Elsevier Editorial System(tm) for Social Science & Medicine
Manuscript Draft

Manuscript Number: SSM-D-19-01509R1

Title: Mind the (inter-rater) gap. An investigation of self-reported versus proxy-reported assessments in the derivation of childhood utility values for economic evaluation: A systematic review

Article Type: Review article

Keywords: Childhood, Utility, Quality of Life, Self-reports, Proxy-reports, multi-attribute instrument, direct valuation methods, Health Economics

Corresponding Author: Dr. Jyoti Khadka, PhD

Corresponding Author's Institution: Flinders University

First Author: Jyoti Khadka, PhD

Order of Authors: Jyoti Khadka, PhD; Joseph Kwon, MSc; Stavros Petrou, PhD; Emily Lancsar, PhD; Julie Ratcliffe, PhD

Manuscript Region of Origin: AUSTRALIA

Abstract: Purpose: Evidence surrounding utilities for health states, derived either directly from the application of preference-based valuation methods or indirectly from the application of preference-based quality of life instruments, is increasingly being utilised to inform the cost-effectiveness of child health interventions. Proxy (parent or health professional) assessments are common in this area. This study sought to investigate the degree of convergence in childhood utilities generated directly or indirectly within dyad child and proxy assessments.
Methods: A systematic literature review was conducted following PRISMA guidelines. A comprehensive literature search strategy conducted across six search engines (PubMed, Embase, Web of Science, PsychoINFO, EconLit, CINAHL and Cochrane Library). Original peer-reviewed articles that reported utilities derived directly or indirectly using simultaneous dyad child and proxy assessments were extracted. Mean and median utilities, correlation coefficients and levels of agreement were extracted, catalogued and assessed.
Results: A total of 35 studies that reported utilities for two or more respondent types were identified. Of these, 29 studies reported dyad childhood self-report and proxy utilities whilst six studies reported levels of agreement and/or correlations only without documenting overall utilities. Proxy assessment was most often conducted by parents with the HUI3 representing the most commonly applied instrument across a range of health conditions. The utilities derived from child and parent proxy assessment were bidirectional with parental proxies tending to underestimate and health professional proxies tending to overestimate relative to child self-reports. Inter-rater agreement between child self-reports and parent-proxy reports were poorer for more subjective attributes (cognition, emotion and pain), relative to physical attributes (mobility, self-care, speech, vision) of health-related quality of life.

Conclusions: Childhood utilities derived from children or proxies are not interchangeable. The choice of self or proxy assessor may have potentially significant implications for economic evaluations of child health interventions.

Ref: SSM-D-19-01509

Dear Dr. Khadka,

Thank you for submitting your manuscript to Social Science & Medicine. We have now completed its review and attach the comments of the reviewers. Your paper is of potential interest, however based on the reviewers' comments, we cannot accept your manuscript in its present form. As you will see our reviewers raised a number of suggestions for revision and clarification.

Response: We would like to sincerely thank the reviewers for their insightful comments and the opportunity to revise the manuscript. We have addressed all the issues raised point-by-point and have also made corresponding changes in the manuscript (please refer to marked manuscript with track changes). We hope that you will find them satisfactory.

In this response letter,

Our responses to the reviewers' comments are in blue

New text added in the manuscript is in green

Reviewer #1:

Comment 1: The introduction was well-researched, however the writing style was highly academic. Much more could have been written in plain writing. Aside from the utility to the reader, this writing approach makes it difficult to see the novelty and necessity for this research. It was not make clear to me what gap this review is filling in the literature after a few readings of the introductory section. If you were casually explaining the topic to an educated audience, what language would you use? This approach needs to be utilized more in the introduction to pull in the reader.

Ans: We would like to thank the reviewer for their comments. Our intention was to frame the introduction in a manner that provides an overview of health economic evaluation methods and potential methodological challenges in childhood utility assessments for those readers who are health economics novices. However, we agree that we may not have provided sufficient rationale for undertaking the study. We have revised and recognised the material in the introduction in several places and additionally added the following text to the introduction (refer to marked manuscript: introduction, p3, last paragraph)

Previous studies of childhood health utilities primarily focused on specific diseases and were limited to investigation of the independent effects of respondent type (Known et al., 2018, Pickard et al., 2004, van Litsenburg et al., 2014). Furthermore, those studies did not assess the self-proxy convergence (or divergence) for overall utilities and individual health domains within MAUIs or how utilities may differ between children and proxies across disease areas. This study therefore aimed to systematically assess the degree of convergence (or divergence) in childhood utilities generated directly or indirectly in studies that included simultaneous dyad self and proxy (parent and/or health professional) assessment of childhood HRQoL outcomes.

Comment 2: The methodology seems appropriately described and the review appears to be comprehensive and well done.

Ans: Thank you.

Comment 3: The results section was comprehensive and I appreciated the subheadings, however I thought further stratification of the studies may have been possible based on the information you describe, perhaps by type of proxy, MAUI, or health condition.

Ans: Thank you for the comments, we have now stratified the results by types of proxies and valuation methods. However, due to the paucity of eligible studies, further stratification by health conditions was not deemed appropriate for drawing meaningful comparisons.

Comment 4: The discussion is great- "The findings reinforce the importance of obtaining self-reports where possible from childhood populations (rather than relying on proxy responses) when considering childhood HRQoL outcomes." This is the first time in the paper I felt like I understood the "why" of the research. Subheadings are needed in the discussion to delineate the different arguments you are making.

Ans: We have added subheadings in the discussion and re-organised it accordingly.

Comment 5: This is a category of paper I find most difficult to review. Technically, it is excellent. There are few obvious errors to correct. The review is comprehensive, and the results substantive. Yet, it was well after the first 2/3 of the paper that I started to understand the need for this study and the implications of this research. Especially when talking about populations such as children, there needs to be a deeper human element to introducing the paper and making the case for the study. Yes, this is an academic paper, but in this type of journal we cannot forget the "social science" part of the equation. I think if the introduction especially is revised with this in mind, it will strengthen the overall paper and draw the reader into your work.

Ans: Please refer to our response to comment 1.

Reviewer #2: This was a very clearly written paper on a systematic review of studies measuring and reporting health utility from both a child and proxy perspective. It is focused on understanding the methods of measurement and the relationship between utility values derived. It represents an enormous amount of work and a very large amount of literature reviewed. In the child health economic evaluation research space this is a very valuable contribution and feeds into one of the most critical areas around accurate and valid measurement of outcomes for children.

Thank you for your insightful comments and suggestions.

I have a couple of comments that might help strengthen the manuscript.

Comment 6: Title: Consider the use of the word 'empirical' in the title. Might it be more helpful to the reader to highlight the systematic review methodology?

Ans: We have revised the title as follows:

“Mind the (inter-rater) gap. An investigation of self-reported versus proxy-reported assessments in the derivation of childhood utility values for economic evaluations: A systematic review”

Comment 7: Introduction paragraph 2. The wording of the first sentence does not seem to match up with the Chen and Ratcliffe 2015 reference. The Chen article focusses on derivation of child health utilities largely using adult preferences whereas the sentence currently talks about most widely used adult MAU's. Perhaps consider an alternative reference or reword the sentence to focus on children.

Response: Apologies for the oversight, a relevant reference has now been added.

Bulamu, N.B., Kaambwa, B., & Ratcliffe, J. (2015). A systematic review of instruments for measuring outcomes in economic evaluation within aged care. Health Qual Life Outcomes, 13, 179.

Comment 8: Methods- Literature Search, first paragraph. Consider justifying why the existing database PEDEs was not used as part of the search strategy. http://pede.ccb.sickkids.ca/pede/index.jsp The PEDE's database states that it includes an inventory of child health state utility weights. Consider cross checking to ensure the existing search strategy has not missed any articles included in PEDES.

Response: We did search the PEDEs database besides the other databases mentioned in the manuscript. In order to indicate that we searched the PEDEs database, we have added the following text in the methods section. (refer to marked manuscript, p4, first paragraph, last two lines)

The Paediatric Economic Database Evaluation (PEDE) was also searched to identify CUAs published between 1990 and 2016 that incorporated primary estimation of health utilities.

Comment 9: Methods- data analysis, p 5 last sentence. It would be helpful to provide more information about the choice to categorise the studies into "cancer-related", "other" and "general population". The whole review is subsequently organized around these three themes and it would be helpful for the reader to better understand the choice and rationale for this.

Response: This categorisation was based upon the focus of the identified studies and for the purpose of synthesising the data to extract the main findings. We have added the following text in the paper to highlight this. (refer to marked manuscript, p6, 2nd paragraph)

In order to achieve a meaningful appraisal and comparison between studies, the included studies were categorized into three main groups based upon the prevalence of the identified health condition(s) and perspective(s), namely cancer-related, other health conditions and general population-related. We used the International Classification of Diseases 10 revision (ICD-10) to aggregate samples into cancer-related and other health condition(s). Any general population sample that acted as a control or comparative group within a study was considered as a unique sample and classified under the 'general population' category.

Comment 10: The review quality appraises each article using a quality of reporting score that was developed by the authors around the 15 extracted variables. The method for assessing quality of reporting seems sensible and the 15 variables are key to the review. However, in Results, p 6 second to last sentence of paragraph 1 and on a couple of subsequent occasions the language slips to 'high quality based on our quality appraisal'. Care should be taken to ensure that when reporting that this is specifically referred to as quality of 'reporting'.

Response: We agree and we have now changed this sentence to the following. (refer to marked manuscript, p7, first paragraph)

"Included articles had a quality score that ranged from 11 to 13, signifying moderate to high quality of reporting based on our quality appraisal."

Comment 11: Study Characteristics, Page 7. Confirm if it is generally accepted clinically to refer to LBW as a congenital disorder.

Response: We thank the reviewer for this observation. We have now deleted the words 'low birth weight' to the examples provided in parentheses.

Comment 12: Study Characteristics, second paragraph page 7. The first sentence refers to 4 studies but then only 3 references are provided.

Response: The study that was not cited is Kulpeng et al (2013). This was an oversight and has been corrected now, thank you. (refer to marked manuscript, p8, 2nd paragraph)

Comment 13: Discussion page 16, last sentence of 2nd paragraph. This sentence states 'difficulty to pinpoint specific reasons for these findings" when talking about the difference in concordance between physical attributes and psychosocial attributes. Then again on page 19 second paragraph this point is taken up again but this time relating to the literature by Jardine et al 2014 providing reasons for the distinction. Might it be worth bringing these two parts of the discussion together?

Response: We have reconciled the two paragraphs into a single paragraph as suggested. Please refer to the marked manuscript (p18, last paragraph and p19, first paragraph)

Despite the implications for childhood healthcare decision making mentioned above, the results from this systematic review indicated that there was no consistent pattern in direction or magnitude of difference in the inter-rater gap (with evidence of both proxy under and over estimation relative to children) in studies of children and adolescents with clinical conditions. However, proxy overestimation of utilities relative to those generated by children and adolescents was clearly evident in healthy general childhood population samples. At the attribute or dimension level, the findings from this study concord with the wider child health and HRQoL literature in that proxy reports generally have fair to good correlations and agreement with those from childhood self-reports for overall utilities and for physical attributes, but there is often poor agreement for psychosocial attributes (Jardine et al., 2014). We also found that there was higher agreement between child and proxy-reports for physical attributes (e.g., vision, ambulation, dexterity, pain) in comparison with more subjective psychosocial attributes (e.g. emotional, cognitive) of HRQoL (Jardine et al., 2014; Petrou, 2003; Tarride et al., 2010). Due to significant heterogeneity among the included studies, it is difficult to pinpoint specific reasons for these findings. However, Jardine and colleagues argue that parents are likely to be more perceptive to large effects on their children's functioning that they can observe day-in and day-out and might not have direct insights into the emotional and psychological well-being of their children (Jardine et al., 2014). Another important issue to highlight is the source of values for the utility algorithms of contributing MAUIs. With the exception of a study in children with asthma that used the PAHOM (a disease-specific MAUI) (Gerald et al., 2012), all other studies used tariffs derived from adults. This demonstrates that there is a widespread and continued use of adult algorithms to estimate childhood utility values despite clear evidence to suggest that young people value identical health states differently to adults (Brazier et al., 2017; Ratcliffe et al., 2012). As mentioned above, these utility divergences may be sufficient to influence reimbursement decisions and healthcare policy (Ratcliffe et al., 2012). Further studies to derive childhood-specific MAUIs and scoring algorithms are required to address this critical issue.

Comment 14: Minor: Page 18, end of first paragraph there is a dash that looks inadvertent

Response: Corrected.

Comment 15: Concluding paragraph. The conclusion states that the two different viewpoints should be treated as complementary. Might it be worth saying that the child perspective is preferred if one can ensure it is validly and reliably captured? Then in cases where that is not possible proxy measurement is needed and to understand how the responses will then be different. The final sentence reads 'it is imperative that researchers reflect alterative perspectives in their utility assessments…" Do the authors really want to recommend the measurement of both perspectives always, or would they rather recommend a place for each and for acknowledging and understanding differences in the results that this will generate?

Response: Thank you for raising this important point. Our intention is not to recommend that proxy-reports should always be collected in parallel to child self-reports. However, we also want to highlight that in those circumstances when child self-reports are not possible or not reliable, proxy-reports do have a value. In response to the reviewer's suggestions, we have revised the conclusions as follows. (refer to marked manuscript, p23, conclusions)

"Utility assessments obtained from children and adolescents, and their corresponding proxies, should be viewed as complementary, rather than necessarily interchangeable. When possible, self-reported utility assessments are preferred over those from proxies if one can ensure that they are valid and reliably captured. In cases where that is not possible, both sets of assessments should be considered in clinical practice and policy settings, including those informing cost-effectiveness based decision-making, and those reports should be interpreted separately. A key finding from this review is that differences exist between utilities elicited via child self-report and their proxies. However, the direction of such differences was not homogeneous. We observed no clear pattern in the inter-rater gap both in terms of the magnitude and/or direction of differences in utilities across valuation methods. The paucity of identified studies and the heterogeneity of methods used further complicated the synthesis of findings. In conclusion, we acknowledge that there is a place for both self and proxy-reported assessments and the inclusion of both set of assessments may in turn provide greater insight into predictors of utility values and relationships between values generated by different respondent types."

Comment 16: Figure 1. Is it possible to place a N on the individual reasons for full text exclusion? Noting that often there are multiple reasons for excluding a single article, but one prominent reason is often selected for each article.

Response: Unfortunately, we did not keep the record of numbers for the individual reasons for study exclusion at the full-text stage. This is because the PRISMA guideline (point 17) recommends giving reasons for exclusions at each stage but does not additionally require disaggregating numbers of exclusions by reason. We acknowledge that this as a limitation of our review process, though it should not affect the results and conclusion of the research. We have added the following in the limitation section. (refer to marked manuscript, strengths and limitations)

Another limitation of our review process was that we have not provided numbers for the individual reasons for study exclusion at the full-text stage (Figure 1). However, this should not affect the results and conclusions of our study. The PRISMA guideline (point 17) also recommends giving reasons for exclusion at each stage, but does not ask for disaggregation by reason.

Comment 17: Table 1. With the formatting of this table it was difficult to follow the subheadings in both description 1 and 2 columns.

Response: We have reformatted Table 1 as recommended by the reviewer so that it is easier to follow.

Comment 18: Table 2. Table 2 summarises the agreement methods used in the column 'results reported'. Could this also be summarized in the text? i.e. how many studies used each method to reflect agreement between proxy and self-report (ICC, Kappa, mean difference etc)

Response: We have now added the following in the text. (refer to marked manuscript, p8, 2nd paragraph)

Of those studies that reported levels of agreements and co-relations with (16 studies) and without (6 studies) utilities, 6 studies used Kappa, 9 studies used intra-class correlations (ICC) and 12 studies used correlations coefficients (Table 2 and 3).

Comment 19: Once again, this was a really well written paper on a critical topic for child health economic evaluation. Reconciling differences in child versus proxy measurement is a core problem for advancing methods in this sub-discipline. This paper represents an important advance in understanding existing utility measurement literature.

Thank you.

*Cover Page

Mind the (inter-rater) gap. An ~~empirical~~ investigation of self-reported versus proxy-reported assessments in the derivation of childhood utility values for economic evaluation: A systematic review

Jyoti Khadka[1,2,3], Joseph Kwon[4], Stavros Petrou[4], Emily Lancsar[5], Julie Ratcliffe[1]

[1]Health and Social Care Economic Group, College of Nursing and Health Science, Flinders University, Bedford Park, South Australia, Australia

[2]Institute for Choice, Business School, University of South Australia, South Australia, Australia

[3]Registry of Older South Australians, Healthy Ageing Research Consortium, South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia

[4]Warwick Medical School, University of Warwick, Coventry, UK

[5]Department of Health Services Research and Policy, School of Population Health, The Australian National University, Canberra, Australia

Corresponding author: Dr Jyoti Khadka, Health and Social Care Economics Group, College of Nursing and Health Sciences, Flinders University, Sturt Road, Bedford Park South Australia 5042, GPO Box 2100 Adelaide SA 5001, P: +61 8 8201 3702, E: Jyoti.khadka@flinders.edu.au ; jk.goldeneye@gmail.com

Declaration of interest: None

Abstract

Purpose: Evidence surrounding utilities for health states, derived either directly from the application of preference-based valuation methods or indirectly from the application of preference-based quality of life instruments, is increasingly being utilised to inform the cost-effectiveness of child health interventions. Proxy (parent or health professional) assessments are common in this area. This study sought to investigate the degree of convergence in childhood utilities generated directly or indirectly within dyad child and proxy assessments.

Methods: A systematic literature review was conducted following PRISMA guidelines. A comprehensive literature search strategy conducted across six search engines (PubMed, Embase, Web of Science, PsychoINFO, EconLit, CINAHL and Cochrane Library). Original peer-reviewed articles that reported utilities derived directly or indirectly using simultaneous dyad child and proxy assessments were extracted. Mean and median utilities, correlation coefficients and levels of agreement were extracted, catalogued and assessed.

Results: A total of 35 studies that reported utilities for two or more respondent types were identified. Of these, 29 studies reported dyad childhood self-report and proxy utilities whilst six studies reported levels of agreement and/or correlations only without documenting overall utilities. Proxy assessment was most often conducted by parents with the HUI3 representing the most commonly applied instrument across a range of health conditions. The utilities derived from child and parent proxy assessment were bidirectional with parental proxies tending to underestimate and health professional proxies tending to overestimate relative to child self-reports. Inter-rater agreement between child self-reports and parent-proxy reports were poorer for more subjective attributes (cognition, emotion and pain), relative to physical attributes (mobility, self-care, speech, vision) of health-related quality of life.

Conclusions: Childhood utilities derived from children or proxies are not interchangeable. The choice of self or proxy assessor may have potentially significant implications for economic evaluations of child health interventions.

Key words: Childhood, Quality of Life, Self-reports, Proxy-reports, multi-attribute instrument, direct valuation methods, Health Economics

Highlights

- Differences exist between utilities elicited by childhood self-reports and their corresponding proxies with no clear direction of such differences.
- Utility assessments obtained from children and their proxies, should be viewed as complementary, rather than necessarily interchangeable.
- Studies should provide greater insight into predictors of utility values and relationships between values generated by different respondent types to reflect alternative perspectives.

Introduction:

Evidence from economic evaluations is increasingly used internationally to inform decisions about how to allocate scarce resources across health and social care systems (Karnon et al., 2018; Stevens & Ratcliffe, 2012). Health economists use a number of approaches for conducting such economic evaluations with cost-utility analysis (CUA) being a widely used approach. ~~CUA is recommended for health technology assessment by pricing and reimbursement authorities in several countries, such as the Pharmaceutical Benefit Advisory Committee (PBAC, Australia), Canadian Agency for Drugs and Technology in Health (CADTH, Canada), National Institute of Health and Care Excellence (NICE, England) and Scottish Medicines Consortium (SMC, Scotland).~~ ~~(Canadian Agency for Drugs and Technologies in Health, 2017; National Institute of Health and Care Excellence, 2013; Pharmaceutical Benefits Advisory Committee, 2016; Scottish Medicines Consortium, 2016). The measurement and valuation of health-related quality of life (HRQoL) represents the ~~Within ~~cornerstone of ~~CUA, ~~whereby ~~outcomes are typically captured and reported in terms of quality-adjusted life years (QALYs). The QALY combines health-related quality of life and length of life into a single metric by applying ~~where ~~preference-based weights (or health utilities) that are typically represented on a cardinal scale anchored at "0" (representing being dead) and "1" (representing full health) ~~are used~~ to value health-related quality of life outcomes experienced at different time intervals.

A widely used approach for calculating utilities within the context of CUA is the use of multi-attribute utility instruments (MAUIs), the most well-known and widely applied of which in adult populations is the EuroQol (EQ-5D) (Bulamu et al., 2015)~~-~~. All MAUIs rely on an indirect method for estimating health utilities, comprising two distinct elements. Firstly, a descriptive system for the measurement of HRQoL using self or proxy completion. Secondly, a preference-based scoring algorithm (or tariff) for generating utilities for each of the health states defined by the descriptive system. The scoring algorithm is typically derived using economic modelling of general population values for a sub-set of health states defined by the descriptive system. In recent years, in the context of economic evaluation of child health programmes, several MAUIs have gained prominence specifically for application in children and adolescents. The Health Utilities Index Mark II and III (HUI2 and HUI3) are the most well established instruments in this regard, and other instruments that have been applied in childhood populations include the Child Health Utility 9D (CHU9D), Quality of Well-Being Scale (QWB), EQ-5D Youth version (EQ-5D-Y), Adolescent Health Utility Measure (AHUM), Assessment of Quality of Life 6D (AQoL-6D), and Seventeen-dimensional measure of Health-related Quality of Life (HRQoL-17D). These instruments may be categorised into childhood-specific (i.e. specifically developed and designed for application in children, e.g. CHU-9D, HUI2&3, AHUM) or non-childhood-specific (i.e. instruments originally designed for application with adults but later modified for use in children, e.g. QWB, EQ-5D-Y, AQoL-6D, HRQoL-17D) (Chen & Ratcliffe, 2015). ~~All MAUIs rely on an indirect method for estimating health utilities, comprising two distinct elements. Firstly, a descriptive system for the measurement of HRQoL using self or proxy completion. Secondly, a preference-based scoring algorithm (or tariff) for generating utilities forutilities for each of the health states defined by the descriptive system. The scoring algorithm is typically derived using economic modelling of general population values for a sub-set of health states defined by the descriptive system.~~

The other approaches to eliciting utility values are direct valuation methods in which respondents are asked to value their own and/or hypothetical health state/s, or in the case of proxy assessment are asked to value the health state under consideration from the perspective of a target individual using scaling or choice-based techniques, such as the visual analogue scale (VAS), standard gamble (SG) or time trade-off (TTO) (Arnold et al., 2009). The appropriateness of applying direction valuation methods in children has been questioned by several researchers who note that children may not have the cognitive abilities to evaluate their own or hypothetical health states on the basis of risk of death (SG) or period of life (TTO) (Ratcliffe et al., 2015; Thorrington & Eames, 2015). More generally, young children in particular are likely to lack the required linguistic skills to articulate their preferences for health states using complex concepts of gain and loss in health economic terms.

In applying preference-based approaches to the measurement of HRQoL, self-report ~~even from children~~ is considered preferable wher~~e~~ever possible, including in child populations (Brazier et al., 2017). However, the psychometric literature suggests that during the early stages of childhood in particular, cognitive and communication competencies are likely to constrain the potential for self-report (Eiser & Morse, 2001; Matza et al., 2013; Stein & Jessop, 1990). Researchers working in this context therefore face the challenge of identifying appropriate age and cognition levels for eliciting self-reported descriptions of HRQoL from children. As a consequence, proxy (parent and/or health professional) reports are often used in lieu of child self-reports despite the fact that self-assessment is generally considered preferable for health economic assessments (Otero et al., 2013). Whilst it has been reported in the HRQoL literature that proxy reports (in particular from parents) may under-estimate more objective aspects of children's' functioning relative to child self-reports (Vetter et al., 2012; Youssef et al., 2006), empirical investigations have largely overlooked the implications for economic evaluations and health technology assessments reliant on utilities (Otero et al., 2013). Beyond the descriptive task, alternative sources of values for descriptions of childhood HRQoL outcomes, whether applied within direct or indirect measurement approaches, may also generate discrepancies in utilities (Petrou, 2003). Consequently, discrepancies between different type of respondents in the description and/or preference-based valuation of childhood HRQoL outcomes have potential implications for the findings of economic evaluations of child health programmes.

Previous studies of childhood health utilities primarily focused on specific diseases and were limited to investigation of the independent effects of respondent type (Kwon et al., 2018; Pickard et al., 2004; van Litsenburg et al., 2014). Furthermore, those studies did not assess the self-proxy convergence (or divergence) for overall utilities and individual health domains within MAUIs or how utilities may differ ~~across disease areas~~ between children and proxies across disease areas. This study therefore aimed to systematically assess the degree of convergence (or divergence) in childhood utilities generated directly or indirectly in studies that included simultaneous dyad self and proxy (parent and/or health professional) assessment of childhood HRQoL outcomes.

Methods

Literature search

The primary objective of this systematic review was to conduct a comprehensive systematic review of primary utility data for childhood conditions with ~~the research question of the~~a specific focus on child-proxy dyad comparisons. The time horizon for this review ~~was~~ covered articles published online before 30th June 2017. The ~~same comprehensive~~ literature search strategy and search terms ~~were~~ mirrored those for a previous~~applied~~ systematic review and meta-analysis of primary utility data for childhood conditions. (Kwon et al., 2018)~~.~~ The search terms (Appendix 1) were applied in six different online databases (PubMed, Embase, Web of Science, PsychoINFO, EconLit, CINAHL and Cochrane Library). The Paediatric Economic Database Evaluation (PEDE) was also searched to identify CUAs published between 1990 and 2016 that incorporated primary estimation of health utilities.

We searched for original peer reviewed articles that reported dyad childhood self- and proxy (adult parent and/or health professional) measured utilities, which were elicited at the same time point using direct or indirect methods. For the sake of completeness, articles reporting VAS scores were also included although it is acknowledged that some disagreement exists in the literature about the validity of VAS scores for reflecting health state preferences (Brazier et al., 2017; Parkin & Devlin, 2006). Primary studies reporting mean or median utilities, correlations and levels of agreement for child-proxy dyads with an identifiable health condition or descriptor were included. We excluded review articles and articles published in languages other than English. We also excluded articles reporting: (i) single- or multi-attribute utilities generated by direct or indirect valuation methods from a single respondent type, e.g. either children or parents alone; (ii) frequency distributions for single-attribute levels but not single- or multi-attribute utilities generated by indirect valuation methods; and (iii) utilities for combined samples of children and proxies (e.g., combined but not disentangled child-parent assessments).(Billson & Walker, 1994; Felder-Puig et al., 2000) Similarly, studies reporting utilities for mixed populations (i.e. combined values for general population and clinical groups combined), where utilities could not be attributed to a single identifiable group, were also excluded.

Data extraction

The initial screening of the titles and abstracts identified by the literature searches was carried out by two independent reviewers (J Khadka and J Kwon) with disagreements referred to a third reviewer (SP) for the final assessment for inclusion. Two reviewers (J Khadka and J Kwon) subsequently reviewed all articles that met the criteria for inclusion in the full article review (Figure 1). At this stage, a fourth reviewer (JR) independently reviewed 25% of articles to confirm decisions to include or exclude based upon the specified criteria. All the authors discussed in detail and agreed on inclusion of the final eligible articles. From each article that met our inclusion criteria, we extracted the following information using a proforma detailing bibliographic information, study design, study setting, valuation method, scoring algorithm/s applied to MAUIs (including source population and valuation method), respondent types, administration mode, sample target age(s), whether the health state/s under consideration were experienced and/or hypothetical, sample size(s), geographic setting, utility or VAS scores (including central statistics and measures of variability), response quality, statistical method(s) for analysing utilities and any methodological concerns reported by the study authors. A more detailed description of the extracted 15 variables is presented in Appendix 2. An overall reporting quality score was calculated by assigning a point to the reporting of each of these characteristics (15 is the maximum score indicating the highest quality), with the total aggregated points interpreted as the overall reporting quality score (Kwon et al., 2018). All data were entered into an Excel database.

Data analysis

We utilised narrative synthesis to describe the findings of the review. Narrative synthesis is a widely applied method in systematic reviews to synthesise findings from multiple studies where meta-analysis is not possible due to a high level of heterogeneity in the contributing evidence (Campbell et al., 2018). Narrative synthesis is particularly useful when analysts rely on the use of words and text to summarize and explain the findings of the synthesis into a coherent textual narrative. This approach involves critical appraisal of the characteristics of the study sample and defining statistical data and study findings using graphs, structural tabulations and using textual narratives to "tell the story" of the findings. This systematic review follows the guidance for narrative synthesis initially produced by the UK's Economic and Social Research Council (Campbell et al., 2018; Popay et al., 2006).

In order to achieve a meaningful appraisal and comparison, the included studies were categorized into three main groups based upon the prevalence of the identified health condition(s) and perspective(s), namely cancer-related, other health conditions and general population-related. We used the International Classification of Diseases 10 revision (ICD-10) to aggregate samples into cancer-related and other health condition. Any general population sample that acted as a control or comparative group within a study was considered as a unique sample and classified under the 'general population' category.

Results

The literature search resulted in 63,733 individual articles (Figure 1). A total of 29,025 articles were excluded following deduplication. A further 34,041 articles were excluded at the title and abstract review stage. The main reasons for exclusion at this stage were focus on adult populations only, articles that did not report primary evidence, inaccessible articles and articles published in a language other than English. A total of 667 articles met the criteria for full article review. A detailed review at this stage further excluded 632 articles, primarily because the articles reported utility values for one respondent type only (either child self-report or proxy report) (Figure 1). Other reasons for exclusion at the full article review stage included methodological studies without evidence directly relevant to this review, and only combined child and proxy utility data provided. Finally, data were extracted from 35 articles, of which 32 articles reported utilities generated from the application of MAUIs and 3 articles reported direct valuation methods. Included articles had a quality score that ranged from 11 to 13, signifying moderate to high quality of reporting based on our quality appraisal. Of the 32 articles reporting utilities generated from the application of MAUIs, 29 articles (91%) reported dyad child and proxy-reported overall utilities (Table 1) and three articles (9%) reported only single-attribute utilities (Belfort et al., 2011; Fluchel et al., 2008; Fu et al., 2006) (Table 3) .

Types of valuation methods used

Six different types of MAUIs were represented in the included studies, namely HUI2, HUI3, EQ-5D, QWB, PAHOM and an unnamed utility instrument. The HUI3 (22 out of 32 studies) and variants of the VAS (n=10 studies, including EQ-5D VAS and EQ-5D-Y VAS) represented the most common indirect and direct methods used to generate utilities, respectively. As HUI2 and HUI3 have independent but complimentary descriptive systems, the HUI2 was mostly reported alongside the HUI3 within the same study (n=10 articles) and both instruments, whilst used predominantly in studies of cancers, were also applied in other health conditions and samples of the general population. In contrast, none of the identified studies that used the EQ-5D/~~-~~EQ-5D-Y and other MAUIs focussed on cancers (Table 1). The review included the use of only one condition-specific MAUI, the Paediatric Asthma Health Outcome Measure (PAHOM)(Chiou et al., 2005).

Study characteristics

Studies were initially grouped into those related to cancer (Table 2) and other health conditions (Table 3). In two studies, participants with cancer represented only a small proportion of the total study samples and they were therefore grouped with ~~other~~ studies in Table 3 covering other health conditions (Morrow et al., 2012; Sung et al., 2004). Of the 10 articles focused exclusively on cancers, seven were conducted with cancer survivors and three were conducted with patients with active cancers. Studies in cancers included participants with a variety of cancers, with brain cancer being the most common form of cancer reported (Table 2). Similarly, a wide range of health conditions were reported in the other health conditions group, including congenital disorders (~~low birth weight,~~ cystic fibrosis, fragile x-syndrome), chronic conditions (diabetes, asthma), acute health conditions (sever bacteraemia, pelvic inflammatory diseases), and behavioural disorders.

Studies mostly reported mean or median utilities with only four studies reporting mean differences without reporting means and/or medians for each cohort (Fluchel et al., 2008; Fu et al., 2006; Kulpeng et al., 2013; Robertson et al., 2017) (Table 2 & 3). Of those studies that reported levels of agreements and co-relations with (16 studies) and without (6 studies) utilities, 6 studies used Kappa statistics, 9 studies used intra-class correlations (ICC) and 12 studies used correlation co~~-~~efficients (Table 2 and 3). Many studies did not report the age range for the childhood sample that provided self-reports. Of those studies that reported age ranges, the minimum age of the childhood sample was 6 years (Glaser et al., 1999) and maximum was 18 years (Morrow et al., 2012; Sung et al., 2004). When comparing age groups, studies that reported on the application of direct valuation methods (e.g. VAS, SG, TTO) were all administered to adolescents (aged 13-17 years) and the mode of administration was exclusively face-to-face interviews. In contrast, MAUIs were administered in both children aged ≤12 years and adolescent samples. Types of proxy~~ies~~ used varied and included parents, caregivers and health professionals (doctors, nurses, physiotherapists). Notably, the majority of the identified studies (n=31) used parents and/or caregivers as the primary proxy group(s). Only two studies (Barr et al., 1999; Vermeulen et al., 2017b) exclusively used health professionals as the proxy group, whilst in other studies health professionals were included alongside parents and/or caregiver perspectives (Table 2 and 3). Face-to-face interview was more commonly used as the primary mode of administration relative to self-administered survey. Across the modes of administration, face-to-face interviews had the highest rates of completion whereas online surveys generally had poor response rates (Table 2 & 3).

In comparison to studies on other health conditions (Table 3), studies in cancers (Table 2) were generally based upon smaller sample sizes. All studies applying MAUIs used scoring algorithms derived from general adult populations with the exception of two studies that used the PAHOM and EQ-5D-Y (Gerald et al., 2012; Gusi et al., 2014) that applied scoring algorithms derived from childhood populations. Most of the studies that used the HUI2 and HUI3 applied scoring algorithms, based upon the preferences of adults, specifically, the HUI2 utilities were originally obtained from Canadian parents who were asked to imagine themselves as 10 years children when valuing the states whilst HUI3 utilities were obtained from the members of the general Canadian population (Horsman et al., 2003). In terms of quality of reporting, all studies included in this systematic review scored between 11 and 13 points out of the maximum score of 15, indicating good reporting quality according to our quality appraisal criteria (Appendix 1).

For the studies reporting utilities with a health condition, inconsistency in the direction of discrepancy in inter-rater gap between child-proxy dyads was evident, with proxy underestimation (Barr et al., 1999; Baumann et al., 2016; Brunner et al., 2004; Bull et al., 2014; Chevreul et al., 2016; Morrow et al., 2012; Penn et al., 2011; Saigal et al., 1999; Sung et al., 2004; Verrips et al., 2001), overestimation (Belfort et al., 2011; Chadha et al., 2010; Czyzewski et al., 1994; Fu et al., 2006; Glaser et al., 1999; Iskrov et al., 2015; Lee et al., 2011; Robertson et al., 2017; Vermeulen et al., 2017a; Wolke et al., 2013) or comparable results (Bull et al., 2014; Cheng et al., 2011). Health condition-specific appraisals are detailed below.

Appraisal of cancer-related studies

Types of proxies

Generally, parents or caregivers of children and adolescents with cancers generated lower utilities (indicating more significant health impairments) relative to child self-reports. The magnitude of this difference in utilities was significantly higher for children with severe health conditions and/or in receipt of aggressive cancer treatment (Bull et al., 2014). For children and adolescents with mild cancers, the difference in overall utilities between self and parent proxy-reports ranged from 0 to 0.02 (relative to child self-report), but for severe cancers parents generated utility values for their children that were up to 0.17 lower relative to child self-reports (Bull et al., 2014; Fluchel et al., 2008). In terms of differences in single-attribute utilities between child-proxy pairs, parents and carers significantly underestimated utilities for psychological attributes (emotion and cognition) in contrast to physical function attributes (ambulation, dexterity, vision, hearing, speech) of the HUI3 (Fluchel et al., 2008; Penn et al., 2011). Similarly, nurses acting as proxies for children in their care generated similar patterns of underestimating single-attribute utilities for psychological attributes of the HUI3 (Barr et al., 1999). In terms of other proxy groups, assessment of HRQoL by health professionals, particularly treating physicians, generated significantly higher overall HUI3 utilities, up to 0.17 higher, relative to child self-reports (Fu et al., 2006; Glaser et al., 1999)~~, indicating less severe health impairments~~.

Multi-attribute Utility Instruments

Both children and adolescent self-reports and their corresponding proxy-reports generated higher utilities using the HUI3 when compared to the HUI2 (Fu et al., 2006; Glaser et al., 1999). The magnitude of the mean differences between child self-reports and proxy-reports was generally higher using the HUI3 than for the HUI2 for parent-child dyads (HUI3, 0.08 to 0.09 vs HUI2, 0.03 to 0.06) (Fu et al., 2006; Glaser et al., 1999), physician-child dyads (HUI3, 0.16 to 0.17 VS HUI2, 0.07 to 0.11) (Fu et al., 2006; Glaser et al., 1999), nurse-child dyads (HUI3=0.22 vs HUI2, 0.12) (Barr et al., 1999) and physiotherapist-child dyads (HUI3, 0.10 vs HUI2, 0.05) (Glaser et al., 1999).

~~Cheung and colleagues used several direct valuation methods (including VAS and TTO) in adolescents that developed febrile neutropenia following cancer treatment. The study revealed that parent proxy-reports tended to underestimate utilities relative to child self- reports (Cheng et al., 2011). Only one study disaggregated utility values by level of cancer severity; it reported that disease severity was associated with increased divergence in utilities between those generated by proxy and child-self reports (Bull et al., 2014). We did not find any other study to back-up these findings.~~

For both the HUI2 and HUI3, several studies have reported moderate to high levels of agreement between proxy-reports and child-reports on multi-attribute utilities, as well as single-attribute utilities for physical attributes within HUI classification systems (Banks et al., 2008; Barr et al., 1999; Fluchel et al., 2008; Fu et al., 2006; Glaser et al., 1999). The correlations between proxy and child-reports were also generally reported to be high for multi-attribute utilities and single-attribute utilities for physical attributes of the HUI3 (Penn et al., 2011). In contrast, for psychological attributes (sensation, emotion, cognition), correlations and levels of agreement based on child and parent-proxy reports tended to be lower (Table 4). Interestingly, for the pain attribute within HUI classification systems, there was moderate to high agreement in utility outcomes based on child self-reports and all proxy-reports with the exception of two studies that reported proxy generated single-attribute utilities correlated poorly (Fu et al 2006, proxy= physicians ;ICC=0.08 and Penn et al 2011, proxy= parents, Spearman's rho=0.20) with those from child self-reports (Fu et al., 2006; Penn et al., 2011).

Direct valuation methods

Cheung and colleagues used several direct valuation methods (including VAS and TTO) in adolescents that developed febrile neutropenia following cancer treatment. The study revealed that parent proxy-reports tended to underestimate utilities relative to child self-reports. (Cheng, et al. 2011)~~.~~ Only one study disaggregated utility values by level of cancer severity; it reported that disease severity was associated with increased divergence in utilities between those generated by proxy and child-self reports. (Bull, et al. 2014)~~.~~ We did not find any other study to back-up these findings.

Appraisal of studies in other health conditions

Studies relating to other health conditions used a range of MAUIs (HUI2, HUI3, EQ-5D-3L, EuroQoL-5D-5L, EuroQoL-5D-Y, QWB and PAHOM) as well as vignettes for hypothetical health states. Similar to the studies in cancers, the HUI3 (13 studies) was the most commonly applied instrument in other health conditions, followed by variants of the EQ-5D. The utility values derived from child self-reports and parent proxy-reports in other health conditions were bi-directional, with parent proxies generally more likely to underestimate than overestimate utility values relative to those from child and/or adolescent self-reports. The magnitude and direction of differences in utilities between proxy-child dyads was varied by a range of factors including the type of health condition (chronic vs acute), severity of health condition, mode of administration and age of the childhood sample.

Types of proxies

Studies by Bauman et al (2016) and Verrips et al (2013) in adolescents that were born preterm and/or at very low birth weight reported that parent proxies underestimated multi-attribute utilities and single-attribute utilities for the HUI3 relative to child self-reports (Baumann et al., 2016; Verrips et al., 2001). Comparing modes of administration, face-to-face administration yielded similar utility values between child-parent dyads, whilst for postal surveys and telephone interviews parent proxies generated higher utility values for their children relative to child self-reports (Verrips et al., 2001). In one recent study, adolescents (>12 years) who were born preterm and/or at low birth weight generated significantly lower HUI3 multi-attribute utility values for their health states than proxy-reports by their parents (Wolke et al., 2013). Using a direct valuation approach, Saigal et al (1999) demonstrated that parents overestimated utility values for vignettes of four hypothetical health states compared to child valuations, whereas health professionals underestimated utility values for two hypothetical severe health states, but overestimated for two less severe health states

(Saigal et al., 1999). In a separate study of children and adolescents with serious behavioural disorders, health professionals overestimated but parents underestimated utility values based on the EQ-5D-3L relative to child and adolescent self-reports (Vermeulen et al., 2017a).

Multi-attribute Utility Instruments

For one congenital health condition (Spina bifida), parents underestimated multi-attribute utilities values based on the HUI3 relative to child self-reports (Sims-Williams et al., 2017). In cystic fibrosis, parents significantly overestimated utilities based on both the QWB (Czyzewski et al., 1994) and the EQ-5D-3L (Iskrov et al., 2015) relative to child self-reports. Similarly, Chevereul et al (2016) reported that caregiver proxy-reports generated significantly higher utilities (by 0.26 when assessed with the EQ-5D-3L) relative to child self-reports in a study of children and adolescents living with Fragile X syndrome (an inherited genetic disease) (Chevreul et al., 2016).

In obese children and adolescents who were at risk of developing diabetes or with diabetes, parents overestimated utilities both on the HUI3 and EQ-5D-Y (Lee et al., 2011; Rhodes et al., 2011; Robertson et al., 2017). Hanberger et al (2009) reported similar HUI3 utilities based on parent proxy-reports and self-reports in adolescents with type I diabetes (Hanberger et al., 2009). Parents of asthmatic adolescents underestimated overall utilities based on a condition-specific preference-based instrument (PAHOM). Notably, the PAHOM estimates utilities based on tariffs derived from a childhood population (Gerald et al., 2012).

Studies in a range of chronic health conditions that used variants of the HUI also reported that parents tended to underestimate utility values for their children relative to child self-reports (Brunner et al., 2004; Brunner et al., 2003; Morrow et al., 2012; Sung et al., 2004). When comparing the HUI2 and HUI3, parents generated accentuated underestimates of multi-attribute utilities for the HUI3 (lower by as much as 0.13) than for the HUI2 relative to child self-reports (Morrow et al., 2012; Sung et al., 2004).

Among adolescents with acute health conditions (e.g. severe bacteraemia), Kulpeng and colleagues reported that child self-reports generated higher utilities using both the HUI2 and HUI3, but lower utilities using the EQ-5D-3L, in comparison to those from carer proxy-reports (Kulpeng et al., 2013). In a randomized controlled trial in which Vermueulen and colleagues compared a new intervention (multi-systemic therapy) versus treatment as usual for anti-social behavioural disorders (Vermeulen et al., 2017c), parent proxy estimated utilities for the EQ-5D-3L demonstrated improvements over time for both types of interventions evaluated, whereas adolescent self-reported utilities demonstrated improvements over time for one intervention only (multi-systemic therapy) (Table 5).

Generally, the correlations and levels of agreement between child-proxy dyads for overall utilities and for psychological attributes (emotion, cognition, pain) were low whereas the relationships were found to be moderate to high for attributes of physical functioning (for all types of MAUIs and proxy types) (Table 5).

Direct valuation methods

For the extracted articles including direct valuation methods, parent proxy-reports generated significantly higher utilities using the VAS, TTO and SG elicitation methods relative to self-reports from adolescents (Brunner et al., 2004; Brunner et al., 2003; Lee et al., 2011; Sung et al., 2004; Trent et al., 2011).

~~Generally, the correlations and levels of agreement between child-proxy dyads for overall utilities and for psychological attributes (emotion, cognition, pain) were low whereas the relationships were found to be moderate to high for attributes of physical functioning (for all types of MAUIs and proxy types) (Table 5).~~

Appraisal of studies in general populations

With the exception of two studies conducted exclusively in school-settings (Gusi et al., 2014; Jelsma & Ramma, 2010), the remaining six studies identified by the review utilised general population or convenience samples of healthy children and adolescents as controls or comparator groups to compare against children and adolescents living with health conditions (Table 6). For both direct and indirect methods, parents and carer proxies of children and adolescents from general population samples generally overestimated or generated similar multi-attribute and single-attribute utilities relative to children and adolescent self-reports (Table 6) (Baumann et al., 2016; Belfort et al., 2011; Gusi et al., 2014; Jelsma & Ramma, 2010). An exception was the study by Belfort and colleagues (2011) in which parents underestimated on one HUI3 attribute (pain) relative to adolescents aged 12 years and over (Belfort et al., 2011). Similar to this finding, parent proxies of children and adolescents who were used as controls or comparative samples in two separate studies of cancers also tended to overestimate utilities relative to children and adolescent self-reports (Fluchel et al., 2008; Penn et al., 2011). In a comparable to studies involving children with health conditions, correlation and agreement levels between proxy-self dyads were found to be moderate to high (Fluchel et al., 2008; Penn et al., 2011) overall and for physical attributes, but poor for more subjective attributes (such as emotion and pain) (Gusi et al., 2014).

Discussion

This article systematically reviews published studies reporting childhood utilities generated directly or indirectly from child and proxy (parents, carers or health professionals) dyads. Overall, this review demonstrates that discordance exists between self and proxy-reported utilities, ~~and single attribute utilities in the case of MAUIs~~ suggesting different perspectives on descriptions and/or valuations of childhood health states.

Comparison with other studies

The findings reinforce the importance of obtaining self-reports where possible from childhood populations (rather than relying on proxy responses) when considering childhood HRQoL outcomes. Meeting this prerogative is nevertheless constrained by the practical difficulties in obtaining self-reports from certain childhood groups such as infants and very young children, children with not yet fully developed cognitive abilities and self-perception and those with intellectual disabilities. Caution is therefore required before using child self-reports as the main benchmark for comparison with proxy report (Wolke et al., 2013; Youssef et al., 2006). The l~~L~~iterature is not yet completely clear as to which part of the child-proxy dyad should be considered as a reference measurement to which the other part should be compared (Eiser & Morse, 2001; Petrou, 2003; Ratcliffe et al., 2012). However, insofar as different health utilities have a non-trivial impact on the outcome of cost-utility analyses evaluating childhood healthcare programmes and thereby on healthcare resource allocation, it is important to understand any consistent pattern in divergence between child self-reports and proxy-reports. An analogy can be made to previous analyse~~i~~s of divergent CHU9D utility outcomes by adolescent- or adult-derived tariff sets ~~applied~~ (Ratcliffe et al., 2012): the authors in that study noted the difference in utility may be significant enough to have an impact on health care policy. Therefore, analysts conducting primary collection of childhood health utilities or secondary analysis using the utilities must similarly be aware of the impact that respondent type (in addition to valuation method and tariff set) could have on overall utilities and MAUI-derived single-attribute ~~MAUI~~ scores. Despite this, there is currently little guidance from national health technology assessment agencies on the appropriate use of self- and proxy-reports of childhood health utilities, and most attention is placed on valuation method used. For example, NICE in ~~the UK~~England recommends the use of the EQ-5D-Y for children aged 7-12 years without specifying the appropriate respondent type (National Institute of Health and Care Excellence, 2013). The 2016 US Panel on Cost-effectiveness in Health and Medicine mentions HUI2/3, EQ-5D-Y and CHU9D, again without specifying the preferred respondent type (P.J. Neumann et al., 2016).

This study is the first to systematically document the convergence (or lack of) in overall utilities and single-attribute scores (within MAUIs) for childhood health states and how this may vary by different diseases areas ~~(including~~or in general childhood population samples~~health)~~. Previous systematic reviews of primary studies measuring childhood health utilities focused on specific diseases. For example, van Litsenburg and colleagues document~~ the~~ health utilities for active acute lymphoblastic leukaemia patients and survivors and categorise the overall utilities by respondent type (self-report and parent-, nurse- and physician-reports) (van Litsenburg et al., 2014). Their study does not however document the divergence in single-attribute scores within the HUI2 and HUI3 despite significant disparity in self-proxy divergence by different health domain (see Tables 3 and 4). The review by Pickard and colleagues similarly focus on acute lymphoblastic leukaemia and cover both preference-based and non-preference-based HRQoL measures (Pickard et al., 2004). Likewise, reviews by Kua and Davis (Kua & Davis, 2016), Janssens and colleagues (Janssens et al., 2016) and Brown and colleagues (Brown et al., 2018) focus on childhood asthma, neurodisability (assessed by preference-based and non-preference-based instruments) and childhood obesity, respectively, without concentrated attention on self-proxy dyad divergence. The review by Jardine and colleagues compares child self-reports and proxy-reports for both preference-based and non-preference-based instruments, but restricts the focus to congenital health conditions (Jardine et al., 2014). Moreover, the review by Noyes and Edwards covers all childhood health conditions and reports self-proxy dyad divergence measured by ICC and Kappa statistics but restricts attention to studies using EQ-5D despite the much more widespread use of MAUIs such as HUI2 and HUI3 in childhood populations (Noyes & Edwards, 2011). Finally, the review by Kwon and colleagues cover all childhood health conditions and all preference-based HRQoL instruments and approaches (including direct valuation methods and MAUIs) and investigates the independent effect of respondent type on overall HUI3 utilities and VAS scores by meta-regression. They find that controlling for health condition categories and other methodological factors, proxy-reports by parents result in significantly higher HUI3 utilities than child self-reports and those by parents and healthcare professionals in higher VAS scores than child self-reports (Kwon et al., 2018). However, the review does not investigate the self-proxy divergence for individual health domains within MAUIs or how this divergence may differ across disease areas and within general childhood populations. This ~~study~~ review includes studies which were rejected by Kwon and colleagues at full-article assessment stage for not reporting overall multi-attribute utilities and hence offers a more comprehensive investigation of the issue of self-proxy dyad divergence, which is in turn validated by the finding that the direction and magnitude of divergence differ by health domain~~s~~.

Among the studies that used the HUI2 and HUI3 in combination, the magnitude of difference in utility values generated by child self-reports and proxy-reports was consistently higher for the HUI3 (range: 0.06 to 0.28) than for the HUI2 (range: 0.01 to 0.12). The larger differences in child versus proxy generated utility values for the HUI3 could be partly attributable to its wider attribute and level coverage than the HUI2, and wider utility scale range, the effects of which should be considered in terms of its flexibility and sensitivity for capturing high levels of impairment. Furthermore, despite overlap between the HUI2 and HUI3, there are significant differences in the underlying constructs being measured by these two systems. (Horsman et al., 2003) Studies that have compared the HUI2 and HUI3 report that the two instruments yield substantially different results, particularly for people with serious health conditions. (Marra et al., 2005; P. J. Neumann et al., 2000) Caution is therefore required when recommending HUI2 and HUI3 as interchangeable rather than complementary utility assessment tools especially when proxy assessment is being considered.

Direction and magnitude of inter-rater utility assessments

Despite the implications for childhood healthcare decision making mentioned above, the results from this systematic review ~~suggested there~~indicated that~~suggested~~ there was no consistent pattern in direction or magnitude of difference in the inter-rater gap (with evidence of both proxy under and over estimation relative to children) in studies of children and adolescents with clinical conditions. However, proxy overestimation of utilities relative to those generated by children and adolescents was clearly evident in healthy general childhood population samples. At the attribute or dimension level, the findings from this study concord with the wider child health and HRQoL literature in that proxy reports generally have fair to good correlations and agreement with those from childhood self-reports for overall utilities and for physical attributes, but there is often poor agreement for psychosocial attributes (Jardine et al., 2014). ~~Due to significant heterogeneity among the included studies, it is difficult to pinpoint specific reasons for these findings.~~ We also found that there was higher agreement between child and proxy-reports for physical attributes (e.g., vision, ambulation, dexterity, pain)- in comparison with more subjective psychosocial attributes (e.g. emotional, cognitive) of HRQoL (Jardine et al., 2014; Petrou, 2003; Tarride et al., 2010). Due to significant heterogeneity among the included studies, it is difficult to pinpoint specific reasons for these findings. However, Jardine and colleagues argue that parents are likely to be more perceptive to large effects on their children's functioning that they can observe day-in and day-out and might not have direct insights into the emotional and psychological well-being of their children (Jardine et al., 2014). Another important issue to highlight is the source of values for the utility algorithms of contributing MAUIs. With the exception of a study in children with asthma that used the PAHOM (a disease-specific MAUI) (Gerald et al., 2012), all other studies used tariffs derived from adults. This demonstrates that there is a widespread and continued use of adult algorithms to estimate childhood utility values despite clear evidence to suggest that young people value identical health states differently to adults (Brazier et al., 2017; Ratcliffe et al., 2012). As mentioned above, these utility divergences ~~by tariff set~~ may be sufficient to influence reimbursement decisions and ~~enough to alter~~ healthcare policy (Ratcliffe et al., 2012). Further studies to derive childhood-specific MAUIs and scoring algorithms are required to address this critical issue.

Characteristics of the included studies

This review has highlighted significant heterogeneity in the literature related~~ing~~ to valuation methods, MAUIs used, source of valuations for generating scoring algorithms for MAUIs and reporting approaches in childhood utility assessment. The majority of the included studies were cross-sectional and, due to inherent subjectivity in utility assessment, factors such as context and other relevant experiences at the time of the assessment could have influenced the assessment and corresponding utilities generated (Jardine et al., 2014). Furthermore, the studies included in this review applied a variety of different methods to a myriad of populations and a variety of health conditions (Table 2 and Table 3), hence the utilities generated may not be directly comparable between studies. This further adds~~ed~~ complexities to the interpretation of the findings. Whilst there may be some consistency in the direction and magnitude of the inter-rater gap when examining utilities within specific health conditions (with the exception of cancers), the paucity of studies eliciting both child and proxy utilities means that the evidence base remains small. Longitudinal studies that assess HRQoL at repeated time intervals in dyad samples using identical methods and within a single health condition would provide a stronger evidence base from which to draw more definitive conclusions.

~~Among the studies that used the HUI2 and HUI3 in combination, the magnitude of difference in utility values generated by child self-reports and proxy-reports was consistently higher for the HUI3 (range: 0.06 to 0.28) than for the HUI2 (range: 0.01 to 0.12). The larger differences in child versus proxy generated utility values for the HUI3 could be partly attributable to its wider attribute and level coverage than the HUI2, and wider utility scale range, the effects of which should be considered in terms of its flexibility and sensitivity for capturing high levels of impairment. Furthermore, despite overlap between the HUI2 and HUI3, there are significant differences in the underlying constructs being measured by these two systems. Studies that have compared the HUI2 and HUI3 report that the two instruments yield substantially different results, particularly for people~~

~~with serious health conditions. Caution is therefore required when recommending HUI2 and HUI3 as interchangeable rather than complementary utility assessment tools especially when proxy-assessment is being considered~~.

Direct vs indirect valuation methods

Our findings also highlighted that different valuation methods (direct vs indirect) generate divergent directions of the inter-rater gap in utility values. Studies such as those by Brunner et al (Brunner et al., 2003) and Vermeulen et al (Vermeulen et al., 2017a) show that parent proxies underestimate utility values based on the HUI3 and EQ-5D, but overestimate utility values based on the SG and VAS in cohorts where both sets of measures were applied. In contrast, other studies (Iskrov et al., 2015; Kulpeng et al., 2013; Robertson et al., 2017) reported that parents overestimated utility values in comparison to those from their children irrespective of the valuation method used. The study by Sung et al (Sung et al., 2004) demonstrated that parents consistently underestimated utility values generated by the HUI2, HUI3, VAS and TTO, although this was not replicated for the SG where both parent and child generated utilities were broadly consistent. Tarriade et al (2010) argue that direct and indirect valuation methods may capture HRQoL differently; hence, they could yield different utility values in the same cohorts (Tarride et al., 2010). Although the evidence base in our study was too small for us to confidently attribute the differences in child-proxy utility divergence to the valuation method, it is a further reminder of the caution required in selecting the valuation method and the associated respondent type. Systematic reviews and meta-regressions of health utilities in both childhood (Kwon et al., 2018) and adult populations (McLernon et al., 2006; Peasgood et al., 2010; Wyld et al., 2012) regularly find significant differences in utility estimates between direct valuation methods such as TTO and SG and indirect MAUIs. This review further demonstrates how valuation method choices may produce not only different overall utility estimates for identical health states, but also divergence in self-proxy estimates. Future research should verify whether this dual divergence is observed in adult populations or restricted to childhood populations. ~~-~~

Types of proxies

Included studies used different proxies namely parents, teachers, physicians, physiotherapists and nurses with parents being the most commonly used proxy assessors. Different types of proxies used for the estimation of childhood utilities led to differences in the direction and magnitude of inter-rater effects in utility estimates.(Fluchel et al., 2008) Of the four relevant studies identified by this review, three reported that treating physicians for cancer patients tended to overestimate utilities for childhood health states in comparison to those generated by parents (Fluchel et al., 2008; Fu et al., 2006; Glaser et al., 1999). This reflects the findings in literature that assessments by proxies could be influenced by their knowledge about health, health care and by their own current health status (Petrou, 2003; Tarride et al., 2010).

~~As noted above, we also found that there was higher agreement between child and proxy-reports for physical attributes (e.g., vision, ambulation, dexterity, pain) in comparison with more subjective attributes (e.g. emotional, cognitive) of HRQoL (Jardine et al., 2014; Petrou, 2003; Tarride et al., 2010). Jardine and colleagues argue that parents are likely to be more perceptive to large effects on their children's functioning that they can observe day-in and day-out and might not have direct insights into the emotional and psychological well-being of their children (Jardine et al., 2014). Another important issue to highlight is the source of values for the utility algorithms of contributing MAUIs. With the exception of a study in children with asthma that used the PAHOM (a disease-specific MAUI) (Gerald et al., 2012), all other studies used tariffs derived from adults. This demonstrates that there is a widespread and continued use of adult algorithms to estimate childhood utility values despite clear evidence to suggest that children value identical health states differently to adults (Brazier et al., 2017; Ratcliffe et al., 2012). As mentioned above, the utility divergence by tariff set may be enough to alter healthcare policy (Ratcliffe et al., 2012). Further~~

~~studies to derive childhood-specific MAUIs and scoring algorithms are required to address this critical issue.~~

Strengths and limitations

A strength of this study is that it adhered to the standards of PRISMA guidelines and followed its checklist to guide reporting essential information from the included studies. This review also revealed that there is a paucity of studies and lack of information on dyad child self and proxy (adult parent and/or health professional) utilities elicited at the same time point. Potential limitations include the sole reliance on narrative synthesis owing to the wide variability in the methods (MAUIs and direct valuation methods) used and approaches to statistical analyses, which complicated the synthesis of the findings~~, we relied on a narrative synthesis only~~. The grey literature was also excluded, which may limit the comprehensiveness of the review. Ano~~O~~ther limitation of our review process was that we ~~did~~ have not provided ~~N on~~numbers for the individual reasons for study exclusion at the full-text ~~level~~stage (Figure 1). However, this should not affect the results and conclusions of our study. The PRISMA guideline (point 17) also recommends giving reasons for exclusion at each stage, but does not ask for ~~number by reason category~~disaggregation by reason. In addition, this review only included studies published in the English language.

Conclusions

U~~In conclusion, u~~tility assessments obtained from children and adolescents, and their corresponding proxies, should be viewed as complementary, rather than necessarily interchangeable. When possible, self-reported utility assessments are preferred over those from proxies if one can ensure that they are valid and reliably captured. In cases where that is not possible, both sets of assessments should be considered in clinical practice and policy settings, including those informing cost-effectiveness based decision-making, and those reports should be interpreted separately. A key finding from this review is that differences exist between utilities elicited via child self-report and their proxies. However, the direction of such differences was not homogeneous. We observed no clear pattern in the inter-rater gap both in terms of the magnitude and/or direction of differences in utilities across valuation methods. The paucity of identified studies and the heterogeneity of methods used further complicated the synthesis of findings. ~~. Therefore~~In conclusion, ~~it is also important to~~we acknowledge that there is a place for both self and proxy-reports and the inclusion of both set of assessments ~~imperative that researchers reflect alternative perspectives in their utility assessments where possible, which~~ may~~should~~ in turn provide greater insight into predictors of utility values and relationships between values generated by different respondent types.

References

Arnold, D., Girling, A., Stevens, A., & Lilford, R. (2009). Comparison of direct and indirect methods of estimating health state utilities for resource allocation: review and empirical analysis. *BMJ,* 339, b2688.

Banks, B.A., Barrowman, N.J., & Klaassen, R. (2008). Health-related quality of life: changes in children undergoing chemotherapy. *J Pediatr Hematol Oncol,* 30, 292-297.

Barr, R.D., Simpson, T., Whitton, A., Rush, B., Furlong, W., & Feeny, D.H. (1999). Health-related quality of life in survivors of tumours of the central nervous system in childhood--a preference-based approach to measurement in a cross-sectional study. *Eur J Cancer,* 35, 248-255.

Baumann, N., Bartmann, P., & Wolke, D. (2016). Health-Related Quality of Life Into Adulthood After Very Preterm Birth. *Pediatrics*.

Belfort, M.B., Zupancic, J.A., Riera, K.M., Turner, J.H., & Prosser, L.A. (2011). Health state preferences associated with weight status in children and adolescents. *BMC Pediatr,* 11, 12.

Billson, A.L., & Walker, D.A. (1994). Assessment of health status in survivors of cancer. *Arch Dis Child,* 70, 200-204.

Brazier, J., Ratcliffe, J., Salomon, J., & Tsuchiya, A. (2017). *Measuring and valuing health benefits for economic evaluation*. UK: Oxford University Press.

Brown, V., Tan, E.J., Hayes, A.J., Petrou, S., & Moodie, M.L. (2018). Utility values for childhood obesity interventions: a systematic review and meta-analysis of the evidence for use in economic evaluation. *Obes Rev,* 19, 905-916.

Brunner, H.I., Klein-Gitelman, M.S., Miller, M.J., Trombley, M., Baldwin, N., Kress, A., et al. (2004). Health of children with chronic arthritis: relationship of different measures and the quality of parent proxy reporting. *Arthritis Rheum,* 51, 763-773.

Brunner, H.I., Maker, D., Grundland, B., Young, N.L., Blanchette, V., Stain, A.M., et al. (2003). Preference-based measurement of health-related quality of life (HRQL) in children with chronic musculoskeletal disorders (MSKDs). *Med Decis Making,* 23, 314-322.

Bulamu, N.B., Kaambwa, B., & Ratcliffe, J. (2015). A systematic review of instruments for measuring outcomes in economic evaluation within aged care. *Health Qual Life Outcomes,* 13, 179.

Bull, K.S., Kennedy, C.R., Bailey, S., Ellison, D.W., & Clifford, S.C. (2014). Improved health-related quality of life outcomes associated with SHH subgroup medulloblastoma in SIOP-UKCCSG PNET3 trial survivors. *Acta Neuropathol,* 128, 151-153.

Campbell, M., Katikireddi, S.V., Sowden, A., McKenzie, J.E., & Thomson, H. (2018). Improving Conduct and Reporting of Narrative Synthesis of Quantitative Data (ICONS-Quant): protocol for a mixed methods study to develop a reporting guideline. *Bmj Open,* 8.

Canadian Agency for Drugs and Technologies in Health. (2017). Guidelines for the Economic Evaluation of Health Technologies: Canada (4th Version). [Accessed on October 30, 2018]. Available from https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines. Canada.

Chadha, N.K., Allegro, J., Barton, M., Hawkes, M., Harlock, H., & Campisi, P. (2010). The quality of life and health utility burden of recurrent respiratory papillomatosis in children. *Otolaryngology–Head and Neck Surgery,* 143, 685-690.

Chen, G., & Ratcliffe, J. (2015). A Review of the Development and Application of Generic Multi-Attribute Utility Instruments for Paediatric Populations. *Pharmacoeconomics,* 33, 1013-1028.

Cheng, S., Teuffel, O., Ethier, M.C., Diorio, C., Martino, J., Mayo, C., et al. (2011). Health-related quality of life anticipated with different management strategies for paediatric febrile neutropaenia. *Br J Cancer,* 105, 606-611.

Chevreul, K., Gandre, C., Brigham, K.B., Lopez-Bastida, J., Linertova, R., Oliva-Moreno, J., et al. (2016). Social/economic costs and health-related quality of life in patients with fragile X syndrome in Europe. *Eur J Health Econ,* 17 Suppl 1, 43-52.

Chiou, C.F., Weaver, M.R., Bell, M.A., Lee, T.A., & Krieger, J.W. (2005). Development of the multi-attribute Pediatric Asthma Health Outcome Measure (PAHOM). *Int J Qual Health Care,* 17, 23-30.

Czyzewski, D.I., Mariotto, M.J., Bartholomew, L.K., LeCompte, S.H., & Sockrider, M.M. (1994). Measurement of quality of well being in a child and adolescent cystic fibrosis population. *Med Care,* 32, 965-972.

Eiser, C., & Morse, R. (2001). Quality-of-life measures in chronic diseases of childhood. *Health Technol Assess,* 5, 1-157.

Felder-Puig, R., Frey, E., Sonnleithner, G., Feeny, D., Gadner, H., Barr, R.D., et al. (2000). German cross-cultural adaptation of the Health Utilities Index and its application to a sample of childhood cancer survivors. *Eur J Pediatr,* 159, 283-288.

Fluchel, M., Horsman, J.R., Furlong, W., Castillo, L., Alfonz, Y., & Barr, R.D. (2008). Self and proxy-reported health status and health-related quality of life in survivors of childhood cancer in Uruguay. *Pediatr Blood Cancer,* 50, 838-843.

Fu, L., Talsma, D., Baez, F., Bonilla, M., Moreno, B., Ah-Chu, M., et al. (2006). Measurement of health-related quality of life in survivors of cancer in childhood in Central America: feasibility, reliability, and validity. *J Pediatr Hematol Oncol,* 28, 331-341.

Gerald, J.K., McClure, L.A., Harrington, K.F., Moore, T., Hernandez-Martinez, A.C., & Gerald, L.B. (2012). Measurement characteristics of the pediatric asthma health outcome measure. *J Asthma,* 49, 260-266.

Glaser, A.W., Furlong, W., Walker, D.A., Fielding, K., Davies, K., Feeny, D.H., et al. (1999). Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the U.K. *Eur J Cancer,* 35, 256-261.

Gusi, N., Perez-Sousa, M.A., Gozalo-Delgado, M., & Olivares, P.R. (2014). [Validity and reliability of the spanish EQ-5D-Y proxy version]. *An Pediatr (Barc),* 81, 212-219.

Hanberger, L., Ludvigsson, J., & Nordfeldt, S. (2009). Health-related quality of life in intensively treated young patients with type 1 diabetes. *Pediatr Diabetes,* 10, 374-381.

Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes,* 1, 54.

Iskrov, G.G., Stefanov, R.S., Lopez-Bastida, J., Linertova, R., Oliva-Moreno, J., Serrano-Aguilar, P., et al. (2015). Economic Burden And Health-Related Quality Of Life Of Patients With Cystic Fibrosis In Bulgaria. *Folia Med (Plovdiv),* 57, 56-64.

Janssens, A., Rogers, M., Gumm, R., Jenkinson, C., Tennant, A., Logan, S., et al. (2016). Measurement properties of multidimensional patient-reported outcome measures in neurodisability: a systematic review of evaluation studies. *Dev Med Child Neurol,* 58, 437-451.

Jardine, J., Glinianaia, S.V., McConachie, H., Embleton, N.D., & Rankin, J. (2014). Self-reported quality of life of young children with conditions from early infancy: a systematic review. *Pediatrics,* 134, e1129-1148.

Jelsma, J., & Ramma, L. (2010). How do children at special schools and their parents perceive their HRQoL compared to children at open schools? *Health Qual Life Outcomes,* 8, 72.

Karnon, J., Edney, L., & Afzali, H. (2018). The political economy of the assessment of value of new health technologies. *J Health Serv Res Policy*, 1355819617751816.

Kua, W.S., & Davis, S. (2016). Systematic Review of Health State Utilities in Children with Asthma,. *Value Health,* 19, A557.

Kulpeng, W., Sornsrivichai, V., Chongsuvivatwong, V., Rattanavipapong, W., Leelahavarong, P., Cairns, J., et al. (2013). Variation of health-related quality of life assessed by caregivers and patients affected by severe childhood infections. *BMC Pediatr,* 13, 122.

Kwon, J., Kim, S.W., Ungar, W.J., Tsiplova, K., Madan, J., & Petrou, S. (2018). A Systematic Review and Meta-analysis of Childhood Health Utilities. *Med Decis Making,* 38, 277-305.

Lee, J.M., Rhee, K., O'Grady M, J., Basu, A., Winn, A., John, P., et al. (2011). Health utilities for children and adults with type 1 diabetes. *Med Care,* 49, 924-931.

Marra, C.A., Woolcott, J.C., Kopec, J.A., Shojania, K., Offer, R., Brazier, J.E., et al. (2005). A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med,* 60, 1571-1582.

Matza, L.S., Patrick, D.L., Riley, A.W., Alexander, J.J., Rajmil, L., Pleil, A.M., et al. (2013). Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. *Value Health,* 16, 461-479.

McLernon, D.J., Dillon, J.F., & Donnan, P.T. (2006). Health-state utilities in liver disease: A systematic review and meta-analysis. *Value in Health,* 9, A245-A246.

Morrow, A.M., Hayen, A., Quine, S., Scheinberg, A., & Craig, J.C. (2012). A comparison of doctors', parents' and children's reports of health states and health-related quality of life in children with chronic conditions. *Child Care Health Dev,* 38, 186-195.

National Institute of Health and Care Excellence. (2013). NICE Guide to the methods of technology appraisal 2013. [Accessed October 30, 2018]. Available from: http://www.nice.org.uk/article/pmg9/chapter/foreword. In NICE (Ed.). London.

Neumann, P.J., Sandberg, E.A., Araki, S.S., Kuntz, K.M., Feeny, D., & Weinstein, M.C. (2000). A comparison of HUI2 and HUI3 utility scores in Alzheimer's disease. *Med Decis Making,* 20, 413-422.

Neumann, P.J., Sanders, G.D., Russel, L.B., Siegel, J.E., & Ganiats, T.G. (2016). *Cost-Effectiveness in Health and Medicine* Oxford Oxford University Press.

Noyes, J., & Edwards, R.T. (2011). EQ-5D for the assessment of health-related quality of life and resource allocation in children: a systematic methodological review. *Value Health,* 14, 1117-1129.

Otero, S.C., Eiser, C., Wright, N.P., & Butler, G. (2013). Implications of parent and child quality of life assessments for decisions about growth hormone treatment in eligible children. *Child Care Health Dev,* 39, 782-788.

Parkin, D., & Devlin, N. (2006). Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Econ,* 15, 653-664.

Peasgood, T., Ward, S.E., & Brazier, J. (2010). Health-state utility values in breast cancer. *Expert Rev Pharmacoecon Outcomes Res,* 10, 553-566.

Penn, A., Lowis, S.P., Stevens, M.C., Shortman, R.I., Hunt, L.P., McCarter, R.J., et al. (2011). A detailed prospective longitudinal assessment of health status in children with brain tumors in the first year after diagnosis. *J Pediatr Hematol Oncol,* 33, 592-599.

Petrou, S. (2003). Methodological issues raised by preference-based approaches to measuring the health status of children. *Health Econ,* 12, 697-702.

Pharmaceutical Benefits Advisory Committee. (2016). Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee (Version 5.0). [Accessed October 30, 2018]. Available from https://pbac.pbs.gov.au/. In Australian Government Department of Health (Ed.). Australia.

Pickard, A.S., Topfer, L.A., & Feeny, D.H. (2004). A structured review of studies on health-related quality of life and economic evaluation in pediatric acute lymphoblastic leukemia. *J Natl Cancer Inst Monogr*, 102-125.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., et al. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: product of the ESRC methods programme (Version I).* . Lancaster, UK: University of Lancaster.

Ratcliffe, J., Chen, G., Stevens, K., Bradley, S., Couzner, L., Brazier, J., et al. (2015). Valuing Child Health Utility 9D Health States with Young Adults: Insights from a Time Trade Off Study. *Appl Health Econ Health Policy,* 13, 485-492.

Ratcliffe, J., Stevens, K., Flynn, T., Brazier, J., & Sawyer, M.G. (2012). 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,* 15, 730-736.

Rhodes, E.T., Prosser, L.A., Lieu, T.A., Songer, T.J., Ludwig, D.S., & Laffel, L.M. (2011). Preferences for type 2 diabetes health states among adolescents with or at risk of type 2 diabetes mellitus. *Pediatr Diabetes,* 12, 724-732.

Robertson, W., Fleming, J., Kamal, A., Hamborg, T., Khan, K.A., Griffiths, F., et al. (2017). Randomised controlled trial and economic evaluation of the 'Families for Health' programme to reduce obesity in children. *Arch Dis Child,* 102, 416-426.

Saigal, S., Stoskopf, B.L., Feeny, D., Furlong, W., Burrows, E., Rosenbaum, P.L., et al. (1999). Differences in preferences for neonatal outcomes among health care professionals, parents, and adolescents. *JAMA,* 281, 1991-1997.

Scotish Medicines Consortium. (2016). Guidance to manufacturers for completion of new product assessment form (NPAF) .[Accessed on October 30, 2018]. Available from https://www.scottishmedicines.org.uk/. In NHS Scotland (Ed.). Scotland.

Sims-Williams, H.J., Sims-Williams, H.P., Mbabazi Kabachelor, E., & Warf, B.C. (2017). Quality of life among children with spina bifida in Uganda. *Arch Dis Child,* 102, 1057-1061.

Stein, R.E., & Jessop, D.J. (1990). Functional status II(R). A measure of child health status. *Med Care,* 28, 1041-1055.

Stevens, K., & Ratcliffe, J. (2012). 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,* 15, 1092-1099.

Sung, L., Young, N.L., Greenberg, M.L., McLimont, M., Samanta, T., Wong, J., et al. (2004). Health-related quality of life (HRQL) scores reported from parents and their children with chronic illness differed depending on utility elicitation method. *J Clin Epidemiol,* 57, 1161-1166.

Tarride, J.E., Burke, N., Bischof, M., Hopkins, R.B., Goeree, L., Campbell, K., et al. (2010). A review of health utilities across conditions common in paediatric and adult populations. *Health Qual Life Outcomes,* 8, 12.

Thorrington, D., & Eames, K. (2015). Measuring Health Utilities in Children and Adolescents: A Systematic Review of the Literature. *PLoS One,* 10, e0135672.

Trent, M., Lehmann, H.P., Qian, Q., Thompson, C.B., Ellen, J.M., & Frick, K.D. (2011). Adolescent and parental utilities for the health states associated with pelvic inflammatory disease. *Sex Transm Infect,* 87, 583-587.

van Litsenburg, R.R., Kunst, A., Huisman, J., Ket, J.C., Kaspers, G.J., & Gemke, R.J. (2014). Health status utilities in pediatrics: a systematic review of acute lymphoblastic leukemia. *Med Decis Making,* 34, 21-32.

Vermeulen, K.M., Jansen, D., Buskens, E., Knorth, E.J., & Reijneveld, S.A. (2017a). Serious child and adolescent behaviour disorders; a valuation study by professionals, youth and parents. *BMC Psychiatry,* 17, 208.

Vermeulen, K.M., Jansen, D.E., Knorth, E.J., Buskens, E., & Reijneveld, S.A. (2017b). Cost-effectiveness of multisystemic therapy versus usual treatment for young people with antisocial problems. *Crim Behav Ment Health,* 27, 89-102.

Vermeulen, K.M., Jansen, D.E., Knorth, E.J., Buskens, E., & Reijneveld, S.A. (2017c). Cost-effectiveness of multisystemic therapy versus usual treatment for young people with antisocial problems: a reply to McIntosh. *Crim Behav Ment Health,* 27, 103-104.

Verrips, G.H., Stuifbergen, M.C., den Ouden, A.L., Bonsel, G.J., Gemke, R.J., Paneth, N., et al. (2001). Measuring health status using the Health Utilities Index: agreement between raters and between modalities of administration. *J Clin Epidemiol,* 54, 475-481.

Vetter, T.R., Bridgewater, C.L., & McGwin, G., Jr. (2012). An observational study of patient versus parental perceptions of health-related quality of life in children and adolescents with a chronic pain condition: who should the clinician believe? *Health Qual Life Outcomes,* 10, 85.

Wolke, D., Chernova, J., Eryigit-Madzwamuse, S., Samara, M., Zwierzynska, K., & Petrou, S. (2013). Self and parent perspectives on health-related quality of life of adolescents born very preterm. *J Pediatr,* 163, 1020-1026 e1022.

Wyld, M., Morton, R.L., Hayen, A., Howard, K., & Webster, A.C. (2012). A systematic review and meta-analysis of utility-based quality of life in chronic kidney disease treatments. *PLoS Med,* 9, e1001307.

Youssef, N.N., Murphy, T.G., Langseder, A.L., & Rosh, J.R. (2006). Quality of life for children with functional abdominal pain: a comparison study of patients' and parents' perceptions. *Pediatrics,* 117, 54-59.

Introduction:

Evidence from economic evaluations is increasingly used internationally to inform decisions about how to allocate scarce resources across health and social care systems (Karnon et al., 2018; Stevens & Ratcliffe, 2012). Health economists use a number of approaches for conducting such economic evaluations with cost-utility analysis (CUA) being a widely used approach. Within CUA, outcomes are typically captured and reported in terms of quality-adjusted life years (QALYs). The QALY combines health-related quality of life and length of life into a single metric by applying preference-based weights (or health utilities) that are typically represented on a cardinal scale anchored at “0” (representing being dead) and “1” (representing full health) to value health-related quality of life outcomes experienced at different time intervals.

A widely used approach for calculating utilities within the context of CUA is the use of multi-attribute utility instruments (MAUIs), the most well-known and widely applied of which in adult populations is the EuroQol (EQ-5D) (Bulamu et al., 2015). All MAUIs rely on an indirect method for estimating health utilities, comprising two distinct elements. Firstly, a descriptive system for the measurement of HRQoL using self or proxy completion. Secondly, a preference-based scoring algorithm (or tariff) for generating utilities for each of the health states defined by the descriptive system. The scoring algorithm is typically derived using economic modelling of general population values for a sub-set of health states defined by the descriptive system. In recent years, in the context of economic evaluation of child health programmes, several MAUIs have gained prominence specifically for application in children and adolescents. The Health Utilities Index Mark II and III (HUI2 and HUI3) are the most well established instruments in this regard, and other instruments that have been applied in childhood populations include the Child Health Utility 9D (CHU9D), Quality of Well-Being Scale (QWB), EQ-5D Youth version (EQ-5D-Y), Adolescent Health Utility Measure (AHUM), Assessment of Quality of Life 6D (AQoL-6D), and Seventeen-dimensional measure of Health-related Quality of Life (HRQoL-17D). These instruments may be categorised into childhood-specific (i.e. specifically developed and designed for application in children, e.g. CHU-9D, HUI2&3, AHUM) or non-childhood-specific (i.e. instruments originally designed for application with adults but later modified for use in children, e.g. QWB, EQ-5D-Y, AQoL-6D, HRQoL-17D) (Chen & Ratcliffe, 2015).

The other approaches to eliciting utility values are direct valuation methods in which respondents are asked to value their own and/or hypothetical health state/s, or in the case of proxy assessment are asked to value the health state under consideration from the perspective of a target individual using scaling or choice-based techniques, such as the visual analogue scale (VAS), standard gamble (SG) or time trade-off (TTO) (Arnold et al., 2009). The appropriateness of applying direction valuation methods in children has been questioned by several researchers who note that children may not have the cognitive abilities to evaluate their own or hypothetical health states on the basis of risk of death (SG) or period of life (TTO) (Ratcliffe et al., 2015; Thorrington & Eames, 2015). More generally, young children in particular are likely to lack the required linguistic skills to articulate their preferences for health states using complex concepts of gain and loss in health economic terms.

In applying preference-based approaches to the measurement of HRQoL, self-report is considered preferable wherever possible, including in child populations (Brazier et al., 2017). However, the psychometric literature suggests that during the early stages of childhood in particular, cognitive and communication competencies are likely to constrain the potential for self-report (Eiser & Morse, 2001; Matza et al., 2013; Stein & Jessop, 1990). Researchers working in this context therefore face the challenge of identifying appropriate age and cognition levels for eliciting self-reported descriptions of HRQoL from children. As a consequence, proxy (parent and/or health professional) reports are often used in lieu of child self-reports despite the fact that self-assessment is generally considered preferable for health economic assessments (Otero et al., 2013). Whilst it has been reported in the HRQoL literature that proxy reports (in particular from parents) may under-estimate more objective aspects of children’s’ functioning relative to child self-reports (Vetter et al., 2012; Youssef et al., 2006), empirical investigations have largely overlooked the implications for economic evaluations and health technology assessments reliant on utilities (Otero et al., 2013). Beyond the descriptive task, alternative sources of values for descriptions of childhood HRQoL outcomes, whether applied within direct or indirect measurement approaches, may also generate discrepancies in utilities (Petrou, 2003). Consequently, discrepancies between different type of respondents in the description and/or preference-based valuation of childhood HRQoL outcomes have potential implications for the findings of economic evaluations of child health programmes.

Previous studies of childhood health utilities primarily focused on specific diseases and were limited to investigation of the independent effects of respondent type (Kwon et al., 2018; Pickard et al., 2004; van Litsenburg et al., 2014). Furthermore, those studies did not assess the self-proxy convergence (or divergence) for overall utilities and individual health domains within MAUIs or how utilities may differ between children and proxies across disease areas. This study therefore aimed to systematically assess the degree of convergence (or divergence) in childhood utilities generated directly or indirectly in studies that included simultaneous dyad self and proxy (parent and/or health professional) assessment of childhood HRQoL outcomes.

Methods

Literature search

The primary objective of this systematic review was to conduct a comprehensive systematic review of primary utility data for childhood conditions with a specific focus on child-proxy dyad comparisons. The time horizon for this review covered articles published online before 30th June 2017. The literature search strategy and search terms mirrored those for a previous systematic review and meta-analysis of primary utility data for childhood conditions. (Kwon et al., 2018) The search terms (Appendix 1) were applied in six different online databases (PubMed, Embase, Web of Science, PsychoINFO, EconLit, CINAHL and Cochrane Library). The Paediatric Economic Database Evaluation (PEDE) was also searched to identify CUAs published between 1990 and 2016 that incorporated primary estimation of health utilities.

We searched for original peer reviewed articles that reported dyad childhood self- and proxy (adult parent and/or health professional) measured utilities, which were elicited at the same time point using direct or indirect methods. For the sake of completeness, articles reporting VAS scores were also included although it is acknowledged that some disagreement exists in the literature about the validity of VAS scores for reflecting health state preferences (Brazier et al., 2017; Parkin & Devlin, 2006). Primary studies reporting mean or median utilities, correlations and levels of agreement for child-proxy dyads with an identifiable health condition or descriptor were included. We excluded review articles and articles published in languages other than English. We also excluded articles reporting: (i) single- or multi-attribute utilities generated by direct or indirect valuation methods from a single respondent type, e.g. either children or parents alone; (ii) frequency distributions for single-attribute levels but not single- or multi-attribute utilities generated by indirect valuation methods; and (iii) utilities for combined samples of children and proxies (e.g., combined but not disentangled child-parent assessments).(Billson & Walker, 1994; Felder-Puig et al., 2000) Similarly, studies reporting utilities for mixed populations (i.e. combined values for general population and clinical groups combined), where utilities could not be attributed to a single identifiable group, were also excluded.

Data extraction

The initial screening of the titles and abstracts identified by the literature searches was carried out by two independent reviewers (J Khadka and J Kwon) with disagreements referred to a third reviewer (SP) for the final assessment for inclusion. Two reviewers (J Khadka and J Kwon) subsequently reviewed all articles that met the criteria for inclusion in the full article review (Figure 1). At this stage, a fourth reviewer (JR) independently reviewed 25% of articles to confirm decisions to include or exclude based upon the specified criteria. All the authors discussed in detail and agreed on inclusion of the final eligible articles. From each article that met our inclusion criteria, we extracted the following information using a proforma detailing bibliographic information, study design, study setting, valuation method, scoring algorithm/s applied to MAUIs (including source population and valuation method), respondent types, administration mode, sample target age(s), whether the health state/s under consideration were experienced and/or hypothetical, sample size(s), geographic setting, utility or VAS scores (including central statistics and measures of variability), response quality, statistical method(s) for analysing utilities and any methodological concerns reported by the study authors. A more detailed description of the extracted 15 variables is presented in Appendix 2. An overall reporting quality score was calculated by assigning a point to the reporting of each of these characteristics (15 is the maximum score indicating the highest quality), with the total aggregated points interpreted as the overall reporting quality score (Kwon et al., 2018). All data were entered into an Excel database.

Data analysis

We utilised narrative synthesis to describe the findings of the review. Narrative synthesis is a widely applied method in systematic reviews to synthesise findings from multiple studies where meta-analysis is not possible due to a high level of heterogeneity in the contributing evidence (Campbell et al., 2018). Narrative synthesis is particularly useful when analysts rely on the use of words and text to summarize and explain the findings of the synthesis into a coherent textual narrative. This approach involves critical appraisal of the characteristics of the study sample and defining statistical data and study findings using graphs, structural tabulations and using textual narratives to “tell the story” of the findings. This systematic review follows the guidance for narrative synthesis initially produced by the UK’s Economic and Social Research Council (Campbell et al., 2018; Popay et al., 2006).

In order to achieve a meaningful appraisal and comparison, the included studies were categorized into three main groups based upon the prevalence of the identified health condition(s) and perspective(s), namely cancer-related, other health conditions and general population-related. We used the International Classification of Diseases 10 revision (ICD-10) to aggregate samples into cancer-related and other health condition. Any general population sample that acted as a control or comparative group within a study was considered as a unique sample and classified under the 'general population' category.

Results

The literature search resulted in 63,733 individual articles (Figure 1). A total of 29,025 articles were excluded following deduplication. A further 34,041 articles were excluded at the title and abstract review stage. The main reasons for exclusion at this stage were focus on adult populations only, articles that did not report primary evidence, inaccessible articles and articles published in a language other than English. A total of 667 articles met the criteria for full article review. A detailed review at this stage further excluded 632 articles, primarily because the articles reported utility values for one respondent type only (either child self-report or proxy report) (Figure 1). Other reasons for exclusion at the full article review stage included methodological studies without evidence directly relevant to this review, and only combined child and proxy utility data provided. Finally, data were extracted from 35 articles, of which 32 articles reported utilities generated from the application of MAUIs and 3 articles reported direct valuation methods. Included articles had a quality score that ranged from 11 to 13, signifying moderate to high quality of reporting based on our quality appraisal. Of the 32 articles reporting utilities generated from the application of MAUIs, 29 articles (91%) reported dyad child and proxy-reported overall utilities (Table 1) and three articles (9%) reported only single-attribute utilities (Belfort et al., 2011; Fluchel et al., 2008; Fu et al., 2006) (Table 3) .

Types of valuation methods used

Six different types of MAUIs were represented in the included studies, namely HUI2, HUI3, EQ-5D, QWB, PAHOM and an unnamed utility instrument. The HUI3 (22 out of 32 studies) and variants of the VAS (n=10 studies, including EQ-5D VAS and EQ-5D-Y VAS) represented the most common indirect and direct methods used to generate utilities, respectively. As HUI2 and HUI3 have independent but complimentary descriptive systems, the HUI2 was mostly reported alongside the HUI3 within the same study (n=10 articles) and both instruments, whilst used predominantly in studies of cancers, were also applied in other health conditions and samples of the general population. In contrast, none of the identified studies that used the EQ-5D/EQ-5D-Y and other MAUIs focussed on cancers (Table 1). The review included the use of only one condition-specific MAUI, the Paediatric Asthma Health Outcome Measure (PAHOM)(Chiou et al., 2005).

Study characteristics

Studies were initially grouped into those related to cancer (Table 2) and other health conditions (Table 3). In two studies, participants with cancer represented only a small proportion of the total study samples and they were therefore grouped with studies in Table 3 covering other health conditions (Morrow et al., 2012; Sung et al., 2004). Of the 10 articles focused exclusively on cancers, seven were conducted with cancer survivors and three were conducted with patients with active cancers. Studies in cancers included participants with a variety of cancers, with brain cancer being the most common form of cancer reported (Table 2). Similarly, a wide range of health conditions were reported in the other health conditions group, including congenital disorders (cystic fibrosis, fragile x-syndrome), chronic conditions (diabetes, asthma), acute health conditions (sever bacteraemia, pelvic inflammatory diseases), and behavioural disorders.

Studies mostly reported mean or median utilities with only four studies reporting mean differences without reporting means and/or medians for each cohort (Fluchel et al., 2008; Fu et al., 2006; Kulpeng et al., 2013; Robertson et al., 2017) (Table 2 & 3). Of those studies that reported levels of agreements and co-relations with (16 studies) and without (6 studies) utilities, 6 studies used Kappa statistics, 9 studies used intra-class correlations (ICC) and 12 studies used correlation coefficients (Table 2 and 3). Many studies did not report the age range for the childhood sample that provided self-reports. Of those studies that reported age ranges, the minimum age of the childhood sample was 6 years (Glaser et al., 1999) and maximum was 18 years (Morrow et al., 2012; Sung et al., 2004).

When comparing age groups, studies that reported on the application of direct valuation methods (e.g. VAS, SG, TTO) were all administered to adolescents (aged 13-17 years) and the mode of administration was exclusively face-to-face interviews. In contrast, MAUIs were administered in both children aged ≤12 years and adolescent samples. Types of proxy used varied and included parents, caregivers and health professionals (doctors, nurses, physiotherapists). Notably, the majority of the identified studies (n=31) used parents and/or caregivers as the primary proxy group(s). Only two studies (Barr et al., 1999; Vermeulen et al., 2017b) exclusively used health professionals as the proxy group, whilst in other studies health professionals were included alongside parents and/or caregiver perspectives (Table 2 and 3). Face-to-face interview was more commonly used as the primary mode of administration relative to self-administered survey. Across the modes of administration, face-to-face interviews had the highest rates of completion whereas online surveys generally had poor response rates (Table 2 & 3).

In comparison to studies on other health conditions (Table 3), studies in cancers (Table 2) were generally based upon smaller sample sizes. All studies applying MAUIs used scoring algorithms derived from general adult populations with the exception of two studies that used the PAHOM and EQ-5D-Y (Gerald et al., 2012; Gusi et al., 2014) that applied scoring algorithms derived from childhood populations. Most of the studies that used the HUI2 and HUI3 applied scoring algorithms, based upon the preferences of adults, specifically, the HUI2 utilities were originally obtained from Canadian parents who were asked to imagine themselves as 10 years children when valuing the states whilst HUI3 utilities were obtained from the members of the general Canadian population (Horsman et al., 2003). In terms of quality of reporting, all studies included in this systematic review scored between 11 and 13 points out of the maximum score of 15, indicating good reporting quality according to our quality appraisal criteria (Appendix 1).

For the studies reporting utilities with a health condition, inconsistency in the direction of discrepancy in inter-rater gap between child-proxy dyads was evident, with proxy underestimation

(Barr et al., 1999; Baumann et al., 2016; Brunner et al., 2004; Bull et al., 2014; Chevreul et al., 2016; Morrow et al., 2012; Penn et al., 2011; Saigal et al., 1999; Sung et al., 2004; Verrips et al., 2001), overestimation (Belfort et al., 2011; Chadha et al., 2010; Czyzewski et al., 1994; Fu et al., 2006; Glaser et al., 1999; Iskrov et al., 2015; Lee et al., 2011; Robertson et al., 2017; Vermeulen et al., 2017a; Wolke et al., 2013) or comparable results (Bull et al., 2014; Cheng et al., 2011). Health condition-specific appraisals are detailed below.

Appraisal of cancer-related studies

Types of proxies

Generally, parents or caregivers of children and adolescents with cancers generated lower utilities (indicating more significant health impairments) relative to child self-reports. The magnitude of this difference in utilities was significantly higher for children with severe health conditions and/or in receipt of aggressive cancer treatment (Bull et al., 2014). For children and adolescents with mild cancers, the difference in overall utilities between self and parent proxy-reports ranged from 0 to 0.02 (relative to child self-report), but for severe cancers parents generated utility values for their children that were up to 0.17 lower relative to child self-reports (Bull et al., 2014; Fluchel et al., 2008). In terms of differences in single-attribute utilities between child-proxy pairs, parents and carers significantly underestimated utilities for psychological attributes (emotion and cognition) in contrast to physical function attributes (ambulation, dexterity, vision, hearing, speech) of the HUI3 (Fluchel et al., 2008; Penn et al., 2011). Similarly, nurses acting as proxies for children in their care generated similar patterns of underestimating single-attribute utilities for psychological attributes of the HUI3 (Barr et al., 1999). In terms of other proxy groups, assessment of HRQoL by health professionals, particularly treating physicians, generated significantly higher overall HUI3 utilities, up to 0.17 higher, relative to child self-reports (Fu et al., 2006; Glaser et al., 1999).

Multi-attribute Utility Instruments

Both children and adolescent self-reports and their corresponding proxy-reports generated higher utilities using the HUI3 when compared to the HUI2 (Fu et al., 2006; Glaser et al., 1999). The magnitude of the mean differences between child self-reports and proxy-reports was generally higher using the HUI3 than for the HUI2 for parent-child dyads (HUI3, 0.08 to 0.09 vs HUI2, 0.03 to 0.06) (Fu et al., 2006; Glaser et al., 1999), physician-child dyads (HUI3, 0.16 to 0.17 VS HUI2, 0.07 to 0.11) (Fu et al., 2006; Glaser et al., 1999), nurse-child dyads (HUI3=0.22 vs HUI2, 0.12) (Barr et al., 1999) and physiotherapist-child dyads (HUI3, 0.10 vs HUI2, 0.05) (Glaser et al., 1999).

For both the HUI2 and HUI3, several studies have reported moderate to high levels of agreement between proxy-reports and child-reports on multi-attribute utilities, as well as single-attribute utilities for physical attributes within HUI classification systems (Banks et al., 2008; Barr et al., 1999; Fluchel et al., 2008; Fu et al., 2006; Glaser et al., 1999). The correlations between proxy and child-reports were also generally reported to be high for multi-attribute utilities and single-attribute utilities for physical attributes of the HUI3 (Penn et al., 2011). In contrast, for psychological attributes (sensation, emotion, cognition), correlations and levels of agreement based on child and parent-proxy reports tended to be lower (Table 4). Interestingly, for the pain attribute within HUI classification systems, there was moderate to high agreement in utility outcomes based on child self-reports and all proxy-reports with the exception of two studies that reported proxy generated single-attribute utilities correlated poorly (Fu et al 2006, proxy= physicians ;ICC=0.08 and Penn et al 2011, proxy= parents, Spearman's rho=0.20) with those from child self-reports (Fu et al., 2006; Penn et al., 2011).

Direct valuation methods

Cheung and colleagues used several direct valuation methods (including VAS and TTO) in adolescents that developed febrile neutropenia following cancer treatment. The study revealed that parent proxy-reports tended to underestimate utilities relative to child self-reports. (Cheng, et al. 2011) Only one study disaggregated utility values by level of cancer severity; it reported that disease severity was associated with increased divergence in utilities between those generated by proxy and child-self reports. (Bull, et al. 2014) We did not find any other study to back-up these findings.

Appraisal of studies in other health conditions

Studies relating to other health conditions used a range of MAUIs (HUI2, HUI3, EQ-5D-3L, EuroQoL-5D-5L, EuroQoL-5D-Y, QWB and PAHOM) as well as vignettes for hypothetical health states. Similar to the studies in cancers, the HUI3 (13 studies) was the most commonly applied instrument in other health conditions, followed by variants of the EQ-5D. The utility values derived from child self-reports and parent proxy-reports in other health conditions were bi-directional, with parent proxies generally more likely to underestimate than overestimate utility values relative to those from child and/or adolescent self-reports. The magnitude and direction of differences in utilities between proxy-child dyads was varied by a range of factors including the type of health condition (chronic vs acute), severity of health condition, mode of administration and age of the childhood sample.

Types of proxies

Studies by Bauman et al (2016) and Verrips et al (2013) in adolescents that were born preterm and/or at very low birth weight reported that parent proxies underestimated multi-attribute utilities and single-attribute utilities for the HUI3 relative to child self-reports (Baumann et al., 2016; Verrips et al., 2001). Comparing modes of administration, face-to-face administration yielded similar utility values between child-parent dyads, whilst for postal surveys and telephone interviews parent proxies generated higher utility values for their children relative to child self-reports (Verrips et al., 2001). In one recent study, adolescents (>12 years) who were born preterm and/or at low birth weight generated significantly lower HUI3 multi-attribute utility values for their health states than proxy-reports by their parents (Wolke et al., 2013). Using a direct valuation approach, Saigal et al (1999) demonstrated that parents overestimated utility values for vignettes of four hypothetical health states compared to child valuations, whereas health professionals underestimated utility values for two hypothetical severe health states, but overestimated for two less severe health states

(Saigal et al., 1999). In a separate study of children and adolescents with serious behavioural disorders, health professionals overestimated but parents underestimated utility values based on the EQ-5D-3L relative to child and adolescent self-reports (Vermeulen et al., 2017a).

Multi-attribute Utility Instruments

For one congenital health condition (Spina bifida), parents underestimated multi-attribute utilities values based on the HUI3 relative to child self-reports (Sims-Williams et al., 2017). In cystic fibrosis, parents significantly overestimated utilities based on both the QWB (Czyzewski et al., 1994) and the EQ-5D-3L (Iskrov et al., 2015) relative to child self-reports. Similarly, Chevereul et al (2016) reported that caregiver proxy-reports generated significantly higher utilities (by 0.26 when assessed with the EQ-5D-3L) relative to child self-reports in a study of children and adolescents living with Fragile X syndrome (an inherited genetic disease) (Chevreul et al., 2016).

In obese children and adolescents who were at risk of developing diabetes or with diabetes, parents overestimated utilities both on the HUI3 and EQ-5D-Y (Lee et al., 2011; Rhodes et al., 2011; Robertson et al., 2017). Hanberger et al (2009) reported similar HUI3 utilities based on parent proxy-reports and self-reports in adolescents with type I diabetes (Hanberger et al., 2009). Parents of asthmatic adolescents underestimated overall utilities based on a condition-specific preference-based instrument (PAHOM). Notably, the PAHOM estimates utilities based on tariffs derived from a childhood population (Gerald et al., 2012).

Studies in a range of chronic health conditions that used variants of the HUI also reported that parents tended to underestimate utility values for their children relative to child self-reports (Brunner et al., 2004; Brunner et al., 2003; Morrow et al., 2012; Sung et al., 2004). When comparing the HUI2 and HUI3, parents generated accentuated underestimates of multi-attribute utilities for the HUI3 (lower by as much as 0.13) than for the HUI2 relative to child self-reports (Morrow et al., 2012; Sung et al., 2004).

Among adolescents with acute health conditions (e.g. severe bacteraemia), Kulpeng and colleagues reported that child self-reports generated higher utilities using both the HUI2 and HUI3, but lower utilities using the EQ-5D-3L, in comparison to those from carer proxy-reports (Kulpeng et al., 2013). In a randomized controlled trial in which Vermueulen and colleagues compared a new intervention (multi-systemic therapy) versus treatment as usual for anti-social behavioural disorders (Vermeulen et al., 2017c), parent proxy estimated utilities for the EQ-5D-3L demonstrated improvements over time for both types of interventions evaluated, whereas adolescent self-reported utilities demonstrated improvements over time for one intervention only (multi-systemic therapy) (Table 5). Generally, the correlations and levels of agreement between child-proxy dyads for overall utilities and for psychological attributes (emotion, cognition, pain) were low whereas the relationships were found to be moderate to high for attributes of physical functioning (for all types of MAUIs and proxy types) (Table 5).

Direct valuation methods

For the extracted articles including direct valuation methods, parent proxy-reports generated significantly higher utilities using the VAS, TTO and SG elicitation methods relative to self-reports from adolescents (Brunner et al., 2004; Brunner et al., 2003; Lee et al., 2011; Sung et al., 2004; Trent et al., 2011).

Appraisal of studies in general populations

With the exception of two studies conducted exclusively in school-settings (Gusi et al., 2014; Jelsma & Ramma, 2010), the remaining six studies identified by the review utilised general population or convenience samples of healthy children and adolescents as controls or comparator groups to compare against children and adolescents living with health conditions (Table 6). For both direct and indirect methods, parents and carer proxies of children and adolescents from general population samples generally overestimated or generated similar multi-attribute and single-attribute utilities relative to children and adolescent self-reports (Table 6) (Baumann et al., 2016; Belfort et al., 2011;

Gusi et al., 2014; Jelsma & Ramma, 2010). An exception was the study by Belfort and colleagues (2011) in which parents underestimated on one HUI3 attribute (pain) relative to adolescents aged 12 years and over (Belfort et al., 2011). Similar to this finding, parent proxies of children and adolescents who were used as controls or comparative samples in two separate studies of cancers also tended to overestimate utilities relative to children and adolescent self-reports (Fluchel et al., 2008; Penn et al., 2011). In a comparable to studies involving children with health conditions, correlation and agreement levels between proxy-self dyads were found to be moderate to high (Fluchel et al., 2008; Penn et al., 2011) overall and for physical attributes, but poor for more subjective attributes (such as emotion and pain) (Gusi et al., 2014).

Discussion

This article systematically reviews published studies reporting childhood utilities generated directly or indirectly from child and proxy (parents, carers or health professionals) dyads. Overall, this review demonstrates that discordance exists between self and proxy-reported utilities, suggesting different perspectives on descriptions and/or valuations of childhood health states.

Comparison with other studies

The findings reinforce the importance of obtaining self-reports where possible from childhood populations (rather than relying on proxy responses) when considering childhood HRQoL outcomes. Meeting this prerogative is nevertheless constrained by the practical difficulties in obtaining self-reports from certain childhood groups such as infants and very young children, children with not yet fully developed cognitive abilities and self-perception and those with intellectual disabilities. Caution is therefore required before using child self-reports as the main benchmark for comparison with proxy report (Wolke et al., 2013; Youssef et al., 2006). The literature is not yet completely clear as to which part of the child-proxy dyad should be considered as a reference measurement to which the other part should be compared (Eiser & Morse, 2001; Petrou, 2003; Ratcliffe et al., 2012). However, insofar as different health utilities have a non-trivial impact on the outcome of cost-utility analyses evaluating childhood healthcare programmes and thereby on healthcare resource allocation, it is important to understand any consistent pattern in divergence between child self-reports and proxy-reports. An analogy can be made to previous analyses of divergent CHU9D utility outcomes by adolescent- or adult-derived tariff sets (Ratcliffe et al., 2012): the authors in that study noted the difference in utility may be significant enough to have an impact on health care policy. Therefore, analysts conducting primary collection of childhood health utilities or secondary analysis using the utilities must similarly be aware of the impact that respondent type (in addition to valuation method and tariff set) could have on overall utilities and MAUI-derived single-attribute scores. Despite this, there is currently little guidance from national health technology assessment agencies on the appropriate use of self- and proxy-reports of childhood health utilities, and most attention is placed on valuation method used. For example, NICE in England recommends the use of the EQ-5D-Y for children aged 7-12 years without specifying the appropriate respondent type (National Institute of Health and Care Excellence, 2013). The 2016 US Panel on Cost-effectiveness in Health and Medicine mentions HUI2/3, EQ-5D-Y and CHU9D, again without specifying the preferred respondent type (P.J. Neumann et al., 2016).

This study is the first to systematically document the convergence (or lack of) in overall utilities and single-attribute scores (within MAUIs) for childhood health states and how this may vary by different diseases areas or in general childhood population samples. Previous systematic reviews of primary studies measuring childhood health utilities focused on specific diseases. For example, van Litsenburg and colleagues document health utilities for active acute lymphoblastic leukaemia patients and survivors and categorise the overall utilities by respondent type (self-report and parent-, nurse- and physician-reports) (van Litsenburg et al., 2014). Their study does not however document the divergence in single-attribute scores within the HUI2 and HUI3 despite significant disparity in self-proxy divergence by different health domain (see Tables 3 and 4). The review by Pickard and colleagues similarly focus on acute lymphoblastic leukaemia and cover both preference-based and non-preference-based HRQoL measures (Pickard et al., 2004). Likewise, reviews by Kua and Davis

(Kua & Davis, 2016), Janssens and colleagues (Janssens et al., 2016) and Brown and colleagues (Brown et al., 2018) focus on childhood asthma, neurodisability (assessed by preference-based and non-preference-based instruments) and childhood obesity, respectively, without concentrated attention on self-proxy dyad divergence. The review by Jardine and colleagues compares child self-reports and proxy-reports for both preference-based and non-preference-based instruments, but restricts the focus to congenital health conditions (Jardine et al., 2014). Moreover, the review by Noyes and Edwards covers all childhood health conditions and reports self-proxy dyad divergence measured by ICC and Kappa statistics but restricts attention to studies using EQ-5D despite the much more widespread use of MAUIs such as HUI2 and HUI3 in childhood populations (Noyes & Edwards, 2011). Finally, the review by Kwon and colleagues cover all childhood health conditions and all preference-based HRQoL instruments and approaches (including direct valuation methods and MAUIs) and investigates the independent effect of respondent type on overall HUI3 utilities and VAS scores by meta-regression. They find that controlling for health condition categories and other methodological factors, proxy-reports by parents result in significantly higher HUI3 utilities than child self-reports and those by parents and healthcare professionals in higher VAS scores than child self-reports (Kwon et al., 2018). However, the review does not investigate the self-proxy divergence for individual health domains within MAUIs or how this divergence may differ across disease areas and within general childhood populations. This review includes studies which were rejected by Kwon and colleagues at full-article assessment stage for not reporting overall multi-attribute utilities and hence offers a more comprehensive investigation of the issue of self-proxy dyad divergence, which is in turn validated by the finding that the direction and magnitude of divergence differ by health domain.

Among the studies that used the HUI2 and HUI3 in combination, the magnitude of difference in utility values generated by child self-reports and proxy-reports was consistently higher for the HUI3 (range: 0.06 to 0.28) than for the HUI2 (range: 0.01 to 0.12). The larger differences in child versus proxy generated utility values for the HUI3 could be partly attributable to its wider attribute and level coverage than the HUI2, and wider utility scale range, the effects of which should be considered in terms of its flexibility and sensitivity for capturing high levels of impairment. Furthermore, despite overlap between the HUI2 and HUI3, there are significant differences in the underlying constructs being measured by these two systems. Studies that have compared the HUI2 and HUI3 report that the two instruments yield substantially different results, particularly for people with serious health conditions. Caution is therefore required when recommending HUI2 and HUI3 as interchangeable rather than complementary utility assessment tools especially when proxy assessment is being considered.

Direction and magnitude of inter-rater utility assessments

Despite the implications for childhood healthcare decision making mentioned above, the results from this systematic review indicated that there was no consistent pattern in direction or magnitude of difference in the inter-rater gap (with evidence of both proxy under and over estimation relative to children) in studies of children and adolescents with clinical conditions. However, proxy overestimation of utilities relative to those generated by children and adolescents was clearly evident in healthy general childhood population samples. At the attribute or dimension level, the findings from this study concord with the wider child health and HRQoL literature in that proxy reports generally have fair to good correlations and agreement with those from childhood self-reports for overall utilities and for physical attributes, but there is often poor agreement for psychosocial attributes (Jardine et al., 2014). We also found that there was higher agreement between child and proxy-reports for physical attributes (e.g., vision, ambulation, dexterity, pain) in comparison with more subjective psychosocial attributes (e.g. emotional, cognitive) of HRQoL (Jardine et al., 2014; Petrou, 2003; Tarride et al., 2010). Due to significant heterogeneity among the included studies, it is difficult to pinpoint specific reasons for these findings. However, Jardine and colleagues argue that parents are likely to be more perceptive to large effects on their children's functioning that they can observe day-in and day-out and might not have direct insights into the emotional and psychological well-being of their children (Jardine et al., 2014). Another important issue to highlight is the source of values for the utility algorithms of contributing MAUIs. With the exception of a study in children with asthma that used the PAHOM (a disease-specific MAUI) (Gerald et al., 2012), all other studies used tariffs derived from adults. This demonstrates that there is a widespread and continued use of adult algorithms to estimate childhood utility values despite clear evidence to suggest that young people value identical health states differently to adults (Brazier et al., 2017; Ratcliffe et al., 2012). As mentioned above, these utility divergences may be sufficient to influence reimbursement decisions and healthcare policy (Ratcliffe et al., 2012). Further studies to derive childhood-specific MAUIs and scoring algorithms are required to address this critical issue.

Characteristics of the included studies

This review has highlighted significant heterogeneity in the literature related to valuation methods, MAUIs used, source of valuations for generating scoring algorithms for MAUIs and reporting approaches in childhood utility assessment. The majority of the included studies were cross-sectional and, due to inherent subjectivity in utility assessment, factors such as context and other relevant experiences at the time of the assessment could have influenced the assessment and corresponding utilities generated (Jardine et al., 2014). Furthermore, the studies included in this review applied a variety of different methods to a myriad of populations and a variety of health conditions (Table 2 and Table 3), hence the utilities generated may not be directly comparable between studies. This further adds complexities to the interpretation of the findings. Whilst there may be some consistency in the direction and magnitude of the inter-rater gap when examining utilities within specific health conditions (with the exception of cancers), the paucity of studies eliciting both child and proxy utilities means that the evidence base remains small. Longitudinal studies that assess HRQoL at repeated time intervals in dyad samples using identical methods and within a single health condition would provide a stronger evidence base from which to draw more definitive conclusions.

Direct vs indirect valuation methods

Our findings also highlighted that different valuation methods (direct vs indirect) generate divergent directions of the inter-rater gap in utility values. Studies such as those by Brunner et al (Brunner et al., 2003) and Vermeulen et al (Vermeulen et al., 2017a) show that parent proxies underestimate utility values based on the HUI3 and EQ-5D, but overestimate utility values based on the SG and VAS in cohorts where both sets of measures were applied. In contrast, other studies (Iskrov et al., 2015; Kulpeng et al., 2013; Robertson et al., 2017) reported that parents overestimated utility values in comparison to those from their children irrespective of the valuation method used. The study by Sung et al (Sung et al., 2004) demonstrated that parents consistently underestimated utility values generated by the HUI2, HUI3, VAS and TTO, although this was not replicated for the SG where both parent and child generated utilities were broadly consistent. Tarriade et al (2010) argue that direct and indirect valuation methods may capture HRQoL differently; hence, they could yield different utility values in the same cohorts (Tarride et al., 2010). Although the evidence base in our study was too small for us to confidently attribute the differences in child-proxy utility divergence to the valuation method, it is a further reminder of the caution required in selecting the valuation method and the associated respondent type. Systematic reviews and meta-regressions of health utilities in both childhood (Kwon et al., 2018) and adult populations (McLernon et al., 2006; Peasgood et al., 2010; Wyld et al., 2012) regularly find significant differences in utility estimates between direct valuation methods such as TTO and SG and indirect MAUIs. This review further demonstrates how valuation method choices may produce not only different overall utility estimates for identical health states, but also divergence in self-proxy estimates. Future research should verify whether this dual divergence is observed in adult populations or restricted to childhood populations. –

Types of proxies

Included studies used different proxies namely parents, teachers, physicians, physiotherapists and nurses with parents being the most commonly used proxy assessors. Different types of proxies used for the estimation of childhood utilities led to differences in the direction and magnitude of inter-rater effects in utility estimates.(Fluchel et al., 2008) Of the four relevant studies identified by this review, three reported that treating physicians for cancer patients tended to overestimate utilities for childhood health states in comparison to those generated by parents (Fluchel et al., 2008; Fu et al., 2006; Glaser et al., 1999). This reflects the findings in literature that assessments by proxies could be influenced by their knowledge about health, health care and by their own current health status (Petrou, 2003; Tarride et al., 2010).

Strengths and limitations

A strength of this study is that it adhered to the standards of PRISMA guidelines and followed its checklist to guide reporting essential information from the included studies. This review also revealed that there is a paucity of studies and lack of information on dyad child self and proxy (adult parent and/or health professional) utilities elicited at the same time point. Potential limitations include the sole reliance on narrative synthesis owing to the wide variability in the methods (MAUIs and direct valuation methods) used and approaches to statistical analyses, which complicated the synthesis of the findings. The grey literature was also excluded, which may limit the comprehensiveness of the review. Another limitation of our review process was that we have not provided numbers for the individual reasons for study exclusion at the full-text stage (Figure 1). However, this should not affect the results and conclusions of our study. The PRISMA guideline (point 17) also recommends giving reasons for exclusion at each stage, but does not ask for disaggregation by reason. In addition, this review only included studies published in the English language.

Conclusions

Utility assessments obtained from children and adolescents, and their corresponding proxies, should be viewed as complementary, rather than necessarily interchangeable. When possible, self-reported utility assessments are preferred over those from proxies if one can ensure that they are valid and reliably captured. In cases where that is not possible, both sets of assessments should be considered in clinical practice and policy settings, including those informing cost-effectiveness based decision-making, and those reports should be interpreted separately. A key finding from this review is that differences exist between utilities elicited via child self-report and their proxies. However, the direction of such differences was not homogeneous. We observed no clear pattern in the inter-rater gap both in terms of the magnitude and/or direction of differences in utilities across valuation methods. The paucity of identified studies and the heterogeneity of methods used further complicated the synthesis of findings. In conclusion, we acknowledge that there is a place for both self and proxy-reports and the inclusion of both set of assessments may in turn provide greater insight into predictors of utility values and relationships between values generated by different respondent types.

References

Arnold, D., Girling, A., Stevens, A., & Lilford, R. (2009). Comparison of direct and indirect methods of estimating health state utilities for resource allocation: review and empirical analysis. *BMJ,* 339, b2688.

Banks, B.A., Barrowman, N.J., & Klaassen, R. (2008). Health-related quality of life: changes in children undergoing chemotherapy. *J Pediatr Hematol Oncol,* 30, 292-297.

Barr, R.D., Simpson, T., Whitton, A., Rush, B., Furlong, W., & Feeny, D.H. (1999). Health-related quality of life in survivors of tumours of the central nervous system in childhood--a preference-based approach to measurement in a cross-sectional study. *Eur J Cancer,* 35, 248-255.

Baumann, N., Bartmann, P., & Wolke, D. (2016). Health-Related Quality of Life Into Adulthood After Very Preterm Birth. *Pediatrics*.

Belfort, M.B., Zupancic, J.A., Riera, K.M., Turner, J.H., & Prosser, L.A. (2011). Health state preferences associated with weight status in children and adolescents. *BMC Pediatr,* 11, 12.

Billson, A.L., & Walker, D.A. (1994). Assessment of health status in survivors of cancer. *Arch Dis Child,* 70, 200-204.

Brazier, J., Ratcliffe, J., Salomon, J., & Tsuchiya, A. (2017). *Measuring and valuing health benefits for economic evaluation*. UK: Oxford University Press.

Brown, V., Tan, E.J., Hayes, A.J., Petrou, S., & Moodie, M.L. (2018). Utility values for childhood obesity interventions: a systematic review and meta-analysis of the evidence for use in economic evaluation. *Obes Rev,* 19, 905-916.

Brunner, H.I., Klein-Gitelman, M.S., Miller, M.J., Trombley, M., Baldwin, N., Kress, A., et al. (2004). Health of children with chronic arthritis: relationship of different measures and the quality of parent proxy reporting. *Arthritis Rheum,* 51, 763-773.

Brunner, H.I., Maker, D., Grundland, B., Young, N.L., Blanchette, V., Stain, A.M., et al. (2003). Preference-based measurement of health-related quality of life (HRQL) in children with chronic musculoskeletal disorders (MSKDs). *Med Decis Making,* 23, 314-322.

Bulamu, N.B., Kaambwa, B., & Ratcliffe, J. (2015). A systematic review of instruments for measuring outcomes in economic evaluation within aged care. *Health Qual Life Outcomes,* 13, 179.

Bull, K.S., Kennedy, C.R., Bailey, S., Ellison, D.W., & Clifford, S.C. (2014). Improved health-related quality of life outcomes associated with SHH subgroup medulloblastoma in SIOP-UKCCSG PNET3 trial survivors. *Acta Neuropathol,* 128, 151-153.

Campbell, M., Katikireddi, S.V., Sowden, A., McKenzie, J.E., & Thomson, H. (2018). Improving Conduct and Reporting of Narrative Synthesis of Quantitative Data (ICONS-Quant): protocol for a mixed methods study to develop a reporting guideline. *Bmj Open,* 8.

Canadian Agency for Drugs and Technologies in Health. (2017). Guidelines for the Economic Evaluation of Health Technologies: Canada (4th Version). [Accessed on October 30, 2018]. Available from https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines. Canada.

Chadha, N.K., Allegro, J., Barton, M., Hawkes, M., Harlock, H., & Campisi, P. (2010). The quality of life and health utility burden of recurrent respiratory papillomatosis in children. *Otolaryngology–Head and Neck Surgery,* 143, 685-690.

Chen, G., & Ratcliffe, J. (2015). A Review of the Development and Application of Generic Multi-Attribute Utility Instruments for Paediatric Populations. *Pharmacoeconomics,* 33, 1013-1028.

Cheng, S., Teuffel, O., Ethier, M.C., Diorio, C., Martino, J., Mayo, C., et al. (2011). Health-related quality of life anticipated with different management strategies for paediatric febrile neutropaenia. *Br J Cancer,* 105, 606-611.

Chevreul, K., Gandre, C., Brigham, K.B., Lopez-Bastida, J., Linertova, R., Oliva-Moreno, J., et al. (2016). Social/economic costs and health-related quality of life in patients with fragile X syndrome in Europe. *Eur J Health Econ,* 17 Suppl 1, 43-52.

Chiou, C.F., Weaver, M.R., Bell, M.A., Lee, T.A., & Krieger, J.W. (2005). Development of the multi-attribute Pediatric Asthma Health Outcome Measure (PAHOM). *Int J Qual Health Care,* 17, 23-30.

Czyzewski, D.I., Mariotto, M.J., Bartholomew, L.K., LeCompte, S.H., & Sockrider, M.M. (1994). Measurement of quality of well being in a child and adolescent cystic fibrosis population. *Med Care,* 32, 965-972.

Eiser, C., & Morse, R. (2001). Quality-of-life measures in chronic diseases of childhood. *Health Technol Assess,* 5, 1-157.

Felder-Puig, R., Frey, E., Sonnleithner, G., Feeny, D., Gadner, H., Barr, R.D., et al. (2000). German cross-cultural adaptation of the Health Utilities Index and its application to a sample of childhood cancer survivors. *Eur J Pediatr,* 159, 283-288.

Fluchel, M., Horsman, J.R., Furlong, W., Castillo, L., Alfonz, Y., & Barr, R.D. (2008). Self and proxy-reported health status and health-related quality of life in survivors of childhood cancer in Uruguay. *Pediatr Blood Cancer,* 50, 838-843.

Fu, L., Talsma, D., Baez, F., Bonilla, M., Moreno, B., Ah-Chu, M., et al. (2006). Measurement of health-related quality of life in survivors of cancer in childhood in Central America: feasibility, reliability, and validity. *J Pediatr Hematol Oncol,* 28, 331-341.

Gerald, J.K., McClure, L.A., Harrington, K.F., Moore, T., Hernandez-Martinez, A.C., & Gerald, L.B. (2012). Measurement characteristics of the pediatric asthma health outcome measure. *J Asthma,* 49, 260-266.

Glaser, A.W., Furlong, W., Walker, D.A., Fielding, K., Davies, K., Feeny, D.H., et al. (1999). Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the U.K. *Eur J Cancer,* 35, 256-261.

Gusi, N., Perez-Sousa, M.A., Gozalo-Delgado, M., & Olivares, P.R. (2014). [Validity and reliability of the spanish EQ-5D-Y proxy version]. *An Pediatr (Barc),* 81, 212-219.

Hanberger, L., Ludvigsson, J., & Nordfeldt, S. (2009). Health-related quality of life in intensively treated young patients with type 1 diabetes. *Pediatr Diabetes,* 10, 374-381.

Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes,* 1, 54.

Iskrov, G.G., Stefanov, R.S., Lopez-Bastida, J., Linertova, R., Oliva-Moreno, J., Serrano-Aguilar, P., et al. (2015). Economic Burden And Health-Related Quality Of Life Of Patients With Cystic Fibrosis In Bulgaria. *Folia Med (Plovdiv),* 57, 56-64.

Janssens, A., Rogers, M., Gumm, R., Jenkinson, C., Tennant, A., Logan, S., et al. (2016). Measurement properties of multidimensional patient-reported outcome measures in neurodisability: a systematic review of evaluation studies. *Dev Med Child Neurol,* 58, 437-451.

Jardine, J., Glinianaia, S.V., McConachie, H., Embleton, N.D., & Rankin, J. (2014). Self-reported quality of life of young children with conditions from early infancy: a systematic review. *Pediatrics,* 134, e1129-1148.

Jelsma, J., & Ramma, L. (2010). How do children at special schools and their parents perceive their HRQoL compared to children at open schools? *Health Qual Life Outcomes,* 8, 72.

Karnon, J., Edney, L., & Afzali, H. (2018). The political economy of the assessment of value of new health technologies. *J Health Serv Res Policy*, 1355819617751816.

Kua, W.S., & Davis, S. (2016). Systematic Review of Health State Utilities in Children with Asthma,. *Value Health,* 19, A557.

Kulpeng, W., Sornsrivichai, V., Chongsuvivatwong, V., Rattanavipapong, W., Leelahavarong, P., Cairns, J., et al. (2013). Variation of health-related quality of life assessed by caregivers and patients affected by severe childhood infections. *BMC Pediatr,* 13, 122.

Kwon, J., Kim, S.W., Ungar, W.J., Tsiplova, K., Madan, J., & Petrou, S. (2018). A Systematic Review and Meta-analysis of Childhood Health Utilities. *Med Decis Making,* 38, 277-305.

Lee, J.M., Rhee, K., O'Grady M, J., Basu, A., Winn, A., John, P., et al. (2011). Health utilities for children and adults with type 1 diabetes. *Med Care,* 49, 924-931.

Marra, C.A., Woolcott, J.C., Kopec, J.A., Shojania, K., Offer, R., Brazier, J.E., et al. (2005). A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med,* 60, 1571-1582.

Matza, L.S., Patrick, D.L., Riley, A.W., Alexander, J.J., Rajmil, L., Pleil, A.M., et al. (2013). Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. *Value Health,* 16, 461-479.

McLernon, D.J., Dillon, J.F., & Donnan, P.T. (2006). Health-state utilities in liver disease: A systematic review and meta-analysis. *Value in Health,* 9, A245-A246.

Morrow, A.M., Hayen, A., Quine, S., Scheinberg, A., & Craig, J.C. (2012). A comparison of doctors', parents' and children's reports of health states and health-related quality of life in children with chronic conditions. *Child Care Health Dev,* 38, 186-195.

National Institute of Health and Care Excellence. (2013). NICE Guide to the methods of technology appraisal 2013. [Accessed October 30, 2018]. Available from: http://www.nice.org.uk/article/pmg9/chapter/foreword. In NICE (Ed.). London.

Neumann, P.J., Sandberg, E.A., Araki, S.S., Kuntz, K.M., Feeny, D., & Weinstein, M.C. (2000). A comparison of HUI2 and HUI3 utility scores in Alzheimer's disease. *Med Decis Making,* 20, 413-422.

Neumann, P.J., Sanders, G.D., Russel, L.B., Siegel, J.E., & Ganiats, T.G. (2016). *Cost-Effectiveness in Health and Medicine* Oxford Oxford University Press.

Noyes, J., & Edwards, R.T. (2011). EQ-5D for the assessment of health-related quality of life and resource allocation in children: a systematic methodological review. *Value Health,* 14, 1117-1129.

Otero, S.C., Eiser, C., Wright, N.P., & Butler, G. (2013). Implications of parent and child quality of life assessments for decisions about growth hormone treatment in eligible children. *Child Care Health Dev,* 39, 782-788.

Parkin, D., & Devlin, N. (2006). Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Econ,* 15, 653-664.

Peasgood, T., Ward, S.E., & Brazier, J. (2010). Health-state utility values in breast cancer. *Expert Rev Pharmacoecon Outcomes Res,* 10, 553-566.

Penn, A., Lowis, S.P., Stevens, M.C., Shortman, R.I., Hunt, L.P., McCarter, R.J., et al. (2011). A detailed prospective longitudinal assessment of health status in children with brain tumors in the first year after diagnosis. *J Pediatr Hematol Oncol,* 33, 592-599.

Petrou, S. (2003). Methodological issues raised by preference-based approaches to measuring the health status of children. *Health Econ,* 12, 697-702.

Pharmaceutical Benefits Advisory Committee. (2016). Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee (Version 5.0). [Accessed October 30, 2018]. Available from https://pbac.pbs.gov.au/. In Australian Government Department of Health (Ed.). Australia.

Pickard, A.S., Topfer, L.A., & Feeny, D.H. (2004). A structured review of studies on health-related quality of life and economic evaluation in pediatric acute lymphoblastic leukemia. *J Natl Cancer Inst Monogr*, 102-125.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., et al. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: product of the ESRC methods programme (Version I).* . Lancaster, UK: University of Lancaster.

Ratcliffe, J., Chen, G., Stevens, K., Bradley, S., Couzner, L., Brazier, J., et al. (2015). Valuing Child Health Utility 9D Health States with Young Adults: Insights from a Time Trade Off Study. *Appl Health Econ Health Policy,* 13, 485-492.

Ratcliffe, J., Stevens, K., Flynn, T., Brazier, J., & Sawyer, M.G. (2012). 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,* 15, 730-736.

Rhodes, E.T., Prosser, L.A., Lieu, T.A., Songer, T.J., Ludwig, D.S., & Laffel, L.M. (2011). Preferences for type 2 diabetes health states among adolescents with or at risk of type 2 diabetes mellitus. *Pediatr Diabetes,* 12, 724-732.

Robertson, W., Fleming, J., Kamal, A., Hamborg, T., Khan, K.A., Griffiths, F., et al. (2017). Randomised controlled trial and economic evaluation of the 'Families for Health' programme to reduce obesity in children. *Arch Dis Child,* 102, 416-426.

Saigal, S., Stoskopf, B.L., Feeny, D., Furlong, W., Burrows, E., Rosenbaum, P.L., et al. (1999). Differences in preferences for neonatal outcomes among health care professionals, parents, and adolescents. *JAMA,* 281, 1991-1997.

Scotish Medicines Consortium. (2016). Guidance to manufacturers for completion of new product assessment form (NPAF) .[Accessed on October 30, 2018]. Available from https://www.scottishmedicines.org.uk/. In NHS Scotland (Ed.). Scotland.

Sims-Williams, H.J., Sims-Williams, H.P., Mbabazi Kabachelor, E., & Warf, B.C. (2017). Quality of life among children with spina bifida in Uganda. *Arch Dis Child,* 102, 1057-1061.

Stein, R.E., & Jessop, D.J. (1990). Functional status II(R). A measure of child health status. *Med Care,* 28, 1041-1055.

Stevens, K., & Ratcliffe, J. (2012). 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,* 15, 1092-1099.

Sung, L., Young, N.L., Greenberg, M.L., McLimont, M., Samanta, T., Wong, J., et al. (2004). Health-related quality of life (HRQL) scores reported from parents and their children with chronic illness differed depending on utility elicitation method. *J Clin Epidemiol,* 57, 1161-1166.

Tarride, J.E., Burke, N., Bischof, M., Hopkins, R.B., Goeree, L., Campbell, K., et al. (2010). A review of health utilities across conditions common in paediatric and adult populations. *Health Qual Life Outcomes,* 8, 12.

Thorrington, D., & Eames, K. (2015). Measuring Health Utilities in Children and Adolescents: A Systematic Review of the Literature. *PLoS One,* 10, e0135672.

Trent, M., Lehmann, H.P., Qian, Q., Thompson, C.B., Ellen, J.M., & Frick, K.D. (2011). Adolescent and parental utilities for the health states associated with pelvic inflammatory disease. *Sex Transm Infect,* 87, 583-587.

van Litsenburg, R.R., Kunst, A., Huisman, J., Ket, J.C., Kaspers, G.J., & Gemke, R.J. (2014). Health status utilities in pediatrics: a systematic review of acute lymphoblastic leukemia. *Med Decis Making,* 34, 21-32.

Vermeulen, K.M., Jansen, D., Buskens, E., Knorth, E.J., & Reijneveld, S.A. (2017a). Serious child and adolescent behaviour disorders; a valuation study by professionals, youth and parents. *BMC Psychiatry,* 17, 208.

Vermeulen, K.M., Jansen, D.E., Knorth, E.J., Buskens, E., & Reijneveld, S.A. (2017b). Cost-effectiveness of multisystemic therapy versus usual treatment for young people with antisocial problems. *Crim Behav Ment Health,* 27, 89-102.

Vermeulen, K.M., Jansen, D.E., Knorth, E.J., Buskens, E., & Reijneveld, S.A. (2017c). Cost-effectiveness of multisystemic therapy versus usual treatment for young people with antisocial problems: a reply to McIntosh. *Crim Behav Ment Health,* 27, 103-104.

Verrips, G.H., Stuifbergen, M.C., den Ouden, A.L., Bonsel, G.J., Gemke, R.J., Paneth, N., et al. (2001). Measuring health status using the Health Utilities Index: agreement between raters and between modalities of administration. *J Clin Epidemiol,* 54, 475-481.

Vetter, T.R., Bridgewater, C.L., & McGwin, G., Jr. (2012). An observational study of patient versus parental perceptions of health-related quality of life in children and adolescents with a chronic pain condition: who should the clinician believe? *Health Qual Life Outcomes,* 10, 85.

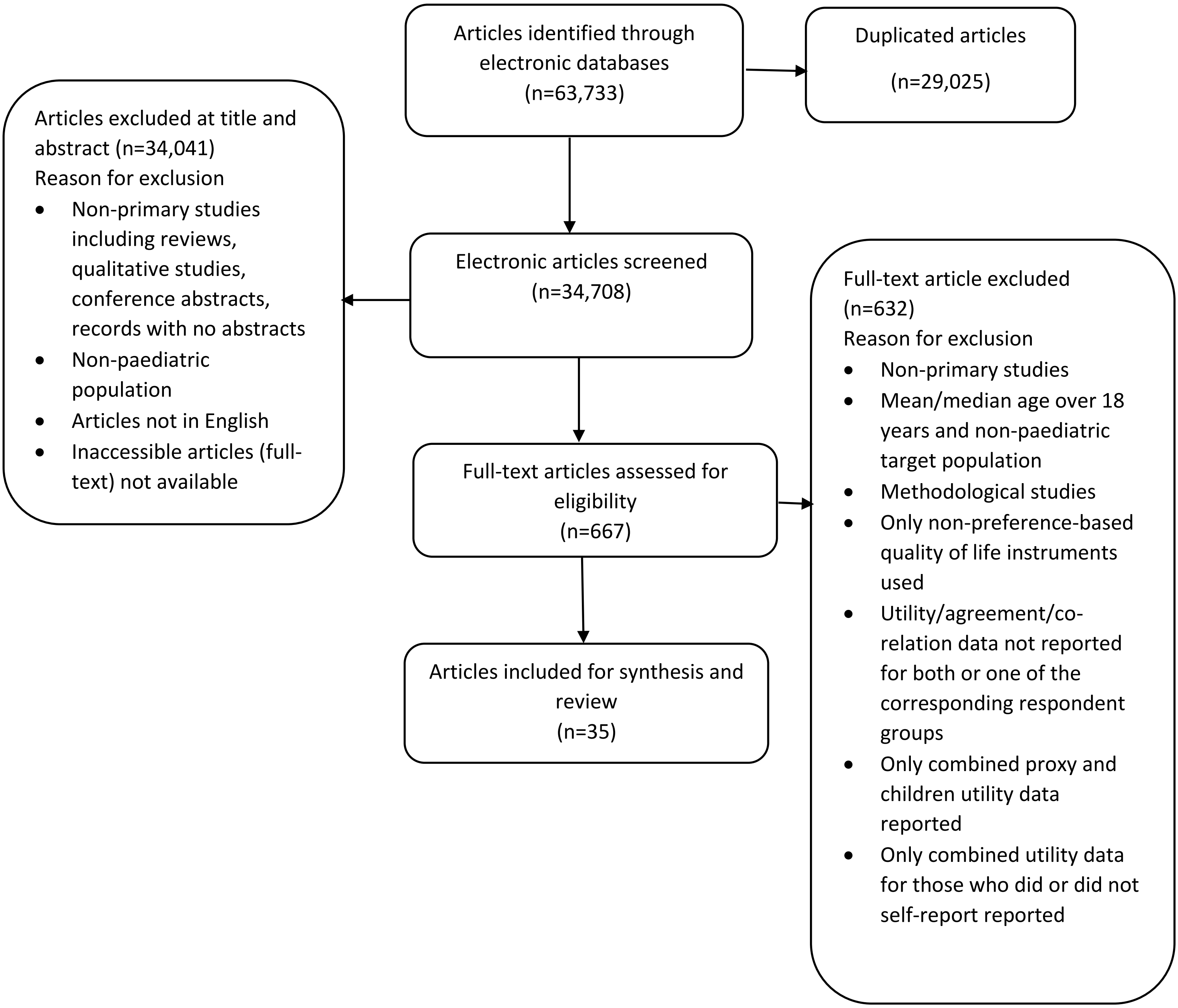
Wolke, D., Chernova, J., Eryigit-Madzwamuse, S., Samara, M., Zwierzynska, K., & Petrou, S. (2013). Self and parent perspectives on health-related quality of life of adolescents born very preterm. *J Pediatr,* 163, 1020-1026 e1022.

Wyld, M., Morton, R.L., Hayen, A., Howard, K., & Webster, A.C. (2012). A systematic review and meta-analysis of utility-based quality of life in chronic kidney disease treatments. *PLoS Med,* 9, e1001307.

Youssef, N.N., Murphy, T.G., Langseder, A.L., & Rosh, J.R. (2006). Quality of life for children with functional abdominal pain: a comparison study of patients' and parents' perceptions. *Pediatrics,* 117, 54-59.

Figures (NO AUTHOR DETAILS)

Figure 1: Preferred Reporting of Items for Systematic review and Meta-analysis (PRISMA) flow diagram showing the process of paper selection for this review



Tables (NO AUTHOR DETAILS)

Table 1: Characteristics of the included studies by valuation method

Description	Number of studies
Total studies included	35
Studies with MAUI/s	32
Studies with HUI3	22
Utility scores reported	19
No scores reported (only agreement or correlation between self-proxy pair)	3
HUI3 on cancers	10
HUI3 on other health condition	12
Studies with HUI2	11
Utility scores reported	9
No scores reported (only agreement between self-proxy pair)	2
HUI2 on cancers	8
HUI2 on other health conditions	3
HUI2 reported with HUI3	10
Studies with EQ-5D (includes EQ-5D-Y; EQ-5D-3L; EQ-5D-5L)	8
Utility scores reported	7
No scores reported (only agreement between self-proxy pair)	1
EQ-5D on cancers	0
EQ-5D on other health conditions	8
Studies with other instruments	3
QWB	1
PAHOM	1
4 Hypothetical Health States Vignette	1
Studies on cancer	0
Studies with VAS (EQ-5D VAS=6)	10
VAS scores reported	10
VAS on cancers	2
VAS on other health conditions	8
Studies with TTO	4
TTO on cancers	2
TTO on health conditions	2
SG	3
SG on Cancers	1
SG on other health conditions	2

MAUI=Multi-attribute utility Index, HUI= Health Utility Index, EQ-5D= EuroQoL-5D, QWB= Quality of Well-Being, PAHOM= Paediatric Asthma Health Outcome Measure, VAS=Visual Analogue Scale, WTP=Willingness to Pay, TTO=Time-trade-off, SG=Standard gamble

Tables (NO AUTHOR DETAILS)

Table 2: Details of the included studies on cancer and/or cancer survivors that reported dyad self and proxy reports using preference-based quality of life instruments

Author, year and country of study	Health Condition	Sample sizes: child (male%) & proxy:	Age of childhood population :mean/median; range (age range of those self-reported)	Instrument / valuation methods used	Response rate (%)	Source of tariff	Proxy type & mode of administration	MAUI utility scores reported	Results reported	Reporting quality score
Banks et al 2008; Canada	Cancer patients: leukaemia, lymphoma, brain tumour	Children=11 (65%); parents=22	Mean: 9.5; 2-18 (10-18 years)	HUI2, HUI3	72.0	Unspecified	Parents. Self-administered paper questionnaire	No	Agreement between self-proxy pair (ICC)	12
Barr et al 1999 Canada	Cancer survivors: brain tumour (PNET, medulloblastoma, astrocytoma, glioma, ependymoma)	Children=15 (46.3%); nurses=15; physicians=12;	Mean: 9.5; 1.7-17.9 (9.5 to 17.9)	HUI2, HUI3	93.2	Canadian general population of adults	Nurses & physicians. Self-administered paper questionnaire	Yes	Mean overall utility scores only & agreement between self-proxy pair for overall (ICC) and dimension scores (Kappa)	12
Bull et al 2014; UK	Cancer survivors: medulloblastoma (WNT molecular group, SHH group, non-WNT/SHH group)	Children=8-31; parents=8-33	Median: 15.7, 14.5, 14.7; 10.3-18.2 (10 to 18 years)	HUI3	49.1 to 50.9	Unspecified	Parents. Unspecified mode of administration.	Yes	Median overall utility scores	12
Cardarelli et al 2006; Italy	Cancer survivors: brain tumour; leukaemia; lymphoma; solid tumour	Children=2-15; parents=30-121	Mean: 13.3; 8-28 (8-28 years)	HUI2	93-100	Canadian general population of adults	Parents. Self-administered paper questionnaire.	Yes	Mean and median overall utility scores	13
Cheng et al (2011) Canada	Cancer patients: leukaemia; lymphoma; solid tumour; brain tumour and others	Children=43 (23.2 to 53.5%); parents=155	Median:15.4; 0-18 (12-18 years)	HUI2, HUI3; VAS,TTO	100	Unspecified	Parents. Face-to-face interview.	Yes	Median overall utility, VAS and TTO scores	12
Fluchel et al, 2008; Uruguay	Cancer survivors and general population as controls: ALL; brain tumour; Wilm’s tumour; retinoblastoma; Hodgkin’s disease; NHL; AML; STS; neuroblastoma; Ewing’s sarcoma; ovarian sarcoma; osteogenic sarcoma; General health (control)	Children=81 Parents=94	Mean: 12.2 to 13.6; 7-27	HUI3	81.4	Canadian general population of adults	Parents, self-assessed by children. Face-to-face interview.	Yes	Mean difference between self and proxy overall utility and attribute scores and agreement (ICC) for self-proxy pair	13

Fu et al 2006; El Salvador, Honduras, Nicaragua, Panama	Cancer survivors: ALL; leukaemia; lymphoma; Renal tumour; germ cell Cancer patients; retinoblastoma; bone tumour; CNS tumour; STS; carcinoma	Children=165-175 (56%); parents=140-180; physicians=200-201	Median: 12.8; 3.4-25.8	HUI2, HUI3	78.2 to 100	Canadian general population of adults	Parents or physicians. Parents, Face-to-face interview or Self-administered paper questionnaire.	Yes	Mean difference between self and proxy overall utility and attribute scores and agreement (ICC or Kappa) for self-proxy pair overall and attribute scores	12
Glaser et al 1999; UK	Cancer survivors: brain tumour (PNET, astrocytoma, medulloblastoma, ependymoma)	Children=25-26; parents= 23-28	Mean: 10.5; 6-16	HUI2, HUI3	90	Canadian general population of adults	Parents, physiotherapist or physicians. Self-administered paper questionnaire.	Yes	Mean overall utility scores and agreement (Kappa) between self-proxy pair for overall and attribute scores	13
Glaser et al 1999b; UK (Canada)	Cancer survivors: brain tumour (PNET, astrocytoma, medulloblastoma, ependymoma)	Children=19-24; parents=19-24	Mean: 10.5; 6-16	HUI2, HUI3	80	Canadian general population of adults	Parents, physiotherapist or physicians. Self-administered paper questionnaire.	No	Agreement (ICC and Kappa) between self-proxy pair for overall and attribute scores	13
Penn et al 2011; UK	Cancer patients: brain tumour (ependymoma, germ cell tumour, astrocytoma, medulloblastoma, meningioma, pineoblastoma, craniopharyngioma); Comorbidity (hydrocephalus); General health (control)	Children=21-22 (48.3 – 53.1%); parents=29-32	Median: 11.1, 10.7; 3.6-18.9	HUI3	60.4	Unspecified	Parents. Face-to-face interview.	Yes	Mean overall and dimension scores; agreement (ICC) between overall and dimension scores	13

MAUI= Multi-attribute utility instruments: HUI= Health Utility Index; ICC= Intra-class correlation; VAS= Visual Analog Scale, WTP=Willingness to pay, TTO=Time Trade Off; QWB= Quality of Well-Being

Table 3: Details of the included studies on health conditions other than cancers that reported dyad self and proxy reports using preference-based quality of life instruments

Author, year and country of study	Health Condition	Sample sizes: child (male%) & proxy: Childhood	Age of childhood population :mean/median; range (age range of those self-reported)	Instrument / valuation methods used	Response rate (%)	Source of tariff	Proxy type & mode of administration	MAUI utility scores reported	Results reported	Reporting quality score
Baumann et al 2016; Germany	Very low birthweight/very preterm birth; General health (normal birthweight/term)	Very pre-term children=190 (45.3%); parents=197 Control children=201 (53.7) & parents=204	Mean: 13; 13-26 (13-26 years)	HUI3	65.0	Canadian general population of adults	Parents. Postal survey	Yes	Mean overall utility scores only	13
Verrips et al 2001; Netherlands	Very low birthweight; Very preterm birth	Children=100 & 203 (46.2%); parents=100 &203	Mean: 14.2	HUI3	67.7	Canadian general population of adults	Parents. Postal survey or face-to-face or telephone interview.	Yes	Mean overall utility scores	13
Wolke et al, 2013; Germany	Very low birthweight; Very preterm birth (VLBW/VP), General health (Full-term, FT, controls)	Children= 260 (VLBW/VP) 282 (FT) (55.2%)	Mean: 13	HUI3	61	Canadian general population of adults	Parents. Postal survey.	Yes	Mean overall utility scores	13
Saigal et al 1999; Canada	Extremely low birthweight or extremely preterm birth children	Children=263	Mean: 8, 14.2	4 hypothetical health states	84.4	Canadian general population of adults	Parents, health professionals (physicians, nurses) Face-to-face interview.	Yes	Mean of hypothetical dimensions	11
Hanberger et al 2009; Sweden	Diabetes type I; Severe hypoglycaemia	Children=148; parents=151	Mean: 13.2	EQ-5D-3L	72.5	Unspecified	Parents. Postal survey.	Yes	Mean overall utility scores	12
Rhodes et al, 2012; USA	Overweight (insulin resistant, prediabetes); Diabetes type II	Children=108 (32.9%); parents=106	Mean: 12.8; 12-18	HUI3	75	Canadian general population of adults	Parents. Face-to-face interview.	Yes	Median overall utility scores and correlation (spearman correlation) between child-proxy pair	13
Lee et al 2011;, USA	Diabetes type I; Complications (blindness, end-stage renal disease, angina,	Children=231; parents=223	Mean: 13.7; 8-18	HUI3; TTO	82	Canadian general population of adults	Parents. Face-to-face interview.	Yes	Mean overall utility and TTO scores; correlation between self-proxy pair for utility and TTO	12

	stroke, amputation)									
Belfort et al 2011 USA	Overweight or obese; General population (healthy weight)	Children=42 overweight (57%) & =34 (47%) (healthy weights); Parents= n (not reported)	Mean: 10.8 (overweight), 11.5 (normal weight); 5-18 (8-18 years)	HUI3	46.2 to 65.3	Canadian general population of adults	Parents. Face-to-face interview	Yes	Mean & median overall utility and 3 dimension scores; correlations between child and proxy scores (Spearman's rho) for the overall and 3 dimensions (pain, emotion & cognition) for self-proxy pair	12
Robertson et al 2017; UK	Obesity	Children=63 (55.4%); parents=63	Mean: 9.46; 6-11	EQ-5D-Y, EQ-5D-Y VAS	100	Unspecified	Parents. Self-administered paper questionnaire	Yes	Mean and mean differences of overall utility and VAS scores	12
Brunner et al 2003; Canada	MSKD: juvenile rheumatoid arthritis; systemic lupus erythematosus; juvenile dermatomyositis; arthropathy secondary to haemophilia	Children=45-55 (23.5%); parents=68-80	1-18 (HUI3 and VAS by 8-18 years & SG by 10-18 years)	HUI3, SG, VAS (Linear analogue scale, Categorical scale),	81.8	Canadian general population of adults	Parents. Face-to-face interview.	Yes	Mean and median overall utility, SG, VAS scores; correlation (Spearman's rho) and agreement (ICC) for self-proxy pair	12
Brunner et al, 2004; USA	MSKD: juvenile rheumatoid arthritis; Castleman syndrome; systemic lupus erythematosus; juvenile dermatomyositis; arthritis associated with IBD; juvenile psoriatic arthritis; juvenile ankylosing spondylitis; sarcoidosis	Children=87 (23.5%); parents=119	3-18 (VAS by 8-18 years & SG by 12-18 years)	VAS (well and pain), SG	46.4	Not applicable	Parents. Face-to-face interview.	No	Mean and median SG and VAS and agreement (ICC) for self-proxy pair	12
Sims-Williams et al 2017; Uganda	Spina bifida; Comorbidities (hydrocephalus, immobility)	Children=62 (56%); caregivers=66	10-14	HUI3, VAS	97.1	Unspecified	Parents. Face-to-face interview; Self-administered paper questionnaire	Yes	Mean overall utility and VAS scores and Pearson correlations between self-proxy pair on utility and VAS scores	12
Chevreul et al 2015b; France (Denmark, England,	Fragile X syndrome	Children=53 (88.7%); caregivers=42	Mean: 10.3 (not reported)	EQ-5D-5L; EQ-5D-5L VAS	36.6	General population in adults in Denmark, England,	Caregivers. Online survey.	Yes	Mean utility and VAS scores	13

Italy, Netherlands, Scotland, Poland)						Italy, Netherlands, Scotland and Poland				
Czyzewski et al 1994, USA (USA)	Cystic fibrosis	Children=55 (48.2%); parents=199	Mean: 8.74; 0.2-17.9 (12 to 18 years)	QWB	70	US population of adults	Parents. Face-to-face interview.	Yes	Mean overall utility scores and agreement (ICC) for self-proxy pair.	13
Iskrov et al 2015; Bulgaria	Cystic fibrosis	Children=32; parents=17	Mean: 16	EQ-5D-3L VAS	69.7	Unspecified	Parents, online survey.	Yes	Median overall utility scores	11
Chadha et al 2010; Canada	Recurrent respiratory papillomatosis	Children=20 (65%); parents=20	Median: 9; 1.42-17 (9 to 17 years)	HUI3; VAS	100	Canadian general population of adults	Caregivers. Face-to-face interview.	Yes	Mean overall utility and VAS scores	13
Unger et al (2012); Canada	Asthma	Children=66 & 91; parents=66 & 91	8-17	HUI2, HUI3	64	Canadian general population of adults	Parent. Face-to-face interviewer administer	No	Agreement between self-proxy pair for overall and 3 dimensions (mobility, ambulation and emotion)	
Gerald et al 2012; USA	Asthma	Children= 64-261 (63.6 to 100); parents=64-261	6-12 (6-12 years)	PAHOM	86.7-93.3	US children with asthma (aged 6-12 years)	Parents. Face-to-face interview.	Yes	Mean overall utility scores and correlation between self-proxy pair	12
Kulpeng et al 2013; Thailand	Severe bacteraemia; Acute otitis media; Severe infection; Permanent sequelae to severe infection (epilepsy, hearing loss, pneumonia, lung disease, severe meningitis, mental retardation)	Children=74 (53.5%); parents=99	Mean: 10; 5-14	EQ-5D-3L VAS, HUI2, HUI3	100	HUI-Canadian general population of adults; EQ-5D-3L- Thai general population of adults	Caregivers. Face-to-face interview.	Yes	Mean difference of overall utility scores and correlations between self-proxy pair	13
Trent et al 2011; USA	Pelvic inflammatory disease; Permanent sequelae (ectopic pregnancy, infertility, chronic abdominal pain)	Children=134 (0%); parents=121	Mean: 16.2; 12-19	VAS, TTO	100	Not applicable	Parents. Online survey.	No	Median VAS and TTO scores	12
Vermeulen et al 2017; Netherlands	Hypothetical behavioural disorders: Oppositional defiant disorder; Conduct disorder; Disruptive behavioural disorder;	Children=51& 65; parents=51&65	Mean: 9, 15	EQ-5D-3L and VAS	100	UK general population of adults	Professionals; Self-assessed by children; Postal survey; Self-administered paper	Yes	Mean overall utility and VAS scores	13

	Comorbidities (ADHD, substance abuse); General health						questionnaire			
Vermeulen et al 2017b; Netherlands	Antisocial personality disorder: Chronically antisocial and seriously violent	Children=15-50 (62-67%); parents=10-16	Mean: 16; 12-18	EQ-5D-3L	100	UK general population of adults	Parents. Self-administered paper questionnaire	Yes	Mean difference of treatment effect on overall utility and VAS scores	13
Gusi et al 2014; Spain	General population	Children=620; parents=477	Mean: 16; 6-17 years (6-17)	EQ-5D-Y & Proxy	100	School aged children in Germany, Italy, South Africa, Spain and Sweden (aged 8-19 years)	Parents, self-administered paper questionnaire	No	Agreement (Kappa) between self-proxy pair	
Morrow et al 2010; Australia	Chronic illness: any cancer, cystic fibrosis, type1 diabetes, cerebral palsy, chronic neurological conditions, inflammatory bowel disease, liver transplant; chronic kidney diseases, autism	Children=69 (94.0%); parents=59; doctors=59	12-18 (12-18)	HUI2, HUI3 (HUI23SUE.15Q & HUI23PUE.15Q)	94	Canadian general population of adults	Parents, doctors	Yes	Mean overall utility scores, agreement (Kappa) between child-parent & child-doctor pair	
Sung et al 2004; Canada	Cancer patients; Arthritis; Haemophilia; Conditions requiring bone marrow transplant; Hypothetical asthma as example of mild disability; Hypothetical stroke as example of moderate disability; Hypothetical injury as example of severe disability	Children=19 & 21 (55%) Parents= 19 & 21	Mean: 13.7; 12-18	HUI2, HUI3, VAS, SG, TTO	81.5	Canadian general population of adults	Parents. Face-to-face interview.	Yes	Mean overall utility, VAS, SG and TTO scores and correlation (Spearman's rho) and agreement between self-proxy pair for overall utility, VAS, SG and TTO.	12

MAUI= Multi-attribute utility instruments: HUI= Health Utility Index; ICC= Intra-class correlation; MSKD= Musculo-skeletal diseases; VAS= Visual Analog Scale, WTP=Willingness to pay, TTO=Time Trade Off; QWB= Quality of Well-Being

Table 4: Details of the studies that reported dyad self and proxy reports on cancer participants : overall utility/attribute scores, agreement and correlations between child and proxy ratings

Author year	Health conditions	Proxy type/ No of child-Proxy pairs	MAUIs	Child and proxy dyad overall utility and/or domain scores: mean/median (SD/SE/(5% CI/range), mean differences where available	Direct valuation methods (VAS, TTO, SG, WP) where available	Child and proxy direct valuation methods based scores: mean/median (SD/SE/CI) where available	Child and proxy reports: Correlation/agreement (ICC, K) where available	Summary of the findings
Banks et al 2008	Cancer patients: leukaemia, lymphoma, brain tumour	Parents/	HUI2, HUI3	Not reported	Not used	-	HUI2: overall score ICC= 0.74 (0.29 to 0.92) HUI3: ICC=0.42 (-0.21 to 0.80)	High and moderate agreement in overall child-proxy scores for HUI2 and HUI3 respectively
Barr et al 1999	Cancer survivors: brain tumour (PNET, medulloblastoma, astrocytoma, glioma, ependymoma)	Health professionals (Nurses)	HUI2, HUI3	Absolute mean overall scores reported for child vs nurse dyad only, HUI2 non-irradiated children=0.91 (0.11) & irradiated children=0.97 (0.04) & HUI3 non-irradiated children=0.85 (0.18) & irradiated children=0.91 (0.12) HUI2 nurse= 0.85 HUI3 nurse=0.69	Not used	-	Reported for HUI2 only Child Vs nurse, ICC=0.85; Chid Vs physician, ICC=0.95 HUI2 domains (Kappa statistics) reported Child Vs Nurse Sensation: K=0.05 Emotion: K=0.13 Cognition: K=0.57 Pain: K=0.71 Child Vs physicians Sensation: K=0.42 Emotion: K=0.13 Cognition: K=0.37 Pain: K=0.73	Nurses' underestimated overall utilities both on HUI2 and HUI3. The mean differences between child vs nurse were larger for HUI3 (0.22 to 0.28) than for HUI2 (0.06 to 0.12). High agreement in HUI2 overall scores for both child-proxy pairs. Emotion and pain domains had the lowest and highest agreement respectively. Sensation domain had the lowest agreement between

								child-nurse pair
Bull et al 2014	Cancer survivors: medulloblastoma (WNT molecular group, SHH group, non-WNT/SHH group)	Parents	HUI3	Median overall utility scores for four molecular sub-groups WNT group (n=14) Child=0.69 (IQR, 047 to 0.97) Parent= 0.68 (IQR, 0.32 to 0.97) SHH group (n=8) Child=0.91 (IQR, 0.65 to 0.99) Parent= 0.91 (IQR, 0.68 to 0.95) Other-non WNT/SHH group (n=31) Child=0.90 (IQR, 0.71 to 0.93) Parent= 0.77 (0.45 to 0.97)	Not used	-	-	Parents of children with mild to moderate health condition rated utility similar to their children. However, parents of children with severe health conditions significantly underestimated the utility scores than their children.
Cardelli et al 2006	Cancer survivors: brain tumour; leukaemia; lymphoma; solid tumour	Parents	HUI2	Mean overall utility scores Overall population Child =0.92 Parent=0.90 Primary brain tumour Child=0.87 Parent=0.84 Acute leukemia/Lymphoma Child=0.96 Parent=0.92 Extra Cerebral solid tumour Child =0.94 Parent=0.91	-	-	-	For all cancer survivors, parents tend to underestimate the utility scores albeit the mean differences not significantly high (0.02 to 0.04)
Cheng et al 2011	Cancer patients: leukaemia; lymphoma; solid tumour; brain tumour and others	155 Parents 43 children	HUI2, HUI3	Median overall utility scores HUI2 Child= 0.8 (IQR, 07 to 0.9) Parent=0.8 (IQR, 0.6 to 0.9) p>0.005 HUI3	VAS,WTP, TTO	Median (IQR)scores VAS Overall population Child =7.2 (5.7 to 8.5) Parent= 7.1 (4.3 to		No notable and significant difference in median utility and VAS scores. Parents

	Child=0.8 (IQR, 0.6 to 0.9) Parent=0.9 (IQR, 0.6 to 0.9) p>0..05	8.2) On hypothetical four different management strategy for febrile neutropenia Inpatient Child=5.7 (4.1, 7.9) Parent= 5.6 (2.8,8.2) Early discharge Child=6.1 (4.6,7.2) Parent= 5.9 (4.4,7.2) Outpatient intravenous Child= 5.6 (4.6,7.0) Parent=5.9(4.4,7.3) Outpatient oral Child= 5.3 (4.3,7.3) Parent= 4.7 (2.3, 7.2) WTP (inpatient reference) Early discharge Child=50.0 (0.0. 150.0) Parent=50.0 (0.0, 200.0) Outpatient intravenous Child=62.5 (0.0,230.0) Parent=20.0 (0.0, 200.0) Outpatient oral Child= 45.0 (0.0,185.0) Parent= 0.0 (0.0,	underestimated health status on WTP and TTO than their children

						200.0) TTO (inpatient reference) Early discharge Child=7.3(0.0, 156.0) Parent=0.0 (0.0, 0.0) Outpatient intravenous Child=10.9 (0.0,156.0) Parent=0.0 (0.0, 0.0) Outpatient oral Child= 7.3 (0.0,104.0) (0.0,185.0) Parent= 0.0 (0.0, 0.0)		
Fluchel et al 2008	Cancer survivors: ALL; brain tumour; Wilm's tumour; retinoblastoma; Hodgkin's disease; NHL; AML; STS; neuroblastoma; Ewing's sarcoma; ovarian sarcoma; osteogenic sarcoma;	92-95 Parents 92-95 Children	HUI3	Mean differences (child – proxy reports) Overall =-0.17 Vision =0.001 Hearing =0.003 Speech =-0.026 Ambulation =0.002 Dexterity =0 Emotion=-0.148 Cognition =-0.104 Pain =0.007	-	-	ICC (95% CI) Overall =0.773 (0.706,0.826) Vision =0.67 (0.584, 0.742) Hearing =0.104 (-0.058,0.262) Speech =0.121 (-0.026 to 0.263) Ambulation =0.422 (0.298, 0.532) Dexterity =0.14 (-0.002, 0.277) Emotion=0.433 (0.297,0.552) Cognition =0.752 (0.683,0.807) Pain =0.752 (0.683 to 0.807)	Children rated lower in overall utility and on psychological domains but higher in physical domains than their parents. Agreement of child-proxy ratings was moderate to excellent on overall utility, pain, vision and emotion domains but domains but poor in other domains.

Fluchel et al 2008	Cancer survivors: ALL; brain tumour; Wilm’s tumour; retinoblastoma; Hodgkin’s disease; NHL; AML; STS; neuroblastoma; Ewing’s sarcoma; ovarian sarcoma; osteogenic sarcoma;	94 Physican 92-95 Children	HUI3	Mean differences (child – proxy reports) Overall =-0.216 Vision =0.002 Hearing =0.011 Speech =-0.024 Ambulation =-0.005 Dexterity =-0.002 Emotion=-0.171 Cognition =-0.106 Pain =-0.028	-	-	ICC (95% CI) Overall =0 Vision =0.102 (-0.143,0.142) Hearing =0.754 (0.686,0.81) Speech =0.08 (-0.063,0.219) Ambulation =0.341 (0.2,0.469) Dexterity =0.841(0.794,0.878) Emotion=0.841 (0.794, 0.878) Cognition =0.616 (0.518,0.697) Pain =0.619 (0.523,0.699)	Children generally rated lower than their physicians except for physical domains. There was moderate agreement on psychological domains but poor or on agreement on other overall scores and other domains.
Fu et al 2006	Cancer survivors: ALL; leukaemia; lymphoma; renal tumour; germ cell tumour; retinoblastoma; bone tumour; CNS tumour; STS; carcinoma	140-261 Parents 165-175 children	HUI2, HUI3	Mean (SD) and median overall scores HUI2 Children; mean = 0.87 (0.17), median= 0.93 (0.23 to 1.00) Parents, mean= 0.90 (0.12), median= 0.95 (0.46 to 1.00) HUI3 Children; mean = 0.73 (0.27), median= 0.79 (-0.01 to 1.00) Parents, mean= 0.84 (0.20), median= 0.93 (0.10 to 1.00)	-	-	ICC (95% CI) HUI2, Child vs parents Overall=0.389 (0.227,0.531) Sensation=0.773 (0.706,0.826) Mobility=0.670 (0.584, 0.742) Emotion= 0.104 (-0.058, 0.262) Cognition= 0.121 (-0.026, 0.263) Self-care=0.422 (0.298,0.532) Pain= 0.140 (-0.002, 0.377) HUI3 Overall=0.433(0.2	Parents tend to overestimate overall utility scores than their children both on HUI2 and HUI3. Poor agreement was observed for psychological domains and moderate to excellent on overall and physical domains.

						97, 0.552) Vision=0.754 (0.683,0.807) Hearing= 0.752 (0.683,0.807) Speech=0.843 (0.795, 0.880) Ambulation= 0.699 (0.619, 0.765) Dexterity= -0.006 (-0.0150, 0.138) Emotion= 0.038 (-0.106, 0.181) Cognition= 0.248 (0.110, 0.378) Pain= 0.188 (0.046, 0.323)	
Fu et al 2006	Cancer survivors: ALL; leukaemia; lymphoma; renal tumour; germ cell tumour; retinoblastoma; bone tumour; CNS tumour; STS; carcinoma	200-201 Physicians 165-175 children	HUI2; HUI3	HUI2 Children; mean = 0.87 (0.17), median= 0.93 (0.23 to 1.00) Physicians, mean=0.94 (0.11), median=1.00 (0.37 to 1.00) HUI3 Children; mean = 0.73 (0.27), median= 0.79 (-0.01 to 1.00) Physicians, mean=0.89 (0.11), median=1.00 (0.37 to 1.00)		ICC (95% CI) HUI2, Child vs parents Overall=0.379 (0.237,0.506) Sensation=0.829 (0.778,0.870) Mobility=0.569 (0.465,0.657) Emotion= 0.000 (-0.143,0.143) Cognition= 0.102 (-0.045,0.245) Self-care=0.754 (0.686,0.810) Pain= 0.080 (-0.063,0.219) HUI3 Overall=0.341	Treating physicians tend to overestimate overall utility scores than children both on HUI2 and HUI3. Poor agreement was observed for psychological domains and moderate to excellent on overall and physical domains.

							(0.200,0.469) Vision=0.882 (0.846,0.910) Hearing= 0.841(0.794,0.878) Speech=0.616 (0.518,0.697) Ambulation= 0.619 (0.523, 0.699) Dexterity= 0.337 (0.205, 0.457) Emotion= 0.026 (-0.177, 0.168) Cognition= 0.208 (0.068,0.340) Pain= 0.307 (0.171, 0.431)	
Glaser et al 1999	Cancer survivors: brain tumour (PNET, astrocytoma, medulloblastoma, ependymoma	30 Parents 28 children	HUI2, HUI3	HUI2 Children, mean (SD)=0.78 (0.18), median (min, max)=0.80 (0.36, 1.00) Parents, mean (SD)=0.82 (0.18), median (min, max)= 0.86 (0.25, 1.00) HUI3 Children; mean (SD)=0.66 (0.28), median (min, max)= 0.80 (0.19, 1.00) Parents, mean (SD)=0.72 (0.29), median (min, max)=0.80 (0.24, 1.00)	-	-	HUI2 (Kappa) Sensation =0.54 Mobility=0.72 Emotion=0.37 Self-care=0.47 Pain=0.62	Both children and parents rated lower utility using HUI3 than HUI2. However, for both the MAUIs children rated lower utility than their parents with the mean difference higher on HUI3. Scores on psychological domains showed poor agreement than physical domains.
Glaser et al 1999	Cancer survivors: brain tumour (PNET,	24 Physicians	HUI2, HUI3	HUI2 Children, mean (SD)=0.78 (0.18), median (min,	-	-	HUI2 (Kappa) Sensation =0.38	Both children and physicians rated lower utility scores

	astrocytoma, medulloblastoma, ependymoma	28 children		max)=0.80 (0.36, 1.00) Physicians, mean (SD)=0.89 (0.11), median (min, max)=0.90 (0.65,1.00) <u>HUI3</u> Children; mean (SD)=0.66 (0.28), median (min, max)= 0.80 (0.19, 1.00) Physicians, mean (SD)=0.83 (0.17), median (min, max)=0.86 (0.57, 1.00)			Mobility=0.77 Self-care=0.78	using HUI3 than HUI2. However, for both the MAUIs children rated lower utility than their parents with the mean difference higher on HUI3. Physical domains showed excellent agreement between child and proxy pairs.
Glaser et al 1999	Cancer survivors: brain tumour (PNET, astrocytoma, medulloblastoma, ependymoma	29 Physiotherapist 28 children	HUI2, HUI3	<u>HUI2</u> Children, mean (SD)=0.78 (0.18), median (min, max)=0.80 (0.36, 1.00) Physiotherapist, mean (SD)=0.85 (0.13), median (min, max)=0.88 (0.56,1.00) <u>HUI3</u> Children; mean (SD)=0.66 (0.28), median (min, max)= 0.80 (0.19, 1.00) Physiotherapist, mean (SD)=0.76 (0.24), median (min, max)=0.80 (0.01, 1.00)	-	-	HUI2 (Kappa) Sensation=0.32 Emotion=0.37 Cognition=0.70 Self-care=0.43	Both children and physiotherapists rated lower utility scores using HUI3 than HUI2. However, for both the MAUIs children rated lower utility than their parents with the mean difference higher on HUI3. Except for cognition domain emotion and sensation domain had poor agreement between child-proxy reports.
Glaser et al 1999b	Cancer survivors: brain tumour (PNET, astrocytoma, medulloblastoma,	30 Parents 25 physioth	HUI2	Not reported	-	-	**<u>HUI2</u>** children Vs parents; ICC=0.57 , Pearson r=0.59,	Children had moderate correlation and agreement with their parents' utility ratings. Physicians

	ependymoma)	erapist 23 physicians 28 Children					p<0.01 Children vs physicians; ICC=0.15; Pearson r=0.37, p=0.12 Children vs physiotherapists, ICC=0.40, Pearson r=0.54, p<0.001	showed poor correlation and agreement with children utility scores.
Penn et al 2011; UK	Cancer patients: brain tumour (ependymoma, germ cell tumour, astrocytoma, medulloblastoma, meningioma, pineoblastoma, craniopharyngioma); Comorbidity (hydrocephalus);	29 parents 21 children	HUI3	Overall utility, domain scores and mean differences at 12 months follow-up provided for the child-parent proxy pairs Children; mean (SD, range)= 0.84 (0.21, 0.06 to 1.00) Parents; mean (SD, range)= 0.74 (0.29, -0.15 to 1.00) <u>Domains</u> <u>Children; mean (SD)</u> Vision =0.98 (0.09) Hearing=0.98 (0.11) Speech=0.98 (0.05) Ambulation=0.94 (0.19) Dexterity= 0.95 (0.12) Emotion= 0.97 (0.04) Cognition= 0.92 (0.11) Pain= 0.94 (0.22) <u>Parents; mean (SD)</u> Vision =0.97 (0.09) Hearing=0.96 (0.13) Speech=0.98 (0.07) Ambulation=0.91 (0.26)	-	-	Spearman correlation coefficient(r) Overall= 0.76 Vision=0.73 Hearing=1.00 Speech=1.00 Ambulation=0.82 Dexterity=0.79 Emotion=0.30 Cognition=0.75 Pain=0.20	Parents and child reports provided at 12 months after diagnosis Children rated significantly higher for overall and most of the attributes. Correlation between parents proxy-report and children self-report for overall and the attribute vision, hearing, speech, ambulation, dexterity and cognition was good to excellent but moderate to poor for emotion and pain respectively.

Dexterity= 0.95 (0.14) Emotion= 0.97 (0.07) Cognition= 0.86 (0.19) Pain= 0.90 (0.21) Mean differences (95% CI, P) overall=0.07 (0.00 to 0.14, <0.001) Vision= 0.02 (-0.01 to 0.04, <0.001) Hearing= 0.00 Speech=0.01 (-0.01to 0.02) Ambulation= 0.02 (-0.05 to 0.09, <0.001) Dexterity= 0.01(-0.04 to 0.06, <0.001) Emotion= 0.02 (-0.02 to 0.06, 0.192) Cognition= 0.04 (0.00 to 0.09, <0.001) Pain= 0.03 (-0.14 to 0.08, 0.393)

IQR= Intra quartile range; VAS= Visual Analogue Scale; TTO= Time trade-off, WTP= Willingness-to-pay, HUI= Health utility Index

Interpretation of the ICC<0.4 poor agreement, fair to good agreement 0.40 ≤ ICC < 0.75, excellent ICC≥0.75; Kappa<0.40 poor agreement, fair to moderate 0.40 ≤ ICC < 0.70, excellent ICC≥0.70)

Table 6: Details of the studies that reported dyad self and proxy reports in well (general population) children and adolescents : overall utility/attribute scores, agreement and correlations between child and proxy ratings

Author year	Health conditions	Proxy type/ No of child-Proxy pairs	MAUIs	Child and proxy dyad overall utility and/or domain scores: mean/median (SD/SE/(5% CI/range), mean differences where available	Direct valuation methods (VAS, TTO, SG, WP) where available	Child and proxy direct valuation methods based scores: mean/median (SD/SE/CI) where available	Child and proxy reports: Correlation/agreement (ICC, K) where available	Summary of the findings
Fluchel et al 2008	General health used as controls in the study with cancer survivors	91 physcians/teachers 91 Children	HUI3	Mean differences of overall and domains (child – proxy reports) Overall =-0.167 Vision =-0.003 Hearing =0 Speech =-0.008 Ambulation = 0 Dexterity =0 Emotion =-0.139 Cognition = -0.068 Pain = -0.009	-	-	ICC (95% CI) Overall =0.843 (0.795, 0.88) Vision =0.699 (0.619,0.765) Hearing =-0.006 (-0.15,0.138) Speech =0.038 (-- -0.106, 0.181) Ambulation =0.248 (0.11, 0.378) Dexterity =0.188 0.046,0.657) Emotion=0.433 (0.297,0.552) Cognition =0.829 (0.778,0.87) Pain =0.569 (0.465,0.657)	Proxies of children with no obvious health issues tend to slightly overestimate or rate similarly as their children. Most of the domains and overall scores showed moderate to high agreement except for hearing, ambulation and dexterity domains.
Penn et al 2011; UK	General health used as controls in the study with cancer survivors	32 parents 22 children	HUI3	Overall utility scores at 12 months follow-up provided for the child-proxy pair Children; mean (SD, range)= 0.94 (0.13, 0.56 to 1.00) Parents; mean (SD, range)= 0.96 (0.09, 0.63 to 1.00) **<u>Domains</u>** **<u>Children, mean (SD)</u>** Vision=0.99 (0.02)	-	-	Spearman correlation coefficient(r) Overall= 0.31	Moderate correlation between parent proxy-report and child-reports. Children rated slightly higher or similar to their parents on overall utility and attributed scores. Ceiling

				Hearing= 1.00 (0.00) Speech=1.00 (0.00) Ambulation= 1.00 (0.00) Dexterity= 0.98 (0.12) Emotion= 0.98 (0.06) Cognition=0.97 (0.09) Pain=0.99 (0.03) **Parents, mean (SD)** Vision=1.00 (0.00) Hearing= 1.00 (0.00) Speech=1.00 (0.00) Ambulation=1.00 (0.00) Dexterity= 0.98 (0.10) Emotion= 0.99 (0.05) Cognition=0.98 (0.06) Pain=0.98 (0.06)				effect was observed as most of the controls (with no establish health conditions) and their parents had a maximum possible score of 1.00.
Baumann et al 2016	Normal birthweight/ full term	Parents	HUI3	Absolute overall utility scores reported Controls (normal birth weight/full term) child=0.89 (CI0.87 to 0.92); Parent= 0.95 (CI=0.95 to 0.96)	Not used	-	-	Parents slightly overestimated (mean difference= -0.06) for normal birthweight or full-term children
Belfort et al 2011	Health weight children (≥8 years)	Parents	HUI3	Children with healthy weight Overall=0.82 (95% CI= 0.77, 0.88) Ambulation= 1 (95% CI=1,1) Emotion= 0.96 (95% CI=0.93, 0.98) Cognition=0.87 (0.81, 0.92) Pain=0.95 (95% CI=0.90, 0.97) Parents (>8 years children) Overall=0.84 (95% CI=	Not used	-	-	Children rated similar or slightly lower than their parents on overall utility and domains. Parents underestimated their children utility on pain domain by 0.04.

				0.74, 0.92) Ambulation= 1 (95% CI=1,1) Emotion= 0.96 (95% CI=0.93,0.99) Cognition=0.88 (0.80, 0.95) Pain=0.91 (95% CI= 0.82,0.97)				
Belfort et al 2011	Healthy weight children (≥12 years)	Parents	HUI3	Healthy weight children (>12 years) Overall=0.85 (95% CI= 0.80, 0.90) Ambulation= 1 (95% CI=1,1) Emotion= 0.98 (95% CI=0.96,1.0) Cognition=0.87 (0.81, 0.92) Pain=0.96 (95% CI=0.93, 0.99) Parents of healthy weight children (>12 years children) Overall=0.86 (95% CI= 0.73, 0.95) Ambulation= 1 (95% CI=1,1) Emotion= 0.97 (95% CI=0.93,0.99) Cognition=0.91 (95% CI=0.81,0.98) Pain=0.90 (95% CI= 0.76,0.95)	Not used	-	-	Children rated similar or slightly lower than their parents on overall utility and domains. Parents underestimated on pain domain by 0.06.
Gusi et al 2014	General population	442 Mothers and children	EQ-5D-Y and EQ-	Not reported	-	-	Kappa Children vs mothers	Mother had poor agreement on emotional domain but moderate on

			5D-Y Proxy				Pain & discomfort=0.68 (p<0.05) Feeling worried, sad or unhappy=0.221 (p<0.05)	pain domain with their children.
Jelsma and Ramma 2010	Children in open school (General school population with no obvious chronic illness)	530 mothers 158 children	EQ-5D-Y	Not reported	EQ-5D-Y VAS	Children from open school; mean (SD)= 87.9 (16.5) Mothers; mean (SD)=88.4 (15.7)	Open school VAS scores Pearson's r=0.33 (p<0.01) Kappa Mobility=0.15 Looking after myself=0.08 Doing usual activities=0.01 Having pain or discomfort=0.20 Feeling worried, sad an unhappy=0.21	Mothers rated similar health status as their children using VAS scale. Poor agreement on all the domains between child and proxy respondents.
Verrips et al 2001	Very low birthweight; Very preterm birth	203 (mail), 100 (telephone) and 103 (face-to-face) parents and children pairs	HUI3	Overall utility scores by mode of administration **Mail** Children; mean (95%CI)=0.90 (0.88to 0.92) Parents; mean (95%CI)=0.93 (0.91 to 0.94) **Telephone** Children; mean (95%CI)=0.87 (0.84 to 0.90) Parents; mean (95%CI)=0.85 (0.81 to 0.88)	-	-	Kappa agreement for domains Vision, k=0.62 Speech, k=0.24 Ambulation, k=0.50 Dexterity, k=0.80 Cognition, k=0.38 Emotion, k=0.35 Pain, k=0.23	Children rated higher than parents when administered via telephone or face-to-face than their parents. For postal mail survey, children rated lower utility than their parents. However, the mean differences were not statistically

				Face-to-face Children; mean (95% CI)= 0.84 (0.81-0.88) Parents; mean (95% CI)= 0.83 (0.79 to 0.87)				significant. Agreement was high in the physical attributed but poor in psychological attributes.
Wolke et al, 2013	General health (control) Full term	282 parents and adolescents	HUI3	Overall mean utilities Adolescent; mean (SD)=0.87 (0.15) Parents; mean (SD)=0.91 (0.11)	-	-	-	Adolescent rated utilities significantly lower than their parents

IQR= Intra quartile range; VAS= Visual Analogue Scale; TTO= Time trade-off, WTP= Willingness-to-pay, HUI= Health utility Index

Interpretation of the ICC<0.4 poor agreement, fair to good agreement 0.40 ≤ ICC < 0.75, excellent ICC≥0.75; Kappa<0.40 poor agreement, fair to moderate 0.40 ≤ ICC < 0.70, excellent ICC≥0.70)

Table 1: Summary of the included studies

Description 1	**Description 2**	**Number of studies**
Total studies included		35
Studies with MAUI/s		32
Studies with HUI3	Total	22
	HUI3 utility scores reported	19
	No scores reported (only agreement or correlation between self-proxy pair)	3
	HUI3 on cancers	10
	HUI3 on other health condition	12
Studies with HUI2	Total	11
	HUI2 utility scores reported	9
	No scores reported (only agreement between self-proxy pair)	2
	HUI2 on cancers	8
	HUI2 on other health conditions	3
	Reported with HUI3	10
Studies with EQ-5D (studies with EQ-5D-Y=3 ; EQ-5D-3L=3; EQ-5D-5L=3)	Total	8
	EQ-5D utility scores reported	7
	No scores reported (only agreement between self-proxy pair)	1
	EQ-5D on cancers	0
	EQ-5D on other health conditions	8
Studies with Other instruments (none of the instrument was used on cancers)	Total	3
	QWB	1
	PAHOM	1
	4 Hypothetical Health States Vignette	1
VAS (EQ-5D VAS= 6 studies)	Total	10
	VAS scores reported	10
	VAS on cancers	2
	VAS on other health conditions	8
TTO	Total	4
	TTO on cancers	2
	TTO on health conditions	2
SG	Total	3
	SG on Cancers	1
	SG on other health conditions	2

MAUI=Multi-attribute utility Index, HUI= Health Utility Index, EQ-5D= EuroQoL-5D, QWB= Quality of Well-Being, PAHOM= Paediatric Asthma Health Outcome Measure, VAS=Visual Analogue Scale, WTP=Willingness to Pay, TTO=Time-trade-off, SG=Standard gamble

Table 2: Details of the included studies on cancer and/or cancer survivors that reported dyad self and proxy reports using preference-based quality of life instruments

Author, year and country of study	Health Condition	Sample sizes: child (male%) & proxy:	Age of childhood population :mean/median; range (age range of those self-reported)	Instrument / valuation methods used	Response rate (%)	Source of tariff	Proxy type & mode of administration	MAUI utility scores reported	Results reported	Reporting quality score
Banks et al 2008; Canada	Cancer patients: leukaemia, lymphoma, brain tumour	Children=11 (65%); parents=22	Mean: 9.5; 2-18 (10-18 years)	HUI2, HUI3	72.0	Unspecified	Parents. Self-administered paper questionnaire	No	Agreement between self-proxy pair (ICC)	12
Barr et al 1999 Canada	Cancer survivors: brain tumour (PNET, medulloblastoma, astrocytoma, glioma, ependymoma)	Children=15 (46.3%); nurses=15; physicians=12;	Mean: 9.5; 1.7-17.9 (9.5 to 17.9)	HUI2, HUI3	93.2	Canadian general population of adults	Nurses & physicians. Self-administered paper questionnaire	Yes	Mean overall utility scores only & agreement between self-proxy pair for overall (ICC) and dimension scores (Kappa)	12
Bull et al 2014; UK	Cancer survivors: medulloblastoma (WNT molecular group, SHH group, non-WNT/SHH group)	Children=8-31; parents=8-33	Median: 15.7, 14.5, 14.7; 10.3-18.2 (10 to 18 years)	HUI3	49.1 to 50.9	Unspecified	Parents. Unspecified mode of administration.	Yes	Median overall utility scores	12
Cardarelli et al 2006; Italy	Cancer survivors: brain tumour; leukaemia; lymphoma; solid tumour	Children=2-15; parents=30-121	Mean: 13.3; 8-28 (8-28 years)	HUI2	93-100	Canadian general population of adults	Parents. Self-administered paper questionnaire.	Yes	Mean and median overall utility scores	13
Cheng et al (2011) Canada	Cancer patients: leukaemia; lymphoma; solid tumour; brain tumour and others	Children=43 (23.2 to 53.5%); parents=155	Median:15.4; 0-18 (12-18 years)	HUI2, HUI3; VAS,TTO	100	Unspecified	Parents. Face-to-face interview.	Yes	Median overall utility, VAS and TTO scores	12
Fluchel et al, 2008; Uruguay	Cancer survivors and general population as controls: ALL; brain tumour; Wilm's tumour; retinoblastoma; Hodgkin's disease; NHL; AML; STS; neuroblastoma; Ewing's sarcoma; ovarian sarcoma; osteogenic sarcoma; General health (control)	Children=81 Parents=94	Mean: 12.2 to 13.6; 7-27	HUI3	81.4	Canadian general population of adults	Parents, self-assessed by children. Face-to-face interview.	Yes	Mean difference between self and proxy overall utility and attribute scores and agreement (ICC) for self-proxy pair	13

Fu et al 2006; El Salvador, Honduras, Nicaragua, Panama	Cancer survivors: ALL; leukaemia; lymphoma; Renal tumour; germ cell Cancer patients; retinoblastoma; bone tumour; CNS tumour; STS; carcinoma	Children=165-175 (56%); parents=140-180; physicians=200-201	Median: 12.8; 3.4-25.8	HUI2, HUI3	78.2 to 100	Canadian general population of adults	Parents or physicians. Parents, Face-to-face interview or Self-administered paper questionnaire.	Yes	Mean difference between self and proxy overall utility and attribute scores and agreement (ICC or Kappa) for self-proxy pair overall and attribute scores	12
Glaser et al 1999; UK	Cancer survivors: brain tumour (PNET, astrocytoma, medulloblastoma, ependymoma)	Children=25-26; parents= 23-28	Mean: 10.5; 6-16	HUI2, HUI3	90	Canadian general population of adults	Parents, physiotherapist or physicians. Self-administered paper questionnaire.	Yes	Mean overall utility scores and agreement (Kappa) between self-proxy pair for overall and attribute scores	13
Glaser et al 1999b; UK (Canada)	Cancer survivors: brain tumour (PNET, astrocytoma, medulloblastoma, ependymoma)	Children=19-24; parents=19-24	Mean: 10.5; 6-16	HUI2, HUI3	80	Canadian general population of adults	Parents, physiotherapist or physicians. Self-administered paper questionnaire.	No	Agreement (ICC and Kappa) between self-proxy pair for overall and attribute scores	13
Penn et al 2011; UK	Cancer patients: brain tumour (ependymoma, germ cell tumour, astrocytoma, medulloblastoma, meningioma, pineoblastoma, craniopharyngioma); Comorbidity (hydrocephalus); General health (control)	Children=21-22 (48.3 – 53.1%); parents=29-32	Median: 11.1, 10.7; 3.6-18.9	HUI3	60.4	Unspecified	Parents. Face-to-face interview.	Yes	Mean overall and dimension scores; agreement (ICC) between overall and dimension scores	13

MAUI= Multi-attribute utility instruments: HUI= Health Utility Index; ICC= Intra-class correlation; VAS= Visual Analog Scale, WTP=Willingness to pay, TTO=Time Trade Off; QWB= Quality of Well-Being

Table 3: Details of the included studies on health conditions other than cancers that reported dyad self and proxy reports using preference-based quality of life instruments

Author, year and country of study	Health Condition	Sample sizes: child (male%) & proxy: Childhood	Age of childhood population :mean/median; range (age range of those self-reported)	Instrument / valuation methods used	Response rate (%)	Source of tariff	Proxy type & mode of administration	MAUI utility scores reported	Results reported	Reporting quality score
Baumann et al 2016; Germany	Very low birthweight/very preterm birth; General health (normal birthweight/term)	Very pre-term children=190 (45.3%); parents=197 Control children=201 (53.7) & parents=204	Mean: 13; 13-26 (13-26 years)	HUI3	65.0	Canadian general population of adults	Parents. Postal survey	Yes	Mean overall utility scores only	13
Verrips et al 2001; Netherlands	Very low birthweight; Very preterm birth	Children=100 & 203 (46.2%); parents=100 &203	Mean: 14.2	HUI3	67.7	Canadian general population of adults	Parents. Postal survey or face-to-face or telephone interview.	Yes	Mean overall utility scores	13
Wolke et al, 2013; Germany	Very low birthweight; Very preterm birth (VLBW/VP), General health (Full-term, FT, controls)	Children= 260 (VLBW/VP) 282 (FT) (55.2%)	Mean: 13	HUI3	61	Canadian general population of adults	Parents. Postal survey.	Yes	Mean overall utility scores	13
Saigal et al 1999; Canada	Extremely low birthweight or extremely preterm birth children	Children=263	Mean: 8, 14.2	4 hypothetical health states	84.4	Canadian general population of adults	Parents, health professionals (physicians, nurses) Face-to-face interview.	Yes	Mean of hypothetical dimensions	11
Hanberger et al 2009; Sweden	Diabetes type I; Severe hypoglycaemia	Children=148; parents=151	Mean: 13.2	EQ-5D-3L	72.5	Unspecified	Parents. Postal survey.	Yes	Mean overall utility scores	12
Rhodes et al, 2012; USA	Overweight (insulin resistant, prediabetes); Diabetes type II	Children=108 (32.9%); parents=106	Mean: 12.8; 12-18	HUI3	75	Canadian general population of adults	Parents. Face-to-face interview.	Yes	Median overall utility scores and correlation (spearman correlation) between child-proxy pair	13
Lee et al 2011;, USA	Diabetes type I; Complications (blindness, end-stage renal disease, angina,	Children=231; parents=223	Mean: 13.7; 8-18	HUI3; TTO	82	Canadian general population of adults	Parents. Face-to-face interview.	Yes	Mean overall utility and TTO scores; correlation between self-proxy pair for utility and TTO	12

	stroke, amputation)									
Belfort et al 2011 USA	Overweight or obese; General population (healthy weight)	Children=42 overweight (57%) & =34 (47%) (healthy weights); Parents= n (not reported)	Mean: 10.8 (overweight), 11.5 (normal weight); 5-18 (8-18 years)	HUI3	46.2 to 65.3	Canadian general population of adults	Parents. Face-to-face interview	Yes	Mean & median overall utility and 3 dimension scores; correlations between child and proxy scores (Spearman's rho) for the overall and 3 dimensions (pain, emotion & cognition) for self-proxy pair	12
Robertson et al 2017; UK	Obesity	Children=63 (55.4%); parents=63	Mean: 9.46; 6-11	EQ-5D-Y, EQ-5D-Y VAS	100	Unspecified	Parents. Self-administered paper questionnaire	Yes	Mean and mean differences of overall utility and VAS scores	12
Brunner et al 2003; Canada	MSKD: juvenile rheumatoid arthritis; systemic lupus erythematosus; juvenile dermatomyositis; arthropathy secondary to haemophilia	Children=45-55 (23.5%); parents=68-80	1-18 (HUI3 and VAS by 8-18 years & SG by 10-18 years)	HUI3, SG, VAS (Linear analogue scale, Categorical scale),	81.8	Canadian general population of adults	Parents. Face-to-face interview.	Yes	Mean and median overall utility, SG, VAS scores; correlation (Spearman's rho) and agreement (ICC) for self-proxy pair	12
Brunner et al, 2004; USA	MSKD: juvenile rheumatoid arthritis; Castleman syndrome; systemic lupus erythematosus; juvenile dermatomyositis; arthritis associated with IBD; juvenile psoriatic arthritis; juvenile ankylosing spondylitis; sarcoidosis	Children=87 (23.5%); parents=119	3-18 (VAS by 8-18 years & SG by 12-18 years)	VAS (well and pain), SG	46.4	Not applicable	Parents. Face-to-face interview.	No	Mean and median SG and VAS and agreement (ICC) for self-proxy pair	12
Sims-Williams et al 2017; Uganda	Spina bifida; Comorbidities (hydrocephalus, immobility)	Children=62 (56%); caregivers=66	10-14	HUI3, VAS	97.1	Unspecified	Parents. Face-to-face interview; Self-administered paper questionnaire	Yes	Mean overall utility and VAS scores and Pearson correlations between self-proxy pair on utility and VAS scores	12
Chevreul et al 2015b; France (Denmark, England,	Fragile X syndrome	Children=53 (88.7%); caregivers=42	Mean: 10.3 (not reported)	EQ-5D-5L; EQ-5D-5L VAS	36.6	General population in adults in Denmark, England,	Caregivers. Online survey.	Yes	Mean utility and VAS scores	13

Italy, Netherlands, Scotland, Poland)						Italy, Netherlands, Scotland and Poland				
Czyzewski et al 1994, USA (USA)	Cystic fibrosis	Children=55 (48.2%); parents=199	Mean: 8.74; 0.2-17.9 (12 to 18 years)	QWB	70	US population of adults	Parents. Face-to-face interview.	Yes	Mean overall utility scores and agreement (ICC) for self-proxy pair.	13
Iskrov et al 2015; Bulgaria	Cystic fibrosis	Children=32; parents=17	Mean: 16	EQ-5D-3L VAS	69.7	Unspecified	Parents, online survey.	Yes	Median overall utility scores	11
Chadha et al 2010; Canada	Recurrent respiratory papillomatosis	Children=20 (65%); parents=20	Median: 9; 1.42-17 (9 to 17 years)	HUI3; VAS	100	Canadian general population of adults	Caregivers. Face-to-face interview.	Yes	Mean overall utility and VAS scores	13
Unger et al (2012); Canada	Asthma	Children=66 & 91; parents=66 & 91	8-17	HUI2, HUI3	64	Canadian general population of adults	Parent. Face-to-face interviewer administer	No	Agreement between self-proxy pair for overall and 3 dimensions (mobility, ambulation and emotion)	
Gerald et al 2012; USA	Asthma	Children= 64-261 (63.6 to 100); parents=64-261	6-12 (6-12 years)	PAHOM	86.7-93.3	US children with asthma (aged 6-12 years)	Parents. Face-to-face interview.	Yes	Mean overall utility scores and correlation between self-proxy pair	12
Kulpeng et al 2013; Thailand	Severe bacteraemia; Acute otitis media; Severe infection; Permanent sequelae to severe infection (epilepsy, hearing loss, pneumonia, lung disease, severe meningitis, mental retardation)	Children=74 (53.5%); parents=99	Mean: 10; 5-14	EQ-5D-3L VAS, HUI2, HUI3	100	HUI- Canadian general population of adults; EQ-5D-3L- Thai general population of adults	Caregivers. Face-to-face interview.	Yes	Mean difference of overall utility scores and correlations between self-proxy pair	13
Trent et al 2011; USA	Pelvic inflammatory disease; Permanent sequelae (ectopic pregnancy, infertility, chronic abdominal pain)	Children=134 (0%); parents=121	Mean: 16.2; 12-19	VAS, TTO	100	Not applicable	Parents. Online survey.	No	Median VAS and TTO scores	12
Vermeulen et al 2017; Netherlands	Hypothetical behavioural disorders: Oppositional defiant disorder; Conduct disorder; Disruptive behavioural disorder;	Children=51& 65; parents=51&65	Mean: 9, 15	EQ-5D-3L and VAS	100	UK general population of adults	Professionals; Self-assessed by children; Postal survey; Self-administered paper	Yes	Mean overall utility and VAS scores	13

	Comorbidities (ADHD, substance abuse); General health						questionnaire			
Vermeulen et al 2017b; Netherlands	Antisocial personality disorder: Chronically antisocial and seriously violent	Children=15-50 (62-67%); parents=10-16	Mean: 16; 12-18	EQ-5D-3L	100	UK general population of adults	Parents. Self-administered paper questionnaire	Yes	Mean difference of treatment effect on overall utility and VAS scores	13
Gusi et al 2014; Spain	General population	Children=620; parents=477	Mean: 16; 6-17 years (6-17)	EQ-5D-Y & Proxy	100	School aged children in Germany, Italy, South Africa, Spain and Sweden (aged 8-19 years)	Parents, self-administered paper questionnaire	No	Agreement (Kappa) between self-proxy pair	
Morrow et al 2010; Australia	Chronic illness: any cancer, cystic fibrosis, type1 diabetes, cerebral palsy, chronic neurological conditions, inflammatory bowel disease, liver transplant; chronic kidney diseases, autism	Children=69 (94.0%); parents=59; doctors=59	12-18 (12-18)	HUI2, HUI3 (HUI23SUE.15Q & HUI23PUE.15Q)	94	Canadian general population of adults	Parents, doctors	Yes	Mean overall utility scores, agreement (Kappa) between child-parent & child-doctor pair	
Sung et al 2004; Canada	Cancer patients; Arthritis; Haemophilia; Conditions requiring bone marrow transplant; Hypothetical asthma as example of mild disability; Hypothetical stroke as example of moderate disability; Hypothetical injury as example of severe disability	Children=19 & 21 (55%) Parents= 19 & 21	Mean: 13.7; 12-18	HUI2, HUI3, VAS, SG, TTO	81.5	Canadian general population of adults	Parents. Face-to-face interview.	Yes	Mean overall utility, VAS, SG and TTO scores and correlation (Spearman’s rho) and agreement between self-proxy pair for overall utility, VAS, SG and TTO.	12

MAUI= Multi-attribute utility instruments: HUI= Health Utility Index; ICC= Intra-class correlation; MSKD= Musculo-skeletal diseases; VAS= Visual Analog Scale, WTP=Willingness to pay, TTO=Time Trade Off; QWB= Quality of Well-Being

Table 4: Details of the studies that reported dyad self and proxy reports on cancer participants : overall utility/attribute scores, agreement and correlations between child and proxy ratings

Author year	Health conditions	Proxy type/ No of child-Proxy pairs	MAUIs	Child and proxy dyad overall utility and/or domain scores: mean/median (SD/SE/(5% CI/range), mean differences where available	Direct valuation methods (VAS, TTO, SG, WP) where available	Child and proxy direct valuation methods based scores: mean/median (SD/SE/CI) where available	Child and proxy reports: Correlation/agreement (ICC, K) where available	Summary of the findings
Banks et al 2008	Cancer patients: leukaemia, lymphoma, brain tumour	Parents/	HUI2, HUI3	Not reported	Not used	-	HUI2: overall score ICC= 0.74 (0.29 to 0.92) HUI3: ICC=0.42 (-0.21 to 0.80)	High and moderate agreement in overall child-proxy scores for HUI2 and HUI3 respectively
Barr et al 1999	Cancer survivors: brain tumour (PNET, medulloblastoma, astrocytoma, glioma, ependymoma)	Health professionals (Nurses)	HUI2, HUI3	Absolute mean overall scores reported for child vs nurse dyad only, HUI2 non-irradiated children=0.91 (0.11) & irradiated children=0.97 (0.04) & HUI3 non-irradiated children=0.85 (0.18) & irradiated children=0.91 (0.12) HUI2 nurse= 0.85 HUI3 nurse=0.69	Not used	-	Reported for HUI2 only Child Vs nurse, ICC=0.85; Chid Vs physician, ICC=0.95 HUI2 domains (Kappa statistics) reported Child Vs Nurse Sensation: K=0.05 Emotion: K=0.13 Cognition: K=0.57 Pain: K=0.71 Child Vs physicians Sensation: K=0.42 Emotion: K=0.13 Cognition: K=0.37 Pain: K=0.73	Nurses' underestimated overall utilities both on HUI2 and HUI3. The mean differences between child vs nurse were larger for HUI3 (0.22 to 0.28) than for HUI2 (0.06 to 0.12). High agreement in HUI2 overall scores for both child-proxy pairs. Emotion and pain domains had the lowest and highest agreement respectively. Sensation domain had the lowest agreement between

								child-nurse pair
Bull et al 2014	Cancer survivors: medulloblastoma (WNT molecular group, SHH group, non-WNT/SHH group)	Parents	HUI3	Median overall utility scores for four molecular sub-groups WNT group (n=14) Child=0.69 (IQR, 047 to 0.97) Parent= 0.68 (IQR, 0.32 to 0.97) SHH group (n=8) Child=0.91 (IQR, 0.65 to 0.99) Parent= 0.91 (IQR, 0.68 to 0.95) Other-non WNT/SHH group (n=31) Child=0.90 (IQR, 0.71 to 0.93) Parent= 0.77 (0.45 to 0.97)	Not used	-	-	Parents of children with mild to moderate health condition rated utility similar to their children. However, parents of children with severe health conditions significantly underestimated the utility scores than their children.
Cardelli et al 2006	Cancer survivors: brain tumour; leukaemia; lymphoma; solid tumour	Parents	HUI2	Mean overall utility scores Overall population Child =0.92 Parent=0.90 Primary brain tumour Child=0.87 Parent=0.84 Acute leukemia/Lymphoma Child=0.96 Parent=0.92 Extra Cerebral solid tumour Child =0.94 Parent=0.91	-	-	-	For all cancer survivors, parents tend to underestimate the utility scores albeit the mean differences not significantly high (0.02 to 0.04)
Cheng et al 2011	Cancer patients: leukaemia; lymphoma; solid tumour; brain tumour and others	155 Parents 43 children	HUI2, HUI3	Median overall utility scores HUI2 Child= 0.8 (IQR, 07 to 0.9) Parent=0.8 (IQR, 0.6 to 0.9) p>0.005 HUI3	VAS,WTP, TTO	Median (IQR)scores VAS Overall population Child =7.2 (5.7 to 8.5) Parent= 7.1 (4.3 to		No notable and significant difference in median utility and VAS scores. Parents

	Child=0.8 (IQR, 0.6 to 0.9) Parent=0.9 (IQR, 0.6 to 0.9) p>0..05	8.2) On hypothetical four different management strategy for febrile neutropenia <u>Inpatient</u> Child=5.7 (4.1, 7.9) Parent= 5.6 (2.8,8.2) <u>Early discharge</u> Child=6.1 (4.6,7.2) Parent= 5.9 (4.4,7.2) Outpatient intravenous Child= 5.6 (4.6,7.0) Parent=5.9(4.4,7.3) Outpatient oral Child= 5.3 (4.3,7.3) Parent= 4.7 (2.3, 7.2) <u>WTP (inpatient reference)</u> Early discharge Child=50.0 (0.0. 150.0) Parent=50.0 (0.0, 200.0) Outpatient intravenous Child=62.5 (0.0,230.0) Parent=20.0 (0.0, 200.0) Outpatient oral Child= 45.0 (0.0,185.0) Parent= 0.0 (0.0,	underestimated health status on WTP and TTO than their children

						200.0) <u>TTO (inpatient reference)</u> Early discharge Child=7.3(0.0, 156.0) Parent=0.0 (0.0, 0.0) Outpatient intravenous Child=10.9 (0.0,156.0) Parent=0.0 (0.0, 0.0) Outpatient oral Child= 7.3 (0.0,104.0) (0.0,185.0) Parent= 0.0 (0.0, 0.0)		
Fluchel et al 2008	Cancer survivors: ALL; brain tumour; Wilm's tumour; retinoblastoma; Hodgkin's disease; NHL; AML; STS; neuroblastoma; Ewing's sarcoma; ovarian sarcoma; osteogenic sarcoma;	92-95 Parents 92-95 Children	HUI3	Mean differences (child – proxy reports) Overall =-0.17 Vision =0.001 Hearing =0.003 Speech =-0.026 Ambulation =0.002 Dexterity =0 Emotion=-0.148 Cognition =-0.104 Pain =0.007	-	-	ICC (95% CI) Overall =0.773 (0.706,0.826) Vision =0.67 (0.584, 0.742) Hearing =0.104 (-0.058,0.262) Speech =0.121 (-0.026 to 0.263) Ambulation =0.422 (0.298, 0.532) Dexterity =0.14 (-0.002, 0.277) Emotion=0.433 (0.297,0.552) Cognition =0.752 (0.683,0.807) Pain =0.752 (0.683 to 0.807)	Children rated lower in overall utility and on psychological domains but higher in physical domains than their parents. Agreement of child-proxy ratings was moderate to excellent on overall utility, pain, vision and emotion domains but domains but poor in other domains.

Fluchel et al 2008	Cancer survivors: ALL; brain tumour; Wilm's tumour; retinoblastoma; Hodgkin's disease; NHL; AML; STS; neuroblastoma; Ewing's sarcoma; ovarian sarcoma; osteogenic sarcoma;	94 Physican 92-95 Children	HUI3	Mean differences (child – proxy reports) Overall =-0.216 Vision =0.002 Hearing =0.011 Speech =-0.024 Ambulation =-0.005 Dexterity =-0.002 Emotion=-0.171 Cognition =-0.106 Pain =-0.028	-	-	ICC (95% CI) Overall =0 Vision =0.102 (-0.143,0.142) Hearing =0.754 (0.686,0.81) Speech =0.08 (-0.063,0.219) Ambulation =0.341 (0.2,0.469) Dexterity =0.841(0.794,0.878) Emotion=0.841 (0.794, 0.878) Cognition =0.616 (0.518,0.697) Pain =0.619 (0.523,0.699)	Children generally rated lower than their physicians except for physical domains. There was moderate agreement on psychological domains but poor or on agreement on other overall scores and other domains.
Fu et al 2006	Cancer survivors: ALL; leukaemia; lymphoma; renal tumour; germ cell tumour; retinoblastoma; bone tumour; CNS tumour; STS; carcinoma	140-261 Parents 165-175 children	HUI2, HUI3	Mean (SD) and median overall scores <u>HUI2</u> Children; mean = 0.87 (0.17), median= 0.93 (0.23 to 1.00) Parents, mean= 0.90 (0.12), median= 0.95 (0.46 to 1.00) <u>HUI3</u> Children; mean = 0.73 (0.27), median= 0.79 (-0.01 to 1.00) Parents, mean= 0.84 (0.20), median= 0.93 (0.10 to 1.00)	-	-	ICC (95% CI) <u>HUI2,</u> Child vs parents Overall=0.389 (0.227,0.531) Sensation=0.773 (0.706,0.826) Mobility=0.670 (0.584, 0.742) Emotion= 0.104 (-0.058, 0.262) Cognition= 0.121 (-0.026, 0.263) Self-care=0.422 (0.298,0.532) Pain= 0.140 (-0.002, 0.377) <u>HUI3</u> Overall=0.433(0.2	Parents tend to overestimate overall utility scores than their children both on HUI2 and HUI3. Poor agreement was observed for psychological domains and moderate to excellent on overall and physical domains.

						97, 0.552) Vision=0.754 (0.683,0.807) Hearing= 0.752 (0.683,0.807) Speech=0.843 (0.795, 0.880) Ambulation= 0.699 (0.619, 0.765) Dexterity= -0.006 (-0.0150, 0.138) Emotion= 0.038 (-0.106, 0.181) Cognition= 0.248 (0.110, 0.378) Pain= 0.188 (0.046, 0.323)	
Fu et al 2006	Cancer survivors: ALL; leukaemia; lymphoma; renal tumour; germ cell tumour; retinoblastoma; bone tumour; CNS tumour; STS; carcinoma	200-201 Physicians 165-175 children	HUI2; HUI3	HUI2 Children; mean = 0.87 (0.17), median= 0.93 (0.23 to 1.00) Physicians, mean=0.94 (0.11), median=1.00 (0.37 to 1.00) HUI3 Children; mean = 0.73 (0.27), median= 0.79 (-0.01 to 1.00) Physicians, mean=0.89 (0.11), median=1.00 (0.37 to 1.00)		ICC (95% CI) HUI2, Child vs parents Overall=0.379 (0.237,0.506) Sensation=0.829 (0.778,0.870) Mobility=0.569 (0.465,0.657) Emotion= 0.000 (-0.143,0.143) Cognition= 0.102 (-0.045,0.245) Self-care=0.754 (0.686,0.810) Pain= 0.080 (-0.063,0.219) HUI3 Overall=0.341	Treating physicians tend to overestimate overall utility scores than children both on HUI2 and HUI3. Poor agreement was observed for psychological domains and moderate to excellent on overall and physical domains.

							(0.200,0.469) Vision=0.882 (0.846,0.910) Hearing= 0.841(0.794,0.878) Speech=0.616 (0.518,0.697) Ambulation= 0.619 (0.523, 0.699) Dexterity= 0.337 (0.205, 0.457) Emotion= 0.026 (-0.177, 0.168) Cognition= 0.208 (0.068,0.340) Pain= 0.307 (0.171, 0.431)	
Glaser et al 1999	Cancer survivors: brain tumour (PNET, astrocytoma, medulloblastoma, ependymoma	30 Parents 28 children	HUI2, HUI3	HUI2 Children, mean (SD)=0.78 (0.18), median (min, max)=0.80 (0.36, 1.00) Parents, mean (SD)=0.82 (0.18), median (min, max)= 0.86 (0.25, 1.00) HUI3 Children; mean (SD)=0.66 (0.28), median (min, max)= 0.80 (0.19, 1.00) Parents, mean (SD)=0.72 (0.29), median (min, max)=0.80 (0.24, 1.00)	-	-	HUI2 (Kappa) Sensation =0.54 Mobility=0.72 Emotion=0.37 Self-care=0.47 Pain=0.62	Both children and parents rated lower utility using HUI3 than HUI2. However, for both the MAUIs children rated lower utility than their parents with the mean difference higher on HUI3. Scores on psychological domains showed poor agreement than physical domains.
Glaser et al 1999	Cancer survivors: brain tumour (PNET,	24 Physicians	HUI2, HUI3	HUI2 Children, mean (SD)=0.78 (0.18), median (min,	-	-	HUI2 (Kappa) Sensation =0.38	Both children and physicians rated lower utility scores

	astrocytoma, medulloblastoma, ependymoma	28 children		max)=0.80 (0.36, 1.00) Physicians, mean (SD)=0.89 (0.11), median (min, max)=0.90 (0.65,1.00) HUI3 Children; mean (SD)=0.66 (0.28), median (min, max)= 0.80 (0.19, 1.00) Physicians, mean (SD)=0.83 (0.17), median (min, max)=0.86 (0.57, 1.00)			Mobility=0.77 Self-care=0.78	using HUI3 than HUI2. However, for both the MAUIs children rated lower utility than their parents with the mean difference higher on HUI3. Physical domains showed excellent agreement between child and proxy pairs.
Glaser et al 1999	Cancer survivors: brain tumour (PNET, astrocytoma, medulloblastoma, ependymoma	29 Physiotherapist 28 children	HUI2, HUI3	HUI2 Children, mean (SD)=0.78 (0.18), median (min, max)=0.80 (0.36, 1.00) Physiotherapist, mean (SD)=0.85 (0.13), median (min, max)=0.88 (0.56,1.00) HUI3 Children; mean (SD)=0.66 (0.28), median (min, max)= 0.80 (0.19, 1.00) Physiotherapist, mean (SD)=0.76 (0.24), median (min, max)=0.80 (0.01, 1.00)	-	-	HUI2 (Kappa) Sensation=0.32 Emotion=0.37 Cognition=0.70 Self-care=0.43	Both children and physiotherapists rated lower utility scores using HUI3 than HUI2. However, for both the MAUIs children rated lower utility than their parents with the mean difference higher on HUI3. Except for cognition domain emotion and sensation domain had poor agreement between child-proxy reports.
Glaser et al 1999b	Cancer survivors: brain tumour (PNET, astrocytoma, medulloblastoma,	30 Parents 25 physioth	HUI2	Not reported	-	-	**HUI2** children Vs parents; ICC=0.57 , Pearson r=0.59,	Children had moderate correlation and agreement with their parents' utility ratings. Physicians

	ependymoma)	erapist 23 physicians 28 Children					p<0.01 Children vs physicians; ICC=0.15; Pearson r=0.37, p=0.12 Children vs physiotherapists, ICC=0.40, Pearson r=0.54, p<0.001	showed poor correlation and agreement with children utility scores.
Penn et al 2011; UK	Cancer patients: brain tumour (ependymoma, germ cell tumour, astrocytoma, medulloblastoma, meningioma, pineoblastoma, craniopharyngioma); Comorbidity (hydrocephalus);	29 parents 21 children	HUI3	Overall utility, domain scores and mean differences at 12 months follow-up provided for the child-parent proxy pairs Children; mean (SD, range)= 0.84 (0.21, 0.06 to 1.00) Parents; mean (SD, range)= 0.74 (0.29, -0.15 to 1.00) <u>Domains</u> <u>Children; mean (SD)</u> Vision =0.98 (0.09) Hearing=0.98 (0.11) Speech=0.98 (0.05) Ambulation=0.94 (0.19) Dexterity= 0.95 (0.12) Emotion= 0.97 (0.04) Cognition= 0.92 (0.11) Pain= 0.94 (0.22) <u>Parents; mean (SD)</u> Vision =0.97 (0.09) Hearing=0.96 (0.13) Speech=0.98 (0.07) Ambulation=0.91 (0.26)	-	-		Spearman correlation coefficient(r) Overall= 0.76 Vision=0.73 Hearing=1.00 Speech=1.00 Ambulation=0.82 Dexterity=0.79 Emotion=0.30 Cognition=0.75 Pain=0.20	Parents and child reports provided at 12 months after diagnosis Children rated significantly higher for overall and most of the attributes. Correlation between parents proxy-report and children self-report for overall and the attribute vision, hearing, speech, ambulation, dexterity and cognition was good to excellent but moderate to poor for emotion and pain respectively.

		Dexterity= 0.95 (0.14) Emotion= 0.97 (0.07) Cognition= 0.86 (0.19) Pain= 0.90 (0.21) Mean differences (95% CI, P) overall=0.07 (0.00 to 0.14, <0.001) Vision= 0.02 (-0.01 to 0.04, <0.001) Hearing= 0.00 Speech=0.01 (-0.01to 0.02) Ambulation= 0.02 (-0.05 to 0.09, <0.001) Dexterity= 0.01(-0.04 to 0.06, <0.001) Emotion= 0.02 (-0.02 to 0.06, 0.192) Cognition= 0.04 (0.00 to 0.09, <0.001) Pain= 0.03 (-0.14 to 0.08, 0.393)	

IQR= Intra quartile range; VAS= Visual Analogue Scale; TTO= Time trade-off, WTP= Willingness-to-pay, HUI= Health utility Index

Interpretation of the ICC<0.4 poor agreement, fair to good agreement 0.40 ≤ ICC < 0.75, excellent ICC≥0.75; Kappa<0.40 poor agreement, fair to moderate 0.40 ≤ ICC < 0.70, excellent ICC≥0.70)

Table 6: Details of the studies that reported dyad self and proxy reports in well (general population) children and adolescents : overall utility/attribute scores, agreement and correlations between child and proxy ratings

Author year	Health conditions	Proxy type/ No of child-Proxy pairs	MAUIs	Child and proxy dyad overall utility and/or domain scores: mean/median (SD/SE/(5% CI/range), mean differences where available	Direct valuation methods (VAS, TTO, SG, WP) where available	Child and proxy direct valuation methods based scores: mean/median (SD/SE/CI) where available	Child and proxy reports: Correlation/agreement (ICC, K) where available	Summary of the findings
Fluchel et al 2008	General health used as controls in the study with cancer survivors	91 physcians/teachers 91 Children	HUI3	Mean differences of overall and domains (child – proxy reports) Overall =-0.167 Vision =-0.003 Hearing =0 Speech =-0.008 Ambulation = 0 Dexterity =0 Emotion =-0.139 Cognition = -0.068 Pain = -0.009	-	-	ICC (95% CI) Overall =0.843 (0.795, 0.88) Vision =0.699 (0.619,0.765) Hearing =-0.006 (-0.15,0.138) Speech =0.038 (---0.106, 0.181) Ambulation =0.248 (0.11, 0.378) Dexterity =0.188 0.046,0.657) Emotion=0.433 (0.297,0.552) Cognition =0.829 (0.778,0.87) Pain =0.569 (0.465,0.657)	Proxies of children with no obvious health issues tend to slightly overestimate or rate similarly as their children. Most of the domains and overall scores showed moderate to high agreement except for hearing, ambulation and dexterity domains.
Penn et al 2011; UK	General health used as controls in the study with cancer survivors	32 parents 22 children	HUI3	Overall utility scores at 12 months follow-up provided for the child-proxy pair Children; mean (SD, range)= 0.94 (0.13, 0.56 to 1.00) Parents; mean (SD, range)= 0.96 (0.09, 0.63 to 1.00) **Domains** **Children, mean (SD)** Vision=0.99 (0.02)	-	-	Spearman correlation coefficient(r) Overall= 0.31	Moderate correlation between parent proxy-report and child-reports. Children rated slightly higher or similar to their parents on overall utility and attributed scores. Ceiling

				Hearing= 1.00 (0.00) Speech=1.00 (0.00) Ambulation= 1.00 (0.00) Dexterity= 0.98 (0.12) Emotion= 0.98 (0.06) Cognition=0.97 (0.09) Pain=0.99 (0.03) **Parents, mean (SD)** Vision=1.00 (0.00) Hearing= 1.00 (0.00) Speech=1.00 (0.00) Ambulation=1.00 (0.00) Dexterity= 0.98 (0.10) Emotion= 0.99 (0.05) Cognition=0.98 (0.06) Pain=0.98 (0.06)				effect was observed as most of the controls (with no establish health conditions) and their parents had a maximum possible score of 1.00.
Baumann et al 2016	Normal birthweight/ full term	Parents	HUI3	Absolute overall utility scores reported Controls (normal birth weight/full term) child=0.89 (CI0.87 to 0.92); Parent= 0.95 (CI=0.95 to 0.96)	Not used	-	-	Parents slightly overestimated (mean difference= -0.06) for normal birthweight or full-term children
Belfort et al 2011	Health weight children (≥8 years)	Parents	HUI3	Children with healthy weight Overall=0.82 (95% CI= 0.77, 0.88) Ambulation= 1 (95% CI=1,1) Emotion= 0.96 (95% CI=0.93, 0.98) Cognition=0.87 (0.81, 0.92) Pain=0.95 (95% CI=0.90, 0.97) Parents (>8 years children) Overall=0.84 (95% CI=	Not used	-	-	Children rated similar or slightly lower than their parents on overall utility and domains. Parents underestimated their children utility on pain domain by 0.04.

				0.74, 0.92) Ambulation= 1 (95% CI=1,1) Emotion= 0.96 (95% CI=0.93,0.99) Cognition=0.88 (0.80, 0.95) Pain=0.91 (95% CI= 0.82,0.97)				
Belfort et al 2011	Healthy weight children (≥12 years)	Parents	HUI3	Healthy weight children (>12 years) Overall=0.85 (95% CI= 0.80, 0.90) Ambulation= 1 (95% CI=1,1) Emotion= 0.98 (95% CI=0.96,1.0) Cognition=0.87 (0.81, 0.92) Pain=0.96 (95% CI=0.93, 0.99) Parents of healthy weight children (>12 years children) Overall=0.86 (95% CI= 0.73, 0.95) Ambulation= 1 (95% CI=1,1) Emotion= 0.97 (95% CI=0.93,0.99) Cognition=0.91 (95% CI=0.81,0.98) Pain=0.90 (95% CI= 0.76,0.95)	Not used	-	-	Children rated similar or slightly lower than their parents on overall utility and domains. Parents underestimated on pain domain by 0.06.
Gusi et al 2014	General population	442 Mothers and children	EQ-5D-Y and EQ-	Not reported	-	-	Kappa Children vs mothers	Mother had poor agreement on emotional domain but moderate on

			5D-Y Proxy				Pain & discomfort=0.68 (p<0.05) Feeling worried, sad or unhappy=0.221 (p<0.05)	pain domain with their children.
Jelsma and Ramma 2010	Children in open school (General school population with no obvious chronic illness)	530 mothers 158 children	EQ-5D-Y	Not reported	EQ-5D-Y VAS	Children from open school; mean (SD)= 87.9 (16.5) Mothers; mean (SD)=88.4 (15.7)	Open school VAS scores Pearson's r=0.33 (p<0.01) Kappa Mobility=0.15 Looking after myself=0.08 Doing usual activities=0.01 Having pain or discomfort=0.20 Feeling worried, sad an unhappy=0.21	Mothers rated similar health status as their children using VAS scale. Poor agreement on all the domains between child and proxy respondents.
Verrips et al 2001	Very low birthweight; Very preterm birth	203 (mail), 100 (telephone) and 103 (face-to-face) parents and children pairs	HUI3	Overall utility scores by mode of administration **Mail** Children; mean (95%CI)=0.90 (0.88to 0.92) Parents; mean (95%CI)=0.93 (0.91 to 0.94) **Telephone** Children; mean (95%CI)=0.87 (0.84 to 0.90) Parents; mean (95%CI)=0.85 (0.81 to 0.88)	-	-	Kappa agreement for domains Vision, k=0.62 Speech, k=0.24 Ambulation, k=0.50 Dexterity, k=0.80 Cognition, k=0.38 Emotion, k=0.35 Pain, k=0.23	Children rated higher than parents when administered via telephone or face-to-face than their parents. For postal mail survey, children rated lower utility than their parents. However, the mean differences were not statistically

				Face-to-face Children; mean (95% CI)= 0.84 (0.81-0.88) Parents; mean (95% CI)= 0.83 (0.79 to 0.87)				significant. Agreement was high in the physical attributed but poor in psychological attributes.
Wolke et al, 2013	General health (control) Full term	282 parents and adolescents	HUI3	Overall mean utilities Adolescent; mean (SD)=0.87 (0.15) Parents; mean (SD)=0.91 (0.11)	-	-	-	Adolescent rated utilities significantly lower than their parents

IQR= Intra quartile range; VAS= Visual Analogue Scale; TTO= Time trade-off, WTP= Willingness-to-pay, HUI= Health utility Index

Interpretation of the ICC<0.4 poor agreement, fair to good agreement $0.40 \leq ICC < 0.75$, excellent ICC≥0.75; Kappa<0.40 poor agreement, fair to moderate $0.40 \leq ICC < 0.70$, excellent ICC≥0.70)

ETHICS APPROVAL STATEMENT & AUTHOR DECLARATION

The manuscript is a systematic review, therefore an ethics approval is not required.
We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome. We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). She is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from Julie.ratcliffe@flinders.edu.au or jyoti.khadka@flinders.edu.au

Signed by all authors as follows:

Dr Jyoti Khadka ………………………… 29/04/2019

Joesph Kwon ………………………… 24/04/2019

Prof Stavros Petrou ………………………… 30/4/2019

Prof Emily Lancsar ………………………… 30/4/19

Prof Julie Ratcliffe ………………………… 29/04/2019