The RETRIEVE Checklist for Studies Reporting the Elicitation of Stated Preferences for Child Health-Related Quality of Life

Cate Bailey[1] · Martin Howell[2,3] · Rakhee Raghunandan[2,3] · Kim Dalziel[1] · Kirsten Howard[2,3] · Brendan Mulhern[4] · Stavros Petrou[5] · Donna Rowen[6] · Amber Salisbury[2,3] · Rosalie Viney[4] · Emily Lancsar[7] · Nancy Devlin[1]

Abstract

Background Recent systematic reviews show varying methods for eliciting, modelling, and reporting preference-based values for child health-related quality-of-life (HRQoL) outcomes, thus producing value sets with different characteristics. Reporting in many of the reviewed studies was found to be incomplete and inconsistent, making them difficult to assess. Checklists can help to improve standards of reporting; however, existing checklists do not address methodological issues for valuing child HRQoL. Existing checklists also focus on reporting methods and processes used in developing HRQoL values, with less focus on reporting of the values' key characteristics and properties. We aimed to develop a checklist for studies generating values for child HRQoL, including for disease-specific states and value sets for generic child HRQoL instruments.

Development A conceptual model provided a structure for grouping items into five modules. Potential items were sourced from an adult HRQoL checklist review, with additional items specific to children developed using recent reviews. Checklist items were reduced by eliminating duplication and overlap, then refined for relevance and clarity via an iterative process. Long and short checklist versions were produced for different user needs. The resulting long RETRIEVE contains 83 items, with modules for reporting methods (A–D) and characteristics of values (E), for researchers planning and reporting child health valuation studies. The short RETRIEVE contains 14 items for decision makers or researchers choosing value sets.

Conclusion Applying the RETRIEVE checklists to relevant studies suggests feasibility. RETRIEVE has the potential to improve completeness in the reporting of preference-based values for child HRQOL outcomes and to improve assessment of preference-based value sets.

Key Points

Reporting of values for child health-related quality-of-life (HRQoL) was found to be incomplete and inconsistent in a series of systematic reviews.

To improve reporting, a checklist for studies reporting the elicitation of stated preferences for child HRQoL was developed, including long and short versions.

The checklists are shown to be feasible to use for both paediatric HRQoL value sets and for values for HRQoL vignettes, and should improve the reporting of HRQoL values for children.

Cate Bailey and Martin Howell are joint first authors on this work.

Emily Lancsar and Nancy Devlin are joint senior authors on this work.

Extended author information available on the last page of the article

1 Introduction

Economic evaluation is a cornerstone of health economics and is used to inform resource allocation decisions across technologies, such as medicines, services, and tests [1, 2]. When considering interventions targeted at children and young people, the development of health-related quality of life (HRQoL) instruments specific to the measurement of child health [3, 4], and the valuation of child HRQoL anchored on a 0–1 scale required for estimation of quality-adjusted life-years (QALYs) [5], are key elements. There is a lack of consensus, though, about fundamental aspects of these research methods used in valuing child HRQoL [6, 7].

It is crucial that those choosing which preference-based values for child HRQoL to use for QALY estimation and subsequent application in economic evaluation, and those using that evidence in decision making, are aware of the underlying characteristics of the values. There are numerous characteristics that might affect and limit

the comparability of evidence on HRQoL and QALYs in children and young people. However, in a recent systematic review of measurement and valuation of child HRQoL [3], a review of the psychometric performance of generic childhood multi-attribute utility instruments [4], and a review of the methods used to value child HRQoL [8], authors concluded that the reporting of such studies is often incomplete and inconsistent. Poor reporting of methods used to value child HRQoL and the values derived makes it difficult for users of these instruments to make informed choices and for decision makers to use the evidence in an informed way.

Various checklists are available for reporting the estimation of adult HRQoL values, including CREATE [9] and SpRUCE [10]. While the methodological considerations relevant to producing values for adult HRQoL are also relevant to HRQoL of children and young people, there are additional considerations that are unique to valuing childhood HRQoL. These considerations relate to fundamental aspects of the valuation task; for instance, whose stated preferences are considered relevant when valuing child HRQoL, and from what perspective are they asked to imagine the states they are requested to value? These questions are more complicated for valuing child value sets than for adults, who are generally asked about their own preferences. The choice of duration of the state used in the valuation task is an issue that arises in valuing adult HRQoL, but this choice interacts with other choices specific to child HRQoL in complex ways, particularly when related to the age of a hypothetical child (as is often used for child valuation). As such, existing checklists do not provide an adequate basis for guiding the reporting and assessment of values and/or value sets for childhood HRQoL.

Existing checklists also tend to focus on reporting the methods and processes used in developing HRQoL values. There has been much less focus on reporting of the values themselves and their key characteristics and properties. This issue is particularly important for child values because of the wide range of methods used to value child HRQoL for QALYs, resulting in utility values for child health that have notably different properties depending on the methods used in their generation [3]. Comprehensive reporting would enable users to understand the methods and process issues in developing child HRQoL values with more confidence.

The aim of this study was to develop a checklist to support the reporting of methods and results from studies of values for childhood HRQoL. The checklist will be applicable to a broad range of studies that aim to produce values for childhood HRQoL. The checklist can be used to assess studies that produce value sets for child HRQoL instruments (both generic and disease-specific) for QALY estimation, as well as studies that seek to produce values for a limited number of specific child health states, e.g., described by

vignettes, or a selection of states from a disease-specific child patient-reported outcome measure. These types of studies (such as those included in the systematic review by Bailey et al. [8]) have been used in cost-effectiveness models considered by decision-making groups such as the Australian Pharmaceutical Benefits Advisory Committee (PBAC) [5], and other health technology assessment agencies such as the National Institute for Health and Care Excellence (NICE) and the Canadian Agency for Drugs and Technologies in Health (CADTH).

Improved reporting of the methods used to generate HRQoL values for children will allow users to better select values and to evaluate and compare results across studies. Improved reporting will also aid decision makers to better understand the sources of values and the implications of differences in values for the interpretation of cost-effectiveness evidence. In this study, we have used the term 'values' for preference weights (often referred to as utilities, values, or QALY weights) in line with our previous study [8]. Child and young person (or for brevity, child) is used here to describe a person under 18 years of age.

2 Development

2.1 EQUATOR Guidelines for Developing the Checklist

Our methodology for the development of the checklist has been adapted from the EQUATOR Network guidelines for developing reporting checklists [11], such as identifying the need for a checklist via systematic reviews, and around our and others' recent work (Sections 1 and 2 of the EQUATOR Network guidelines) [3, 8, 12]. The reporting checklist was then developed following the EQUATOR toolkit, including generating a list of items and conducting a series of meetings (Section 3). We have however provided a single paper rather than follow the process recommended by EQUATOR (Section 4), which suggests a short explanatory paper alongside a longer 'Explanation and Elaboration document'. Dissemination methods as suggested by the EQUATOR network guidelines are outlined in the discussion section.

2.2 Developing a Conceptual Framework to Provide a Foundation for the Checklist

A conceptual framework for the checklist was developed to ensure its relevance for reporting values for child HRQoL, whether the values are for individual health states (e.g., described via vignettes) or value sets, such as reporting values for all health states described by a HRQoL instrument. Given the differences between these study types, a modular

approach was developed to allow flexibility for application to different study types. The modular approach also allowed us to differentiate between checklist items specific to valuation of child HRQoL and those that are important to include but are also common to reporting of adult HRQoL, thereby providing a standalone comprehensive checklist for children. An initial conceptual framework was developed by the authors to identify relevant modules, informed by existing checklists for adult HRQoL values [12] and reviews of methods for valuing child HRQoL [3, 8]. This was refined through checklist item development and testing, using an iterative process (expanded on below).

2.3 The Conceptual Model

The conceptual model for the RETRIEVE (REporting inven-ToRy chIld hEalth ValuEs) checklists is shown in Fig. 1. The checklists are structured using five 'top level' headline groupings (modules) of items. Four of the modules contain items relating to key aspects of the methods used to obtain child HRQoL values (A–D), with the fifth (E) comprising checklist items relating to the characteristics of the values themselves. The modules are not necessarily hierarchical, as decisions relevant to some modules are made simultaneously rather than sequentially and are often iterative. Figure 1 is therefore non-hierarchical. We note that there are likely to be interactions between methods decisions in each module, such as between population and anchoring or method and perspective.

Modules A1–A3 are specific to considerations relating to child HRQoL values. The items they contain are not derived from any of the existing checklists for adult HRQoL values. Modules B2 and B3 are alternative modules that users select depending on whether the values they are considering are value sets (B1) or values for specific states or vignettes (B2). Modules A4 and C contain general methods and sample considerations. These are not necessarily specific to values for childhood HRQoL but are an important part of what users of values would need to check and developers to report. Module D relates to considerations relevant to modelling value sets for an HRQoL descriptive system, so are further relevant considerations to B1 (value sets for patient-reported outcome measures) but not B2 (direct valuation of disease-specific states or vignettes).

Checklists developed for adult HRQoL values have tended to focus on reporting the methods used to produce a given set of values, or on the clarity of reporting the final value-set model (i.e., like checklist Modules A–D described above). We considered it important that our checklist included a module focusing on the characteristics of the values, to ensure users are aware of these, and the relevant differences in values when choosing between instruments and value-sets. Including this module would help decision makers be aware of the potential implications of such differences when interpreting cost-effectiveness evidence based on them, and to encourage more complete reporting of these value characteristics by study teams (Module E).
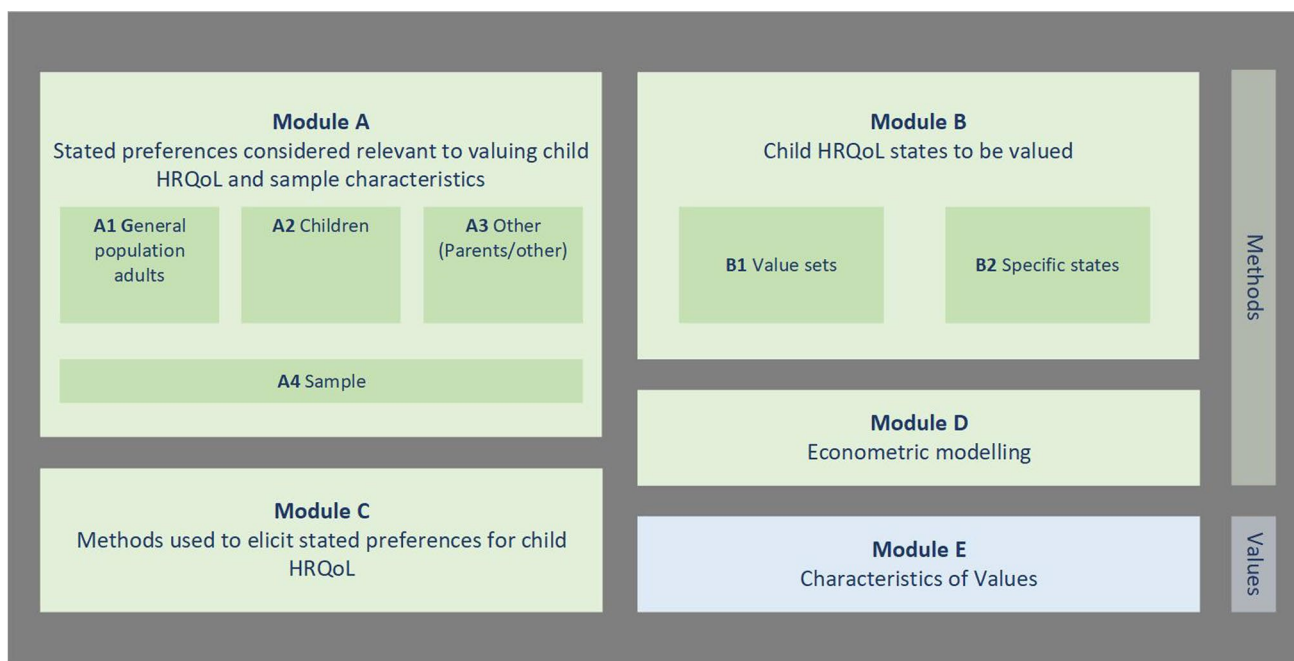

Fig. 1 A conceptual framework for the RETRIEVE modular checklists for reporting values for child HRQoL. Note: the modules are intended to be non-hierarchical. *HRQoL* health-related quality of life

2.4 Establishing Potential Items for Each Module

A review of items for reporting values for adult HRQoL [12] was used to identify items common to both adult and child HRQoL. Two sets of checklists were included in the study by Zoratti et al., i.e. those intended primarily for use in economic evaluation and those primarily intended for use for health utility studies (see Tables 1–6 and 7–12, respectively, in the article by Zoratti et al. [12]. Items from the latter were considered for our checklist, with potential items also identified from Table 7 from Brazier et al. [13], Table 8 from Stalmeier et al. [14], Table 10 from CREATE [9], Table 11 from Nerich et al. [15] and Table 12 from SpRUCE [10]. We did not include Table 9—MAPS [16], as that checklist is relevant to studies mapping across instruments and thus outside the scope of our checklist. Items from the included checklists provided a pool of potential items. These items were grouped by the modules in the conceptual framework by two members of the team (CB and RR) and then further independently reviewed (EL and ND).

We supplemented this pool of potential items with additional items specific to valuation of child HRQoL. The latter items were generated based on (1) methods issues relating to valuation of child HRQoL as identified by Rowen et al. [7], and (2) information from two systematic reviews [3, 8] on aspects of methods specific to valuation of child HRQoL and what was viewed as missing or unclear from the papers reporting values for child HRQoL that were included in those reviews. Combined, this process yielded a list of candidate items under each module. The original list of items, and subsequent versions created through the review process described in the following section, are available from the authors on request.

2.5 Creating an Initial List of Items for Each Module (Long Version)

A series of five meetings were held with a subset of the study team (CB, MH, ND, EL, RV, RR), where items in each module were each considered, with the objective of identifying redundancy or overlap between modules and to check for relevance. Meetings were structured, with an agenda circulated to the team by the first author (CB) prior to the meeting. Decision making was through consensus. Where gaps were identified, new items were created and/or wording clarified. Changes arose most often in the items specific to child HRQoL rather than those also applicable to adults. This collaborative and iterative process led to the creation of an initial draft checklist of 147 items grouped into five modules.

The process of eliminating redundant items and checking relevance yielded a first draft that was considered potentially usable. During this process, the conceptual model was reviewed to ensure the checklist items were grouped appropriately. The first draft of the checklist items was then distributed to the entire authorship team who were invited to comment. The commentary was compiled and the checklist items were edited accordingly (MH, CB).

2.6 Reducing Items for the Short Version

To produce the short version, all authors were asked to review the proposed items using a numbering system (1 = include, 2 = maybe include, 3 = do not include), providing specific comment on the items and to recommend revised or additional items for inclusion (if any). All responses were coded to facilitate refinement of the checklist. A first version of the short version contained 18 items, with a further 15 items as alternatives containing different wording. This was revised to 14 final items, with the format modelled on the CHEERS checklist [17], where, instead of questions, users are asked to indicate where the relevant information is located in the manuscript by page number.

2.7 Testing the Checklists: An Application to Four Studies of Child Health-Related Quality of Life Values

The checklists were evaluated using a sample of studies that report child HRQoL values. We selected four studies published between 2010 and 2021 that had been included in our earlier systematic review [8]. These papers were selected to check that the module approach worked for both value sets and vignettes, were spaced over a range of years, and featured value sets from the two most widely used child HRQoL instruments. The two papers on value sets were on the EQ-5D-Y-3L [18] and the CHU9D [19]. Two papers used vignettes [20, 21]. In each case, two members of the authorship team independently used the checklists to review and summarise the study (CB, MH, RR, KD). These reviews were compared and reported to the wider study team for discussion. Any need for refinement of the checklists was identified and implemented (MH, CB) via an iterative process.

2.8 Expert Review of the Checklists

The authors invited input from senior international health economic researchers who are part of the wider QUOKKA and TORCH project teams ('Associate Investigators') using an online survey. These researchers were from Canada, the UK, Australia, Spain and Singapore. Participants were asked to indicate whether items were relevant, redundant, or required wording changes. Information from the

reviews was compiled and a final workshop was held (CB, MH, ND, EL) to review and address survey responses and inclusion of new items. We received six expert reviews. The reviewers commented chiefly on wording and recommended possible extra questions. The review comments were incorporated (CB, MH) and decisions on any extra question suggestions were workshopped (CB, MH, ND, EL). The final short and long versions of the checklist were then completed. After this final review, we updated the examples as described in Section 2.5.

2.9 Long and Short Versions of RETRIEVE

The resulting RETRIEVE checklists contain modules aimed at reporting methods (A–D) and the characteristics of values (E). The long version of RETRIEVE (Table 1) is populated with a total of 83 items (noting that because of the modular structure, not all items are relevant to all valuation studies) in question form with specified or open-ended response format. The short version of RETRIEVE (Table 2) has 14 items where the user notes where in the paper the information is contained, similar to the CHEERS checklist [17]. Electronic supplementary material (ESM) Table S1 contains a formatted version of the long RETRIEVE; ESM Table S2 contains examples of the use of the long and short RETRIEVE checklists; and ESM Table S3 contains a table of descriptive comments for each included item in the long RETRIEVE version. We also include editable excel versions of both versions in the ESM.

2.10 Differential Use of the Long and Short Versions of RETRIEVE

We considered the different needs of two broad sets of potential users of the checklist—decision makers and researchers. A longer list of items in each module was considered relevant for researchers undertaking and reporting on child valuation studies, to improve completeness of reporting. This list has more extensive descriptions of the aspects being identified and goes into more detail regarding valuation in the last module. A more concise version of the checklist was considered to be more appropriate for decision makers or other users of values wishing to assess, compare and choose between values for child HRQoL. This checklist is presented as statements to check off, similar to that used in the CHEERS checklist [17].

3 Discussion

This paper reports the first checklist for studies reporting values for HRQoL of children. There has been a notable increase in research aimed at producing values for child HRQoL in recent years; for instance, many value-sets for the EQ-5D-Y-3L (for example [18, 22–26]) instrument have commenced or been completed since 2020 [27]. However, the methods being used to value the EQ-5D-Y-3L and other childhood HRQoL instruments vary widely [3, 8], are not always fully reported, and the values can have quite different characteristics. The short version of RETRIEVE will allow users to better understand and be aware of the implications of methods differences when choosing which published values to use. The long version of RETRIEVE is relevant to those designing and reporting studies of values for child HRQoL and will encourage more complete and consistent reporting of methods and results. Our objective was to develop a checklist for studies generating values for child HRQoL, including for disease-specific states or vignettes and value-sets for generic child HRQoL instruments and thereby fill a key gap. RETRIEVE is intended as a standalone checklist and therefore includes items that are also relevant to studies developing adult HRQoL values.

The conceptual framework and selection of modules was based on the combined expert views of the authors drawn from our two research teams, QUOKKA and TORCH, comprising expert health economists across Australia, the UK, and North America. Similar to the authors, expert reviewers were health economists, but from a wider range of geographic locations. Both research teams were funded simultaneously by the Australian Government's Medical Research Future Fund (MRFF) to improve the measurement and valuation of child health to strengthen decision-making in Australia. The process of initial item generation for the long version, and refinement leading to the checklist items reported in Table 1, reflects our individual and collective experiences and opinions as researchers. There is inevitably a degree of subjectivity and judgement involved in all such checklists, and different ways of grouping and presenting the relevant checklist items would be possible. We have been mindful of this, and as a team have reflected on the possible biases that are introduced throughout the process of developing the checklist. We therefore resolved on wider consultation and feedback among the research community, which we achieved through the six external expert reviews.

Similarly, we are mindful of the challenge in striking a balance between (a) providing a full account of relevant features of methods and values, and (b) providing a checklist that is sufficiently concise to be readily used by others. The checklist reported in Table 1 contains a larger number of items than other checklists (e.g. CREATE; [9]), although its modular structure means not all these items will be relevant to all study types. While the checklist was feasible for our team to use, we recognise that there are a range of different potential users (e.g., those designing clinical trials; or choosing between available value-sets for a given instrument to use in economic evaluation) for whom the correct

Table 1 RETRIEVE long checklist (please note that this checklist is modular and not all sections/questions will apply to all papers)

Section A: Stated preferences considered relevant to valuing child HRQoL and sample characteristics

A1: Stated preferences

A1a: Whose preferences were sought? (Adults/children and young people [CYP] <18 years/mixed adults and CYP)

A1b: Did the authors provide a rationale for whose preferences were sought? (Yes/No)

A2: Adults' stated preferences

A2a: Which adults were the focus of preference elicitation? (General population/parent or caregiver of child/healthcare professionals/adult with a health condition/other adults, please specify: ____)

A2b: What perspective were adults asked to take in considering the child states to be valued? (e.g. thinking about the health states as experienced by: own child [parent]/another child they know/a hypothetical child/their own health, thinking back to when they were a child/their own health, as if they were a child now/their own health, but blinded to the states under consideration being specific to children/persons with a health condition [e.g. a health professional asked to take the person with a health condition's perspective]/other, please specify: ____)

A2c: Was the age of the child, for whom respondents were asked to imagine health states to be valued, specified? (Yes/No/Not applicable)

A2d: If yes, what was the age of the child? _____

A2e: Was the rationale for the choice of the age of child provided? (Yes/No)

A3: Children and young people's stated preferences

A3a: From which child/young person were preferences elicited? (General population/person with a health condition/other child, please specify: ____)

A3b: What perspective was the (child/young person) respondent asked to take? (e.g. thinking about the health states as experienced by: themselves [i.e. their own perspective]/another known child/a hypothetical child/other, please specify:____)

A3c: Was the age of the child/young person, for whom respondents were asked to imagine health states to be valued, specified? (Not applicable [i.e. own perspective/themselves]/it was applicable but not stated/yes)

A3d: If the age was specified, what was the age?_____

A3e: Was the rationale for the choice of the age of child/young person provided? (Yes/No)

A4: Sample

A4a: Was the population or sample frame defined from which the sample was drawn, e.g., country, age, condition? (Yes/No)

A4b: Is information provided on how the sample was recruited, e.g., field based-recruitment, online panel, convenience sample? (Yes/Partial/No)

A4c: If data were collected online, were efforts made to avoid online panel fraud? (e.g, related to bots or automated software posing as participants and completing surveys). (Yes/No/Not applicable)

A4d: Was there a target sample size (or sample sizes if by block, e.g. number of tasks per block [e.g. DCE] or health state [e.g. TTO])? (Yes/No)

A4e: Was the target sample justified? (Yes/No)

A4f: Was the target sample achieved? (Yes/No/Unclear)

A4g: Were the characteristics of the final sample described? (Yes/No)

A4h: Did the sample characteristics match the intended population? (Yes/No/Unclear)

A4i: Was the year the data collected stated? (Yes – what year(s) were the data collected?/No)

A4j: Was information provided on missing data (non-completion, withdrawals)? (Yes/Partial/No)

Section B: Child HRQoL states to be valued

B1: Type of study

B1: Did the values reported in this paper comprise: (a value set?/values for a limited number of health states [e.g. vignette]?)

B2: Value Sets

B2a: Which HRQoL instrument was valued? _____

B2b: Were the domains and response options of the instrument clearly described? (Yes/No)

B2c: What experimental design approach was used to choose the health states (combination of dimension levels) to be valued?

B2d: How were the health states assigned to respondents?

B3: Specific health states

B3a: How were the health states described? (Disease-specific vignettes/from a disease-specific HRQoL instrument/other, please specify ____)

B3b: How many health states were preferences elicited for? _____

B3c: Was the rationale for the selection of these health states specified? (Yes – what was the rationale?/No)

Section C: Methods used to elicit stated preferences for child HRQoL

C1: Which method or methods were used to elicit stated preferences? DCE/TTO/SG/BWS/VAS/other, please specify____)

C2: Was a rationale for the choice of method(s) provided? (Yes/No)

Table 1 (continued)

C2a: If yes, what was the rationale?____

C3: Was the duration of the states to be valued reported (e.g. 'x years in this state, followed by death')? (Yes/No)

C3a: Was the duration fixed? (Yes/No)

C3b: What duration(s) was used? _____

C4: Did the method(s) allow values to be elicited that were <0 ('worse than dead')? (Yes/No)

C4a: How were values <0 elicited?

C4b: What was the minimum value possible (may vary according to the method used, so should be clearly stated)?

C4c: What determined how the task was terminated? _____

C5: How were the values anchored on a utility scale? _____

C6: What was the mode of administration for the stated preference tasks? (Online self-completion by the respondent/self-completion of mailed questionnaires/online computer-assisted personal interview (CAPI)/in person CAPI/in person interview/other, please specify ____)

C7: How was the quality of stated preference data assessed? ____

C8: Were any exclusions made to the preference data (e.g. used to represent average preferences)? (Yes/No/Unclear)

C8a: Were reasons for the exclusions provided? (Yes/No/Unclear)

C9: Were the health states randomly assigned? (Yes/No/Unclear)

C10: Was ethics approval for the study obtained from an appropriate research ethics committee? (Yes/No/Unclear/Not stated)

C11: Were sources of funding and non-monetary support and the role of the funder(s) in the design described? (Yes/No)

Section D: Econometric modelling and statistical methods

D1: Did the values reported comprise: (a value-set?/values for a limited number of health states [vignette or condition-specific]?)

D2: Econometric modelling of value-sets for HRQoL instruments

D2a: What was the theoretical model? OR What models were estimated, e.g., OLS, Tobit etc.?

D2b: Were the main assumptions of the model stated (e.g., assumptions about preference homogeneity/heterogeneity)? (Yes/No/Unclear)

D2c: How was the constant term treated (if included)?

D2d: How were missing data handled (e.g. imputation, complete case analysis)

D2e: Were subgroup analyses completed? (Yes/No/Not applicable)

D2f: Were interaction terms included? (Yes/No)

D2g: Were details of the interactions provided? (Yes/No/Not applicable)

D2h: Were non-linear specifications considered? (Yes/No)

D2i: Was more than one model described? (Yes/No)

D2j: Were goodness-of-fit statistics for each model reported? (Yes/No)

D2k: Was the preferred model clearly stated? (Yes/No)

D2l: Were the criteria used to select the preferred model described? (Yes/No)

D2m: Do the preference parameters for the health states follow a logical order (monotonic)? (Yes/No)

D2n: Was any post estimation undertaken to force monotonicity (e.g. collapsing levels)? (Yes/No/Unclear or not stated)

D2o: How were insignificant differences between adjacent levels managed (e.g. collapsed/forced to be different)?___

D2p: Were robustness checks conducted? (Yes/No)

D2q: Was uncertainty around values reported? (Yes/No)

D3 Analysis of values for specific HRQoL states

D3a: Have the statistical methods been described? (Yes/No)

D3b: Have the statistical methods been justified? (Yes/No)

D3c: How were missing data handled (e.g. imputation, complete case analysis)? ___

D3d: Have subgroup analyses and interactions been undertaken? (Yes/No)

D3e: Were subgroups and interaction variables chosen for assessment justified? (Yes/No)

D3f: Were sensitivity analyses undertaken? (Yes/No)

D3g: Were sensitivity analyses described?

Section E: Characteristics of values

E1: Was qualitative or quantitative evidence reported that demonstrates the extent to which respondents engaged with and understood the valuation tasks? (Yes/No)

Table 1 (continued)

E2: Where a value was reported, were the values generated by the final model logically consistent? (Yes/No/Unclear)

E3: Did the authors report the distribution of values over all states defined by the HRQoL instrument (e.g. as per Fig. 1 in the study by Pan et al.) (Yes/No)

E4: Key characteristics of the values

E4a: How many values less than zero were possible? _____

E4b: What was the maximum possible value less than 1? _____

E4c: Where in the descriptive system does the biggest change in values occur, when shifting between adjacent states? _____

E5: Was the order of importance of dimensions (domains) suggested by the value set discussed? (Yes/No)

E6: Did the authors report on specific requirements of users and decision makers about how such values are produced, e.g., as set out in the methods guides of local HTA bodies? (Yes/No/Not applicable)

Table 2 RETRIEVE short checklist (please note that this checklist is modular and not all sections/questions will apply to all papers)

MODULE A	Stated preferences considered relevant to valuing child HRQoL and sample characteristics	Location
1	Whose preferences were sought was stated	Page *x*, etc.
2	Whose perspective was used was stated	
3	If the perspective was as a child, the child's age was stated	
4	The population from which the sample was drawn was described and justified	
5	The target sample size was provided and achieved	
MODULE B	**Child HRQoL states to be valued**	
6	The HRQoL instrument or health states being valued were described	
7	The choice of health states being valued was stated and justified	
MODULE C	**Methods used to elicit stated preferences for child HRQoL**	
8	The valuation methods used to value health states were described and justified (e.g. cTTO, DCE, etc.)	
9	The mode of administration for the valuation tasks was stated (e.g. face-to-face, online, in person, etc.)	
10	How values were anchored at 1 = full health and 0 = dead was stated	
MODULE D	**Econometric modelling and statistical methods**	
11	The modelling methods applied to the data were stated and justified	
12	The basis for choosing the final model and any post-model decisions were clearly stated and justified	
MODULE E	**Characteristics of values**	
13	The characteristics and distributions of values for all health states relevant to the study were reported	
14	If a value set is derived for an HRQoL instrument, there was sufficient information to enable readers to estimate utility scores for all health states described by the instrument	

balance between depth and brevity may be different than for researchers reviewing others' (or reporting their own) valuation studies. This was our rationale for developing the concise version of this checklist (Table 2). Thus, the two checklists are suitable for those who 'demand' value-sets and may need only a high-level overview of the methods and results, as well as a nested set of more detailed items aimed at use by those who 'supply' value-sets, to aid comprehensive reporting.

Understanding the characteristics of values is important for users. The properties of the values which are produced from valuation studies have tended to be under-reported, yet substantial differences in the characteristics and properties of values could have non-trivial implications for estimates of QALYs and cost effectiveness generated from their use.

The differences in values may reflect a myriad of different methods choices, and motivations, of the instrument developers (Pickles et al., 2019). While some papers accurately report 'basic' aspects of this, such as the minimum value and the proportion of negative values in value-sets, we found that the reporting of characteristics of values is inconsistent and sometimes inadequate [3, 8]. We additionally suggest in item E4 in the RETRIEVE long version that authors supply the distribution of values over all the states defined by the instrument, which is currently not commonly reported. An example summarising a distribution of 'theoretical' values for an (adult) HRQoL descriptive system can be found in Figure 1 of Pan et al. (2022). We have produced these figures for the two value sets we reported on, see Figs. S1 and S2 in the supplementary files.

We chose to focus module E on HRQoL value characteristics, rather than judge the validity of values based on stated preferences data, as the basis for judging the validity of those values is challenging to determine. Devlin [6] notes that it is difficult to validate HRQoL values in the same way that we can validate stated preferences in other applications and sectors, as "there are few opportunities to observe 'real' choices people make about HRQoL, so we lack the kind of revealed preferences data that would allow us to check that values are meaningful representations of the preferences embodied in decisions" (page 1087). In the absence of revealed preference data on HRQoL, it might be tempting to think that judging validity requires some other kind of external standard or benchmark. If that were true, is not clear what the source of that external standard should be, and where its legitimacy might be derived from.

Given that no value-set can claim to represent a 'gold standard', judging the validity of any value-set based on its similarity to previous values could risk circularity. There are some criteria that might be applied, such as where the object of valuation is a HRQoL instrument. These criteria arise from the properties of HRQoL descriptive systems in the instruments: within these, there are (some) states that are logically ordered and unequivocally (i.e., descriptively, and independent of preferences) better or worse than others. Where one state is descriptively better than another, its value should be higher. This is a *de minimis* criterion of modelled values, but it may be worth checking that value-sets, of the type that Module B2 is concerned with, have this property. This issue is less likely to be relevant for values from vignettes (Module B3). Lancsar and Swait [29] argue, specifically in relation to DCEs, that while external validity has tended to centre on the question of whether people behave in real markets as they state they would in hypothetical markets, it can also be thought of more broadly in terms of process validity. We consider that process validity is analogously relevant when considering validity in the context of HRQoL values (and values for child HRQoL specifically). Many aspects of process validity are captured in modules A–D. For example, the validity of values may be questioned if there are concerns about the quality of the data, regardless of the characteristics of the value-set they yield. Thus, understanding what processes were in place for handling quality assurance (Module C) provides important information for users.

A key aspect of process validity that we considered, but did not include, is whether the methods and processes for obtaining stated preferences are consistent with any requirements for value-sets or values stated by end users of those values. This could include local decision-makers, such as in the methods guides of Health Technology Assessment (HTA) bodies. Given the normative aspects of methods choices regarding valuation of child HRQoL (for example

asking whose values are considered relevant, and from what perspective), a key aspect of process validity is arguably whether these methods are valid when considered from the perspective of the decision-maker and their views on these value judgements. Currently no HTA body has guidelines on the methods to use in valuing child health. However, existing HTA methods guides may contain guidance on general methods choices that, while not specific to child HRQoL valuation, are nevertheless relevant to it; an example of this is NICE's recommendation that values are obtained using 'choice-based methods' [30]. Further, NICE is currently developing guidelines on methods for measuring and valuing child HRQoL and there is growing awareness of the issues around child HRQoL across HTA bodies. Explicitly considering the extent to which the methods used in valuing child HRQoL match HTA bodies' emerging requirements will therefore be important in the future. This aspect is a key consideration for researchers, as their work must be relevant to local bodies. For this reason, the long checklist includes item E6 on whether specific requirements have been acknowledged.

The RETRIEVE checklist focuses on studies reporting stated preferences that are aimed at producing values for child HRQoL. The intention is that the checklist can be applied to any paper with this aim, whether that be to establish single mean values for a small number of specific states described by vignettes or disease-specific instruments, or modelled values for all states defined by a generic childhood HRQoL instrument. Nonetheless, there are aspects of methods used to obtain values for child HRQoL states that are not covered by the checklist. For example, the checklists were not intended to be applicable to studies that report mapping from a disease specific instrument to a generic instrument, or approaches other than direct stated preference methods used to assign values to disease specific states.

The EQUATOR guidelines note that production of a checklist, on its own, will not necessarily result in its use (section 5). In respect to the suggestions included in the guidelines, the long and short forms of the checklist are available [Supplementary files 2, 3] to all in a format that is editable, to submit the checklist for consideration for inclusion on the EQUATOR website, and to present the checklist at conferences and meetings (at the time of publication, this paper has been presented three times). If included on the EQUATOR website, as a minimum we anticipate that authors will include a statement saying that reporting has followed RETRIEVE and preferably include a completed checklist as supplementary material. The impact of the checklist could be followed through citations in relevant papers, HTA decision-making, and extensions or adjustments to the checklist may be undertaken as required (Sect. 6). Further to this, the research team are considering establishing a RETRIEVE database where reporting of value sets using the checklist are

lodged and made publicly available to others. It should also be noted that when designing and/or assessing a study, it is important to consider the geographic location of the authors and reviewers of the RETREIVE checklist and whether there are items not included in the RETRIVE checklist that need to be addressed to meet requirements specific to the country or jurisdiction.

4 Conclusion

RETRIEVE is the first checklist for reporting preference-based values for child HRQoL. We have developed both long and short versions that are targeted at different audiences who we envisage will use the checklists for different purposes. Importantly, RETRIEVE includes items relating to the characteristics of reported values. Existing checklists (such as for values for adult HRQoL) have tended to focus on the adequacy of the reporting of methods used for obtaining values. Going beyond methods to address the characteristics and properties of the values themselves is clearly important from the point of view of the users who are choosing between value-sets. Relatively few papers reporting value-sets for HRQoL (whether for children or adults) detail the full characteristics of the distribution of values, despite that information arguably being crucial for those interpreting evidence from their use. However, going beyond description of the properties of these distributions, to judgements about the validity of the values, remains contentious. We hope our work provides the basis for the further dialogue needed to establish criteria for judging values. This dialogue might include the legitimacy of the process used to generate values, and *ex ante* judgements about the empirical characteristics. Such discussion should also include the extent to which values (and methods used to obtain them) comply with the stated requirements of end users including government decision-makers—which could be regarded as 'context validity'.

Declarations

Conflicts of interest Cate Bailey, Martin Howell, Rakhee Raghunandan, Kim Dalziel, Kirsten Howard, Brendan Mulhern, Stavros Petrou, Donna Rowen, Amber Salisbury, Rosalie Viney, Emily Lancsar, and Nancy Devlin have no conflicts of interest to declare in relation to this work. It should be noted that Nancy Devlin, Brendan Mulhern, Rosalie Viney and Donna Rownen are members of the EuroQol Group, Nancy Devlin is Chair of the Board of the EuroQol Research Foundation, and Rosalie Viney is Chair of the Scientific Executive of the EuroQol Group.

Author contributions Conceptualization: ND, EL, CB, MH, RR. Methodology: EL, ND, CB, MH, RR. Formal analysis and investigation: CB, MH, RR, AS. Writing—original draft preparation: ND, CB. Writing—review and editing: EL, ND, CB, MH, RR, KD, KH, BM, RV, SP, DR, AS. Funding acquisition: ND, EL.

Data availability All data is presented either within the paper or in the supplementary files.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Code availability Not applicable.

References

1. MF Drummond, MJ Sculpher, K Claxton, GL Stoddart, GW Torrance. Methods for the economic evaluation of health care programmes, 4th ed. Oxford: Oxford University Press; 2015. https://global.oup.com/academic/product/methods-for-the-economic-evaluation-of-health-care-programmes-9780199665884?cc=nz&lang=en&

2. PJ Neumann, TG Ganiats, LB Russell, GD Sanders, JE Siegel (eds). Cost-effectiveness in health and medicine. Second edition. New York: Oxford University Press; 2016. https://doi.org/10.1093/acprof:oso/9780190492939.002.0007

3. Kwon J, et al. Systematic review of conceptual, age, measurement and valuation considerations for generic multidimensional childhood patient-reported outcome measures. Pharmacoeconomics. 2022;40(4):477–8. https://doi.org/10.1007/s40273-021-01128-0.

4. Kwon J, et al. Systematic review of the psychometric performance of generic childhood multi-attribute utility instruments. Appl Health Econ Health Policy. 2023;21(4):559–84. https://doi.org/10.1007/s40258-023-00806-8.

5. Bailey C, Dalziel K, Cronin P, Devlin N, Viney R. How are child-specific utility instruments used in decision making in Australia? A review of pharmaceutical benefits advisory committee public summary documents. Pharmacoeconomics. 2021;40(2):157–82. https://doi.org/10.1007/s40273-021-01107-5.

6. Devlin NJ. Valuing Child Health Isn't Child's Play. Value Health. 2022;25(7):1087–9. https://doi.org/10.1016/j.jval.2022.05.009.

7. Rowen D, Rivero-Arias O, Devlin N, Ratcliffe J. Review of Valuation Methods of Preference-Based Measures of Health for Economic Evaluation in Child and Adolescent Populations: Where are We Now and Where are We Going? Pharmacoeconomics. 2020;38(4):325–40. https://doi.org/10.1007/s40273-019-00873-7.

8. Bailey C, et al. Preference elicitation techniques used in valuing children's health-related quality-of-life: a systematic review. Pharmacoeconomics. 2022;40(7):663–98.

9. Xie F, et al. A Checklist for Reporting Valuation Studies of Multi-Attribute Utility-Based Instruments (CREATE). Pharmacoeconomics. 2015;33(8):867–77. https://doi.org/10.1007/s40273-015-0292-9.

10. Brazier J, et al. Identification, review, and use of health state utilities in cost-effectiveness models: an ISPOR Good Practices for Outcomes Research Task Force Report. Value Health. 2019;22(3):267–75. https://doi.org/10.1016/j.jval.2019.01.004.

11. Equator Network. How to develop a reporting guideline. Oxford: Centre for Statistics in Medicine, University of Oxford; 2018. https://www.equator-network.org/toolkits/developing-a-reporting-guideline/.

12. Zoratti MJ, et al. Evaluating the conduct and application of health utility studies: a review of critical appraisal tools and reporting checklists. Eur J Health Econ. 2021;22(5):723–33. https://doi.org/10.1007/s10198-021-01286-0.

13. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. Health Technol Assess. 1999;3(9):174–84. https://doi.org/10.3310/hta3090.

14. Stalmeier PF, Goldstein MK, Holmes AM, Lenert L, Miyamoto GW, Stiggelbout J, et al. What should be reported in a methods section on utility assessment? Med Decis Making. 2001;21(3):200–7.

15. Nerich V, et al. Critical appraisal of health-state utility values used in breast cancer-related cost–utility analyses. Breast Cancer Res Treat. 2017;164(3):527–36. https://doi.org/10.1007/s10549-017-4283-8.

16. Petrou S, et al. Preferred Reporting Items for Studies Mapping onto Preference-Based Outcome Measures: The MAPS Statement. Pharmacoeconomics. 2015;33(10):985–91. https://doi.org/10.1007/s40273-015-0319-2.

17. Husereau D, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. Eur J Health Econ. 2022;23(8):1309–17. https://doi.org/10.1007/s10198-021-01426-6.

18. Prevolnik Rupel V, Ogorevc M, Greiner W, Kreimeier S, Ludwig K, Ramos-Goni JM. EQ-5D-Y Value Set for Slovenia. Pharmacoeconomics. 2021;39(4):463–71. https://doi.org/10.1007/s40273-020-00994-4.

19. Stevens K. Valuation of the child health utility 9D index. Pharmacoeconomics. 2012;30(8):729–47. https://doi.org/10.2165/11599120-000000000-00000.

20. Retzler J, Grand TS, Domdey A, Smith A, Romano Rodriguez M. Utility elicitation in adults and children for allergic rhino-conjunctivitis and associated health states. Qual Life Res. 2018;27(9):2383–91. https://doi.org/10.1007/s11136-018-1910-8.

21. Lloyd A, et al. A valuation of infusion therapy to preserve islet function in type 1 diabetes. Value Health. 2010;13(5):636–42. https://doi.org/10.1111/j.1524-4733.2010.00705.x.

22. Dewilde S, Roudijk B, Tollenaar NH, Ramos-Goñi JM. An EQ-5D-Y-3L Value Set for Belgium. Pharmacoeconomics. 2022;40(S2):169–80. https://doi.org/10.1007/s40273-022-01187-x.

23. Kreimeier S, et al. EQ-5D-Y Value Set for Germany. Pharmacoeconomics. 2022;40(S2):217–29. https://doi.org/10.1007/s40273-022-01143-9.

24. Rencz F, Ruzsa G, Bató A, Yang Z, Finch AP, Brodszky V. Value Set for the EQ-5D-Y-3L in Hungary. Pharmacoeconomics. 2022;40(S2):205–15. https://doi.org/10.1007/s40273-022-01190-2.

25. Roudijk B, Sajjad A, Essers B, Lipman S, Stalmeier P, Finch AP. A Value Set for the EQ-5D-Y-3L in the Netherlands. Pharmacoeconomics. 2022;40(S2):193–203. https://doi.org/10.1007/s40273-022-01192-0.

26. Yang Z, et al. Estimating an EQ-5D-Y-3L Value Set for China. Pharmacoeconomics. 2022;40(S2):147–55. https://doi.org/10.1007/s40273-022-01216-9.

27. Devlin N, et al. 'Valuing EQ-5D-Y: the state of play. Health Qual Life Outcomes. 2022;20:1.

28. Pan T, Mulhern B, Viney R, Norman R, Hanmer J, Devlin N. A Comparison of PROPr and EQ-5D-5L Value Sets. Pharmacoeconomics. 2022;40(3):297–307. https://doi.org/10.1007/s40273-021-01109-3.

29. Lancsar E, Swait J. Reconceptualising the external validity of discrete choice experiments. Pharmacoeconomics. 2014;32(10):951–65. https://doi.org/10.1007/s40273-014-0181-7.

30. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual. Process and Methods. National Institute for Health and Care Excellence; 2022. https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741

Authors and Affiliations

Cate Bailey[1] · Martin Howell[2,3] · Rakhee Raghunandan[2,3] · Kim Dalziel[1] · Kirsten Howard[2,3] · Brendan Mulhern[4] · Stavros Petrou[5] · Donna Rowen[6] · Amber Salisbury[2,3] · Rosalie Viney[4] · Emily Lancsar[7] · Nancy Devlin[1]

✉ Martin Howell
martin.howell@sydney.edu.au

Cate Bailey
cate.bailey@unimelb.edu.au

1   Health Economics Unit, Centre for Health Policy, Melbourne School of Population and Global Health, The University of Melbourne, Level 4, 207 Bouverie St., Carlton, VIC 3053, Australia

2   School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia

3   Menzies Centre for Health Policy and Economics, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia

4   Centre for Health Economics, Research and Evaluation (CHERE), University of Technology Sydney, Sydney, NSW, Australia

5   Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

6   School of Health and Related Research, University of Sheffield, Sheffield, UK

7   Department of Health Services and Policy Research, Research School of Population Health, Australian National University, Acton, ACT, Australia