help identify groups of children expected to change at follow-up. | X | | X | | X | |
| Child's common ongoing health conditions (selected from list), reported by parent/caregiver. Adapted from from the Longitudinal Study of Australian Children (LSAC), Australia's first nationally representative longitudinal study of child development. | Help identify groups of children expected to change at follow-up. | X | | X | | X | |
| Single item global health measure, reported by parent/caregiver and child self-report if able. | Help identify change in health status between surveys. | X | X | X | X | X | X |
| PedsQL | Help identify minimal clinically important changes in health status between surveys. | X | X | X | X | X | X |
| Parent/caregiver report of child's change in general health between initial and follow-up survey. Adapted from similar items within a previously validated instrument, SF-36 [53]. | Help identify change in health status between surveys. | | X | | X | | X |
| Parent/caregiver report of child's change in main health condition (if reported in initial survey that child has online health condition) between initial and follow-up survey. Adapted from similar items within a previously validated instrument, SF-36 [53]. | Help identify change in health status between surveys. | | X | | X | | X |

Table A3. Cont.

| Item Description | Purpose | Sample 1, Sample Recruited via Hospital | | Sample 2, Online Panel Population Sample | | Sample 3, Online Disease Group Sample | |
|---|---|---|-----------|--|-----------|---------------------------------------|-----------|
| | | Initial | Follow-Up | Initial | Follow-Up | Initial | Follow-Up |
| Parent/caregiver report of major health event between initial and follow-up survey (new treatment/therapy, new medication, accident or injury, new diagnosis, new illness, unplanned doctor/hospital visit). If any major health event selected, parent/caregiver asked to report if this event made the child's health better worse or it had no change. | Help identify direction of change in health status between surveys. | | X | | X | | X |

X- indicates the item will be collected from the sample/time point.

Appendix D. Detailed Sample Size Calculations

Appendix D.1. Responsiveness Sample Size Calculation

Using a two-sided paired test with type I error of 1%, it was determined that a sample of 139 would give statistical power of 0.9 to detect a minimal clinically important difference (MCID) of 4.36 in the total score of the PedsQL self-report. This sample size calculated using Stata 14 using the Stata code 'power pairedmeans, altdiff(4.36) sddiff(13.16) alpha(0.01) power (0.9)'. Values inserted into the power calculation are the standard deviation (13.16) and MCID (4.36) PedsQL self-report scores obtained from Table 2 of Varni et al. 2003 [16].

Using a two-sided paired test with type I error of 1%, it was determined that a sample of 190 would give statistical power of 0.9 to detect a minimal clinically important difference (MCID) of 4.5 in the total score of the PedsQL proxy-report. This sensitivity analysis of the responsiveness sample size calculation was done using PedsQL MCID proxy-report scores. The sensitivity calculation was completed using the Stata code 'power pairedmeans, altdiff(4.5) sddiff(15.9) alpha(0.01) power(0.9)'. Values inserted into the sensitivity power calculation are the standard deviation (15.9) and MCID (4.5) PedsQL proxy-report scores obtained from Table 2 of Varni et al. 2003 [16].

Appendix D.2. Known-Group Validity Sample Size Calculation

Using a one-sided test with type I error of 1%, it was determined that a sample of 186 (93 per group) would give statistical power of 0.9 to detect a difference between healthy children and children with a chronic condition on the total score of the PedsQL proxy-report [16]. This sample size calculation was completed using the Stata code 'power twomeans 82.3 73.1, sd1 (15.6) sd2 (16.5) alpha (0.01) power (0.9)'. Values inserted into the power calculation are the mean (82.3, 73.1) and standard deviation (15.6, 16.5) of the PedsQL proxy-report scores for the healthy and chronic condition groups respectively, obtained from Table 4 of Varni et al. 2003 [16].

Using a one-sided test with type I error of 1%, it was determined that a sample of 128 (64 per group) would give statistical power of 0.9 to detect a difference between healthy children and children with a chronic condition on the total score of the PedsQL self-report [16]. This sensitivity analysis of the known-group validity sample size calculation was done using PedsQL self-report scores for healthy and chronically ill children. This sensitivity calculation was completed with the Stata code 'power two means 83.9 74.2, sd1 (12.5) sd2 (15.4) alpha (0.01) power (0.9)'. Values inserted into the power calculation are the mean (83.9, 74.2) and standard deviation (12.5, 15.4) of the PedsQL self-report scores for the healthy and chronic condition groups respectively, obtained from Table 4 of Varni et al. 2003 [16].

References

- Kromm, S.K.; Bethell, J.; Kraglund, F.; Edwards, S.A.; Laporte, A.; Coyte, P.C.; Ungar, W.J. Characteristics and quality of pediatric cost-utility analyses. *Qual. Life Res.* **2012**, *21*, 1315–1325. [\[CrossRef\]](#)
- 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*, 1618661. [\[CrossRef\]](#)
- Dawson, J.; Doll, H.; Fitzpatrick, R.; Jenkinson, C.; Carr, A.J. The routine use of patient reported outcome measures in healthcare settings. *BMJ* **2010**, *340*, c186. [\[CrossRef\]](#)
- Bulfone, L.; Younie, S.; Carter, R. Health Technology Assessment: Reflections from the Antipodes. *Value Health* **2009**, *12*, S28–S38. [\[CrossRef\]](#)
- Services Australia, Commonwealth of Australia. Annual Report 2019–2020. 2020. Available online: <https://www.servicesaustralia.gov.au/sites/default/files/annual-report-2019-20v2.pdf> (accessed on 1 June 2021).
- Services Australia, Commonwealth of Australia. PBS Expenditure and Prescriptions Report 1 July 2019 to 30 June 2020. 2020. Available online: https://www.pbs.gov.au/statistics/expenditure-prescriptions/2019-2020/PBS_Expenditure_and_Prescriptions_Report_1-July-2019_to_30-June-2020.pdf (accessed on 1 June 2021).
- Higginson, I.J.; Carr, A.J. Using quality of life measures in the clinical setting. *BMJ* **2001**, *322*, 1297. [\[CrossRef\]](#)
- Bele, S.; Chugh, A.; Mohamed, B.; Teela, L.; Haverman, L.; Santana, M.J. Patient-Reported Outcome Measures in Routine Pediatric Clinical Care: A Systematic Review. *Front. Pediatrics* **2020**, *8*, 364. [\[CrossRef\]](#)
- Janssens, A.; Thompson Coon, J.; Rogers, M.; Allen, K.; Green, C.; Jenkinson, C.; Tennant, A.; Logan, S.; Morris, C. A systematic review of generic multidimensional patient-reported outcome measures for children, part I: Descriptive characteristics. *Value Health* **2015**, *18*, 315–333. [\[CrossRef\]](#)
- Finch, A.P.; Brazier, J.E.; Mukuria, C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *Eur. J. Health Econ.* **2018**, *19*, 557–570. [\[CrossRef\]](#)
- Ungar, W. *Economic Evaluation in Child Health*; Oxford University Press: Oxford, UK, 2009.
- Chen, G.; Ratcliffe, J. A Review of the Development and Application of Generic Multi-Attribute Utility Instruments for Paediatric Populations. *Pharmacoeconomics* **2015**, *33*, 1013–1028. [\[CrossRef\]](#)
- Rowen, D.; Keetharuth, A.D.; 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 Health* **2021**, *24*, 443–460. [\[CrossRef\]](#)
- Liu, T.; Lingam, R.; Lycett, K.; Mensah, F.K.; Muller, J.; Hiscock, H.; Huque, M.H.; Wake, M. Parent-reported prevalence and persistence of 19 common child health conditions. *Arch. Dis. Child.* **2018**, *103*, 548–556. [\[CrossRef\]](#)
- Anthoine, E.; Moret, L.; Regnault, A.; Sébille, V.; Hardouin, J.-B. Sample size used to validate a scale: A review of publications on newly-developed patient reported outcomes measures. *Health Qual. Life Outcomes* **2014**, *12*, 1–10. [\[CrossRef\]](#)
- Varni, J.W.; Burwinkle, T.M.; Seid, M.; Skarr, D. The PedsQL 4.0 as a pediatric population health measure: Feasibility, reliability, and validity. *Ambul. Pediatr.* **2003**, *3*, 329–341. [\[CrossRef\]](#)
- Devlin, N.J.; Shah, K.K.; Feng, Y.; Mulhern, B.; van Hout, B. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Econ.* **2018**, *27*, 7–22. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mokkink, L.B.; Terwee, C.B.; Patrick, D.L.; Alonso, J.; Stratford, P.W.; Knol, D.L.; Bouter, L.M.; de Vet, H.C. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J. Clin. Epidemiol.* **2010**, *63*, 737–745. [\[CrossRef\]](#)
- Lamping, D.L.; Schroter, S.; Marquis, P.; Marrel, A.; Duprat-Lomon, I.; Sagnier, P.-P. The Community-Acquired Pneumonia Symptom Questionnaire: A New, Patient-Based Outcome Measure to Evaluate Symptoms in Patients with Community-Acquired Pneumonia. *Chest* **2002**, *122*, 920–929. [\[CrossRef\]](#) [\[PubMed\]](#)
- Peasgood, T.; Mukuria, C.; Carlton, J.; Connell, J.; Devlin, N.; Jones, K.; Lovett, R.; Naidoo, B.; Rand, S.; Rejon-Parrilla, J.C.; et al. What is the best approach to adopt for identifying the domains for a new measure of health, social care and carer-related quality of life to measure quality-adjusted life years? Application to the development of the EQ-HWB? *Eur. J. Health Econ.* **2021**, *22*, 1067–1081. [\[CrossRef\]](#)
- Goodman, R.; Ford, T.; Simmons, H.; Gatward, R.; Meltzer, H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br. J. Psychiatry* **2000**, *177*, 534–539. [\[CrossRef\]](#)
- Stone, L.L.; Otten, R.; Engels, R.C.; Vermulst, A.A.; Janssens, J.M. Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: A review. *Clin. Child Fam. Psychol. Rev.* **2010**, *13*, 254–274. [\[CrossRef\]](#)
- Goodman, R. The Strengths and Difficulties Questionnaire: A research note. *J. Child Psychol. Psychiatry* **1997**, *38*, 581–586. [\[CrossRef\]](#)
- Varni, J.W.; Seid, M.; Kurtin, P.S. PedsQL 4.0: Reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Med. Care* **2001**, *39*, 800–812. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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. *J. Patient-Rep. Outcomes* **2020**, *4*, 94. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wille, N.; Badia, X.; Bonsel, G.; Burström, K.; Cavrini, G.; Devlin, N.; Egmar, A.C.; Greiner, W.; Gusi, N.; Herdman, M.; et al. Development of the EQ-5D-Y: A child-friendly version of the EQ-5D. *Qual. Life Res.* **2010**, *19*, 875–886. [\[CrossRef\]](#)

27. Kreimeier, S.; Åström, M.; Burström, K.; Egmar, A.C.; Gusi, N.; Herdman, M.; Kind, P.; Perez-Sousa, M.A.; Greiner, W. EQ-5D-Y-5L: Developing a revised EQ-5D-Y with increased response categories. *Qual. Life Res.* **2019**, *28*, 1951–1961. [[CrossRef](#)] [[PubMed](#)]
28. Stevens, K. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Qual. Life Res.* **2009**, *18*, 1105–1113. [[CrossRef](#)]
29. Ratcliffe, J.; Stevens, K.; Flynn, T.; Brazier, J.; Sawyer, M. An assessment of the construct validity of the CHU9D in the Australian adolescent general population. *Qual. Life Res.* **2012**, *21*, 717–725. [[CrossRef](#)]
30. Stevens, K.; Ratcliffe, J. Measuring and valuing health benefits for economic evaluation in adolescence: An assessment of the practicality and validity of the child health utility 9D in the Australian adolescent population. *Value Health* **2012**, *15*, 1092–1099. [[CrossRef](#)]
31. 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 Health* **2010**, *13*, 1014–1027. [[CrossRef](#)]
32. Ratcliffe, J.; Stevens, K.; Flynn, T.; Brazier, J.; Sawyer, M.G. Whose Values in Health? An Empirical Comparison of the Application of Adolescent and Adult Values for the CHU-9D and AQOL-6D in the Australian Adolescent General Population. *Value Health* **2012**, *15*, 730–736. [[CrossRef](#)] [[PubMed](#)]
33. Keating, C.L.; Moodie, M.L.; Richardson, J.; Swinburn, B.A. Utility-Based Quality of Life of Overweight and Obese Adolescents. *Value Health* **2011**, *14*, 752–758. [[CrossRef](#)]
34. 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. *Med. Care* **1996**, *34*, 702–722. [[CrossRef](#)]
35. Feeny, D.; Furlong, W.; Torrance, G.W.; Goldsmith, C.H.; Zhu, Z.; Depauw, S.; Denton, M.; Boyle, M. Multiattribute and Single-Attribute Utility Functions for the Health Utilities Index Mark 3 System. *Med. Care* **2002**, *40*, 113–128. [[CrossRef](#)] [[PubMed](#)]
36. Gemke, R.J.; Bonsel, G.J. Reliability and validity of a comprehensive health status measure in a heterogeneous population of children admitted to intensive care. *J. Clin. Epidemiol.* **1996**, *49*, 327–333. [[CrossRef](#)]
37. Herdman, M.; Gudex, C.; Lloyd, A.; Janssen, M.F.; Kind, P.; Parkin, D.; Bonsel, G.; Badia, X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual. Life Res.* **2011**, *20*, 1727–1736. [[CrossRef](#)]
38. Northwestern University. Intro to PROMIS®. 2021. Available online: <https://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis> (accessed on 25 July 2021).
39. Varni, J.W.; Burwinkle, T.M.; Rapoff, M.A.; Kamps, J.L.; Olson, N. The PedsQL in pediatric asthma: Reliability and validity of the Pediatric Quality of Life Inventory generic core scales and asthma module. *J. Behav. Med.* **2004**, *27*, 297–318. [[CrossRef](#)]
40. Ravens-Sieberer, U.; Herdman, M.; Devine, J.; Otto, C.; Bullinger, M.; Rose, M.; Klasen, F. The European KIDSCREEN approach to measure quality of life and well-being in children: Development, current application, and future advances. *Qual. Life Res.* **2014**, *23*, 791–803. [[CrossRef](#)]
41. Ravens-Sieberer, U.; Auquier, P.; Erhart, M.; Gosch, A.; Rajmil, L.; Bruil, J.; Power, M.; Duer, W.; Cloetta, B.; Czemy, L.; et al. The KIDSCREEN-27 quality of life measure for children and adolescents: Psychometric results from a cross-cultural survey in 13 European countries. *Qual. Life Res.* **2007**, *16*, 1347–1356. [[CrossRef](#)]
42. Tavernor, L.; Barron, E.; Rodgers, J.; McConachie, H. Finding out what matters: Validity of quality of life measurement in young people with ASD. *Child Care Health Dev.* **2013**, *39*, 592–601. [[CrossRef](#)] [[PubMed](#)]
43. Davis, E.; Young, D.; Gilson, K.-M.; Swift, E.; Chan, J.; Gibbs, L.; Tonmukayakul, U.; Reddihough, D.; Williams, K. A Rights-Based Approach for Service Providers to Measure the Quality of Life of Children with a Disability. *Value Health* **2018**, *21*, 1419–1427. [[CrossRef](#)] [[PubMed](#)]
44. Bruni, O.; Ottaviano, S.; Guidetti, V.; Romoli, M.; Innocenzi, M.; Cortesi, F.; Giannotti, F. The Sleep Disturbance Scale for Children (SDSC). Construction and validation of an instrument to evaluate sleep disturbances in childhood and adolescence. *J. Sleep Res.* **1996**, *5*, 251–261. [[CrossRef](#)]
45. Romeo, D.M.; Bruni, O.; Brogna, C.; Ferri, R.; Galluccio, C.; De Clemente, V.; Di Jorio, M.; Quintiliani, M.; Ricci, D.; Mercuri, E. Application of the Sleep Disturbance Scale for Children (SDSC) in preschool age. *Eur. J. Paediatr. Neurol.* **2013**, *17*, 374–382. [[CrossRef](#)]
46. Thomson, W.M.; Foster Page, L.A.; Robinson, P.G.; Do, L.G.; Traebert, J.; Mohamed, A.R.; Turton, B.J.; McGrath, C.; Bekes, K.; Hirsch, C.; et al. Psychometric assessment of the short-form Child Perceptions Questionnaire: An international collaborative study. *Community Dent. Oral Epidemiol.* **2016**, *44*, 549–556. [[CrossRef](#)] [[PubMed](#)]
47. Gilchrist, F.; Rodd, H.; Deery, C.; Marshman, Z. Assessment of the quality of measures of child oral health-related quality of life. *BMC Oral Health* **2014**, *14*, 40. [[CrossRef](#)]
48. Foster Page, L.A.; Boyd, D.; Thomson, W.M. Do we need more than one Child Perceptions Questionnaire for children and adolescents? *BMC Oral Health* **2013**, *13*, 26. [[CrossRef](#)] [[PubMed](#)]
49. Ebesutani, C.; Korathu-Larson, P.; Nakamura, B.J.; Higa-McMillan, C.; Chorpita, B. The Revised Child Anxiety and Depression Scale 25-Parent Version: Scale Development and Validation in a School-Based and Clinical Sample. *Assessment* **2017**, *24*, 712–728. [[CrossRef](#)] [[PubMed](#)]
50. Swanson, J.M.; Schuck, S.; Porter, M.M.; Carlson, C.; Hartman, C.A.; Sergeant, J.A.; Clevenger, W.; Wasdell, M.; McCleary, R.; Lakes, K.; et al. Categorical and Dimensional Definitions and Evaluations of Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales. *Int. J. Educ. Psychol. Assess* **2012**, *10*, 51–70.

-
51. Serrien, D.J.; Sovijärvi-Spapé, M.M.; Rana, G. Developmental changes in motor control: Insights from bimanual coordination. *Dev. Psychol.* **2014**, *50*, 316–323. [[CrossRef](#)]
 52. Bethell, C.D.; Read, D.; Stein, R.E.K.; Blumberg, S.J.; Wells, N.; Newacheck, P.W. Identifying Children with Special Health Care Needs: Development and Evaluation of a Short Screening Instrument. *Ambul. Pediatrics* **2002**, *2*, 38–48. [[CrossRef](#)]
 53. Stansfeld, S.A.; Roberts, R.; Foot, S.P. Assessing the validity of the SF-36 General Health Survey. *Qual. Life Res.* **1997**, *6*, 217–224. [[CrossRef](#)]