Paper presented at the Scientific Plenary Meeting of the EuroQol Group, Chicago 2022.
Not to be cited without the authors' consent.

Values for child health related quality of life: a checklist for studies reporting the elicitation of stated preferences

Bailey C[1ψ], Howell M[2ψ], Raghunandan R[2], Dalziel K[1], Howard K[2], Mulhern B[3], Petrou S[4], Rowen D[5], Salisbury A[2], Viney R[3], Lancsar E[6ξ], Devlin N[1ξ]

1. Health Economics Unit, University of Melbourne 2. University of Sydney 3. University of Technology Sydney 4. University of Oxford 5. School of Health and Related Research, University of Sheffield. 6. Australian National University, Canberra.

ψ joint first authors

ξ joint senior authors

Abstract

Background: A systematic review of measures of child Health Related Quality-of-Life (HRQoL) and their accompanying value sets (Kwon et al., 2022) shows widely varying methods for eliciting and modelling value sets for child HRQoL, and very different characteristics of the resulting value sets e.g., some contain no values < 0; others contain very high proportions of negative values. That paper, and a follow-on review of the methods used in valuing pediatric HRQoL to date (Bailey et al., 2022), encountered poor and incomplete reporting in many studies. Checklists can play an important role in improving standards of reporting and in helping users to interpret and assess available values. Existing checklists for HRQoL utilities tend to be focussed on valuation of adult HRQoL instruments, and do not include checklist items addressing the methods issues specific to valuing child HRQoL. Further, existing checklists do not address how users should judge the validity of reported values.

Objectives: Our aim was to develop a checklist for studies generating values for child HRQoL. The values used in cost effectiveness models of pediatric interventions often include values directly elicited for disease-specific states or vignettes for children as well as those from value sets for childhood HRQoL instruments; our checklist was developed with the aim of being able to be applied to either.

Methods: A conceptual model was developed that provides a modular structure for the checklist. The modules are grouped by (i) methods (modules A-D) and (ii) values (module E). A longlist of potential items for each 'method' module was obtained from a recent review of checklists for adult HRQoL values (Zoratti

et al 2021), complemented by generation of additional items specific to child HRQoL values, extracted from recent reviews of the relevant methods literature. Checklist items relating to the characteristics of the 'values' were based on theoretical papers on external validity of stated preference data (e.g., Lancsar and Swait 2014) and papers reporting methods for examining the distribution of 'theoretical' values in value sets (e.g., Pan et al 2021). The long list of items in each module was reduced by eliminating duplication and overlap; and then refined to strengthen relevance and clarity via an iterative process. The resulting checklist was tested by applying it to a range of papers selected from those reported in Kwon et al (2022) and Bailey et al (2022), including recently published EQ-5D-Y value sets.

Results: The resulting checklist contains modules aimed at reporting methods (A-D) and the characteristics of values (E). These modules are populated with a total of 81 items; which modules and sub-set of these items are relevant to use depends on the type of valuation study to which the checklist is applied. Its application to a selection of papers reporting child HRQoL value suggests it is feasible to use. Illustrative examples of its application to an EQ-5D-Y-3L value set and a CHU9D value set are provided.

Conclusions: This is the first checklist for child HRQoL values. Its modular structure means that in principle it can be applied to assessing value sets as well as values generated from other types of studies eliciting values for child health states. The inclusion of items relating to characteristics of values is novel and potentially has broader relevance (e.g., to future checklists for adult utilities). The checklist has the potential to improve completeness in the reporting of pediatric values, as well as helping users compare and assess the characteristics of available value sets. This is work in progress: we plan a consultative process to finalise the checklist, and to improve its useability.

1.0 Introduction

There is currently a lack of consensus about fundamental aspects of the research methods to use in valuing child HRQoL (Devlin, 2022; Rowen et al., 2020). Recent reviews of measures of child HRQoL and their accompanying value sets (Bailey et al., 2022; Kwon et al., 2022) show widely varying methods for eliciting and modelling value sets for child HRQoL, and very different characteristics of the resulting value sets. For instance, while some value sets contain no values for health states that are considered worse than being dead (i.e., have values < 0),others have very high proportions of such health states e.g., the Canadian HUI3 value set has 78% values < 0 (Feeny et al., 2002), and the Slovenia value set for EQ-5D-Y-3L has 21% (Prevolnik Rupel et al., 2021). Differences between the available values for child HRQoL could have implications for conclusions about the effectiveness and cost effectiveness of pediatric treatments.

It is therefore crucial that those choosing which child HRQoL values to use to estimate quality adjusted life years (QALYs) for use in economic evaluation - and those using that evidence in decision making - are

aware of the characteristics of the values and how differences between the characteristics of values (whether generated from value sets or vignettes) might affect and limit the comparability of evidence on HRQoL and QALYs in children. However, both Kwon et al (2022) and a review of the methods used to value child HRQoL (Bailey et al., 2022) concluded that the reporting of such studies is often incomplete and inconsistent. Poor reporting of values for child HRQoL makes it difficult for users to make informed choices, and for decision makers to use the evidence in an informed way. 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 (Bailey et al., 2022).

Various checklists are available for reporting the HRQoL values available for use in estimating quality QALYs e.g., CREATE (Xie et al., 2015), and SpRUCE (Brazier et al., 2019). However, these checklists focus on the values for instruments for measuring *adult* HRQoL. All the methodological considerations that arise in producing values for adult HRQoL *also* apply to childhood HRQoL. However, there are additional methodological considerations which arise in valuing childhood HRQoL, which do *not* arise in valuing adult HRQoL. These considerations relate to fundamental aspects of the valuation task; for instance, whose stated preferences are considered relevant in valuing child HRQoL? If adult members of the general public are invited to value child health states, from what perspective are they asked to imagine the states they are requested to value? The choice of duration of the state to be valued is an issue which arises in valuing adult HRQoL; however, this choice interacts with other methods choices specific in valuation of child HRQoL (such as the age of hypothetical child to be considered) in complex ways. For example, the 'life' in the state to be evaluated might comprise a mix of years in childhood and years in young adulthood. As such, existing checklists do not provide an adequate basis for guiding the reporting and assessment of values and 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 is important because, as Kwon et al (2022) note, "Diverse preference elicitation methods were used to elicit values. Practices with respect to anchoring values on the utility scale also varied considerably. *The range and distribution of values reflect these differences, resulting in value sets with notably different properties*."

The aim of this study was to develop a reporting checklist to support the comprehensive reporting of methods and results from studies of values for childhood HRQoL. The resulting checklist will be applicable

to a broad range of studies that aim to produce values for childhood HRQoL. Here, we include studies that produce value algorithms ('value sets') for childhood HRQoL instruments (both generic and disease-specific) as well as studies that seek to produce values for a limited number of specific states e.g., described by vignettes, or a selection of states from a disease-specific pediatric PROM, perhaps motivated by the states for which values are needed in cost-effectiveness modelling. We included these types of studies as they have often been used in cost-effectiveness models in Australian Pharmaceutical Benefits Advisory Committee PBAC decisions (Bailey et al., 2021). The availability of a checklist for reporting child HRQoL values will (a) improve the information available for users choosing which values, or HRQOL measures accompanied by values, to use in paediatric populations and (b) improve the standards of reporting about values for child HRQoL. Improving the reporting of child HRQoL values will allow users to evaluate and compare the results from studies; and decision makers to better understand the sources of values and the implications of differences in values for interpretation of cost-effectiveness evidence.

2.0 Methods

2.1 Developing a conceptual framework to provide a foundation for the checklist.

A conceptual framework was developed by the authors (ND, EL, CB) with the aims of the checklist in mind i.e., that it should be capable of being used to assess values for child HRQoL, whether they be for individual states (e.g., described via vignettes or other means) or are value sets reporting values for all states described by a given childhood HRQoL instrument. Given the differences between these study types, a modular approach was proposed to allow sufficient flexibility to apply the checklist 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. An initial conceptual framework was developed in a brainstorming exercise to identify the relevant modules. This was refined in the light of subsequent checklist item development and testing, via an iterative process.

2.2 Establishing a long list of items for each module

A review of checklist items for reporting values for adult HRQoL (Zoratti et al., 2021) was used as a source of the items common to both adult and child HRQoL. Two sets of checklists had been included in the Zoratti study: 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 Zoratti et al. 2021). Items from the latter were relevant for this stage, and included Table 7 from Brazier et al. (1999), Table 8 from Stalmeier et al (2001), Table 10 from CREATE (Xie et al., 2015), Table 11 from Nerich et al (2017) and Table 12 from SpRUCE (Brazier et al., 2019). We did not code Table 9 – MAPS (Petrou et al., 2015), as that checklist is only relevant to studies mapping across instruments and thus outside the scope of our work. Items from the included checklists provided a starting point for the development of a long list of items. Items were grouped by the relevant modules in our conceptual framework by two members of the team (CB and RR) independently and reviewed by EL and ND.

As noted in the Introduction, there are currently no checklists available for reporting values for child HRQoL. Therefore, we supplemented the long list derived from Zoratti et al (2021) with further items specific to valuation of child HRQoL, with item generation based on (a) methods issues relating to valuation of child HRQoL identified by Rowen et al., (2020) and (b) information from two systematic reviews (Bailey et al., 2022; Kwon et al., 2022) on what aspects of methods were specific to valuation of pediatric HRQoL and what, in practise, was missing or unclear from papers reporting values for child HRQoL. Combined, these sources yielded a long list of candidate items under each module. The original long list of items, and subsequent versions created through the review process described in the following section, are available from the authors on request.

2.3 Creating a short list of items for each module

A series of five workshops were held with a sub-set of the co-authors (CB, MH, ND, EL, RV, RR) where the long-list of items in each module were each carefully considered, with the objective of eliminating redundancy/overlap and to check for relevance. Where gaps were identified, new items were created, or wording clarified. Changes arose most often in the items specific to child HRQoL. This collaborative and iterative process led to the creation of an initial draft checklist of 147 items grouped by the 5 modules.

We had originally anticipated that the reduced long list of items we would produce at this point would require substantial further efforts to prioritise items to make it more concise and had initially planned use of best worst scaling to achieve that. However, the process of eliminating redundant items and checking relevance yielded a first draft short list that was considered potentially useable. During this process, the conceptual model was also 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. This information was then compiled, and the checklist items were edited accordingly.

2.4 Testing and validating the checklist

We aimed to test how well the checklist performed on a sample of studies as a comprehensive yet practical means of aiding the reporting of values for child HRQoL. To achieve this, we selected studies that had been included in our systematic review (Bailey et al., 2022) which covered both value sets and vignette studies, to test the checklist performance on both these study types. Two value set papers (Prevolnik Rupel et al., 2021; Stevens, 2012) were chosen to test a selection of the principal generic child HRQoL measures (known to represent contrasting methods) and two examples of studies generating values for other types of descriptions of child HRQoL (Lloyd et al., 2010; Retzler et al., 2018). Together, these 4 studies allowed us to test the checklist on a range of study types that the checklist aimed to cover, in papers published between 2010 and 2021. In each case, two members of the authorship team independently used the checklist to review and summarise the study. These reviews were then compared and reported to the wider study team for discussion, and refinements to the checklist identified and implemented. This process yielded the checklist presented in this paper.

3.0 Results

3.1 Conceptual model

The conceptual model is shown in Figure 1. The checklist is structured using five 'top level' headline groupings (modules) of items, summarized in Box. 1. 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 empirical characteristics of the values themselves and how we might judge their validity.

Box 1. Summary of checklist modules.

A: *Whose* stated preferences were used to value child HRQoL?

B: What child HRQoL *states* were valued?

C: Which methods were used to *elicit* stated preferences for child HRQoL?

D: How were stated preferences for child HRQoL *modelled* to create a value set?

E. What are the characteristics of the resulting *values* for child HRQoL?

The modules are not necessarily hierarchical since decisions relevant to some modules are made simultaneously rather than sequentially and are often iterative. Thus, Figure 1 is presented in a non-hierarchical manner. We also note that there are likely to be interactions between methods decisions in each module (for example, between population and anchoring or method and perspective). The checklist items noted in Figure 1 are provided for illustration only.

Modules A-D: valuation methods

Modules A1 to A3 are specific to considerations relating to pediatric HRQoL values. The checklist items they contain were not derived from any of the existing checklists for adult HRQoL values. Modules B2 & B3 are alternative modules which users select depending on whether the values they are considering are value sets (B2) or values for specific states (B3).

Modules C and A4 are modules with more general methods and sample considerations. These are not necessarily specific to pediatric values but are an important part of what users of pediatric 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 PROMs) but not B2 (direct valuation of disease specific state or vignettes).

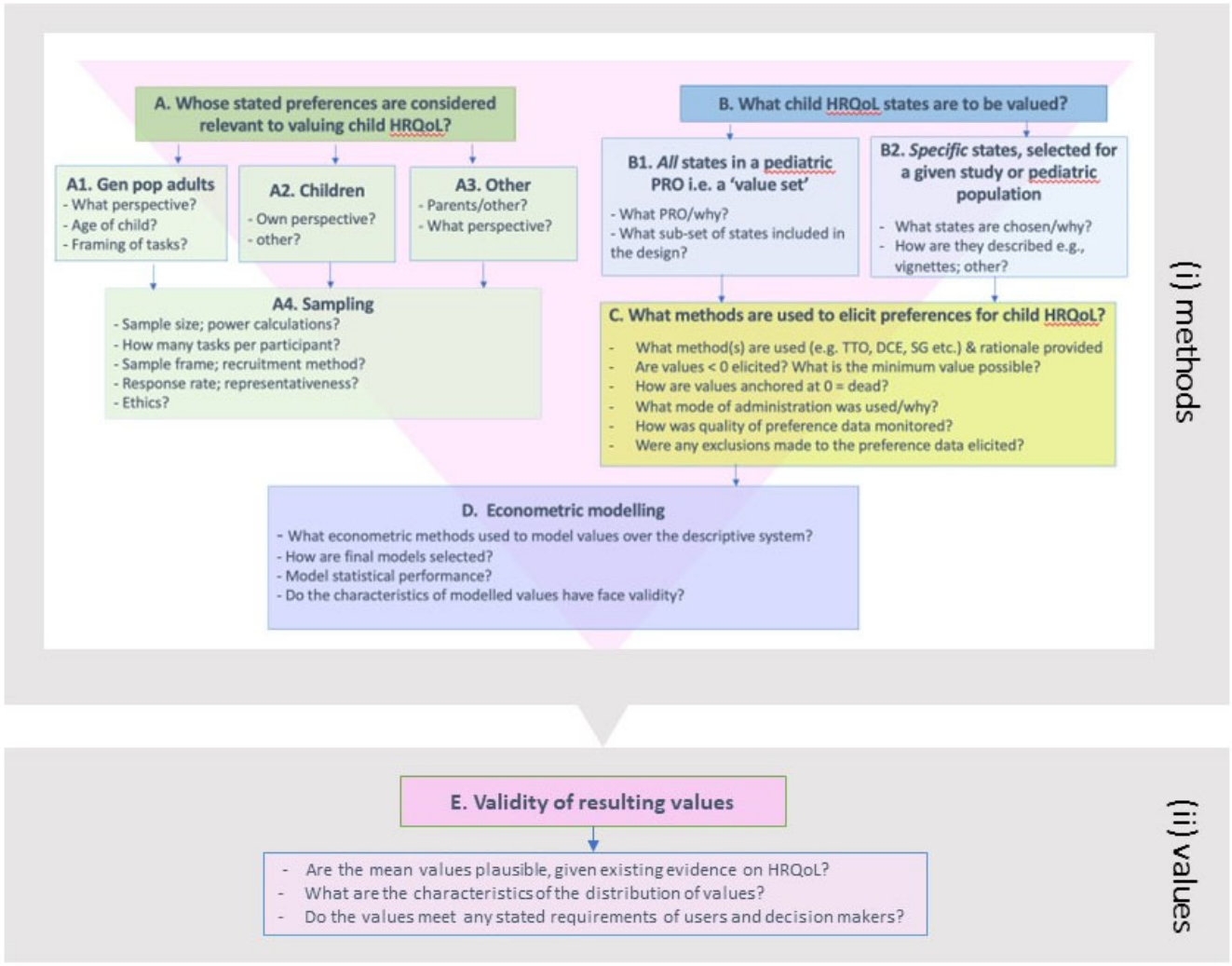

Figure 1. A conceptual framework for a modular checklist for values for child HRQoL.

Module E: characteristics of values

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., similar to the checklist modules A-D described above).

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 reflect a myriad of different methods choices, and motivations, of the instrument developers (Pickles et al., 2019).

For this reason, we considered it important that our checklist include a module focusing on the properties and validity of the values, to ensure users are *aware* of the characteristics of values and the relevant differences in values when choosing between instruments and value sets; to ensure decision makers are aware of the potential implications of these differences when interpreting cost effectiveness evidence based on them; and to encourage more complete reporting of these value characteristics by study teams. However, judging the validity of HRQoL values based on stated preferences data is challenging. We can describe the characteristics of value sets, but the basis for judging the *validity* of those values is much more difficult.

Devlin (2022) 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." 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 derives. We argue here that validity can and should be explored even in the absence of such external benchmark. For example, the characteristics of a new value set might be compared with the characteristics of an existing value set and found to have different properties. 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. However, such comparisons can

nevertheless be helpful and informative in the triangulation and accumulation of evidence – a point we return to below.

There are *some* obvious criteria that can be applied, such as where the object of valuation is a HRQoL instrument. These criteria arise from the properties of HRQoL descriptive systems in such instruments: within these, there are (some) states which 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* criteria of modelled values, but is nevertheless worth checking that value sets (of the type Module B2 is concerned with) have this property. This is less likely to be relevant for values from vignettes (B3).

Lancsar and Swait (2014) 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 process validity is analogously relevant when considering validity in the context of HRQoL values (and values for pediatric HRQoL specifically). Many aspects of process validity are already 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 (QA) (Module C) is important information for users.

A key aspect of process validity not otherwise addressed in modules A-D 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 – e.g. this could include but not be limited to 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 (e.g., whose values are considered relevant; 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 e.g., NICE's recommendation that values are obtained using 'choice-based methods' (Tosh et al., 2011). Further, NICE is currently developing guidelines on methods for measuring and valuing child HRQoL and there is growing awareness of the issues around

pediatric 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.

3.2 Testing the checklist: an application to four studies of child HRQoL values

We selected four papers to review the checklist against. Two papers were on value sets, for the EQ-5D-Y-3L (Prevolnik Rupel et al., 2021) and the CHU9D (Stevens, 2012). Two papers used vignettes (Lloyd et al., 2010; Retzler et al., 2018). The joint first authors reviewed the EQ-5D-Y paper and used this process alongside the other reviewed papers (reviewed by RR, KD, and CB) to make edits to the checklist. Most of the changes involved clarification of wording and removal of duplication. Modules A-D performed well, and Module E required more discussion and adjustment. A summary of the reviews of value sets for EQ-5D-Y and CHU9D is provided for illustrative purposes in Appendix 1.

3.3 Summary of checklist items in each module

The resulting checklist contains modules aimed at reporting methods (A-D) and the characteristics of values (E), populated with a total of 81 items, shown in Table 1. Appendix 2 contains the items accompanied by brief explanations of the rationale for inclusion of each. Appendix 3 contains a summary of the number of items in each module. Note that, because of the modular structure, not all items are relevant to all valuation studies.

Table 1 Checklist items

No.	Item
	Section A – Whose stated preferences were considered relevant to valuing child HRQoL
A1	Whose stated preferences were sought? *Adults – go to A2; Children go to A3; Mixed adult and adolescent: complete both A2 and A3*
A2	**Adult stated preferences**
A2a	Which adults were the focus of preference elicitation?
A2b	What was the rationale for the choice of whose stated preferences were sought?
A2c	What perspective was the adult asked to take in considering the child states to be valued?
A2d	Were adults asked to act as a proxy for a child?
A2e	Was the age of child, for whom respondents were asked to imagine health states to be valued, specified?
A3	**Children's stated preferences**
A3a	Which children were the focus of preference elicitation?
A3b	What perspective was the (child) respondent asked to take?
A3c	If not 'themselves'/own perspective, was the age of child stated?
A4	**Sample**
A4a	What was the population from which the sample was drawn?
A4b	How was the sample frame defined?
A4c	Did the sample frame match the intended population?
A4d	How was the sample recruited?
A4e	Was the sampling method stated and justified?
A4f	What was the target sample size?
A4g	What was the justification for the sample size?
A4h	What sample size was achieved?
A4i	In what year were the data collected?
A4j	Were any of the stated preference data excluded from analysis/ value set modelling?
A4k	Were there any missing data (non-completion, withdrawals)?
A4l	How were missing data addressed?
A4m	Were characteristics of respondents included in the analysis described?
A4n	Were the background characteristics of the respondents whose data were used in analysis representative of the intended population?
	Section B – What child HRQoL states were valued?
B1	Did the values reported in this paper comprise: *A value set, Go to B2; values for a limited number of health states (vignette or condition-specific)? Go to B3*
B2	**Value sets**
B2a	Which HRQoL instrument was valued?

B2b	Were the dimensions of the instrument described?
B2c	Was the number of levels in each dimension (domain) of the instrument described?
B2d	What was the subset of states for which preferences were elicited?
B3	**Specific health states**
B3a	How were the health states described?
B3b	For how many health states were preferences elicited?
B3c	What was the rationale for the selection of these specific states?
B3d	Were the method(s) used in assigning the health states to respondents stated?
	Section C - What methods were used to elicit stated preferences for child HRQoL?
C1	Which methods were used to elicit stated preferences? (*Options: DCE; TTO; SB; BWS; VAS; Other please specify).*
C2	If more than one method was used, what was the intention?
C3	Was a rationale for the choice of method provided?
C4	Was the duration of the states to be valued reported?
C5	How were the survival durations characterized?
C6	Did the methods allow values to be elicited that were < 0 ('worse than dead')?
C6a	How are these values elicited?
C6b	What was the minimum value possible?
C6c	If TTO/SG, was the task description adequate?
C7	How were values anchored at 0 = dead?
C8	What experimental design approach was used to choose the health states (combination of dimension levels) to be valued?
C8a	Were interactions included?
C9	What was the mode of administration for the stated preference tasks?
C10	How was the quality of stated preference data assessed?
C11	Were any exclusions made to the preference data?
C11a	Were reasons for excluding any respondents or observation provided?
C11	Was the order in which the health states were presented to respondents in valuation tasks randomized?
	Ethics and Conflict of interest
C12	Was ethics consent for the study obtained?
C13	Were sources of funding and non-monetary support and the role of the funder(s) in the design described?
C14	Was conflict of interest reported?
	Section D – Analysis
D1	Did the values reported in this paper comprise: *A value set, Go to D2; OR values for a limited number of health states (vignette or condition-specific)? Go to D3*
D2	**Econometric modelling of value sets for HRQoL instruments**
D2a	What was the theoretical model?
D2b	What were the main assumptions of this model?
D2c	How was the constant term treated?

D2d	Were interaction terms included?
D2e	Were non-linear specifications considered?
D2f	Were covariates/confounders included?
D2g	Was more than one model described?
D2h	Were goodness-of-fit statistics for each model reported?
D2i	Was the preferred model clearly stated?
D2j	How was the final model was selected?
D2k	Do the preference parameters for the health states follow a logical order (monotonic)?
D2l	Was any post estimation undertaken to force monotonicity?
D2m	How were insignificant differences between adjacent levels managed?
D2n	What robustness checks were conducted?
D2o	How is uncertainty around the values reported?
D3	**Analysis of values for specific HRQoL states**
D3a	Have the statistical methods been described?
D3b	Have statistical methods been justified?
D3c	Have subgroup analyses and interactions been undertaken?
D3e	Were sub-groups and interaction variable chosen for assessment justified?
D3f	Were sensitivity analyses undertaken and described?
	Section E Characteristics and validity of values
E1	Was there evidence the respondents engaged with and understood the valuation tasks?
E2	Were the mean values plausible, compared to values for the same or other relevant states?
E3	Where a value was reported, were the values generated by the final model logically consistent?
E4	Did authors report the distribution of values over all states defined by the HRQoL instrument?
E5	What were the key characteristics of the values?
E5a	What percentage of values were < 0?
E5b	What was the maximum value <1?
E6	As levels worsen in any one item, holding all other items constant, did values decrease in a linear or non-linear manner?
E7	What was the order of dimension (domain) importance suggested by the value set?
E8	Was uncertainty around the estimated values reported?
E9	Do the values meet any stated 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.

4.0 Discussion

This paper reports the first checklist for studies reporting values for child HRQoL. There has been a notable increase in research aimed at producing values for child HRQoL in recent years – for the EQ-5D-Y-3L alone, 17 country value sets have commenced or been completed since 2020 (Devlin et al., 2022). But the methods being used to value EQ-5D-Y and other pediatric HRQoL instruments vary widely (Bailey et al., 2022; Kwon et al., 2022) and the resulting values for child HRQoL also have very different characteristics. The availability of a checklist is hoped to allow users to better understand and be aware of the implications of these differences when choosing between available values. The better that results of HRQoL studies are reported, the more informed choices users of such results can make, in turn potentially increasing the usefulness of such results in decision making.

The checklist is also useful for those designing and reporting studies of values for child HRQoL, to encourage more complete and consistent reporting of methods and results. While our objective was to develop a reporting checklist for HRQoL to help fill a key gap, we also expect the checklist could prove helpful for those developing or using adult HRQoL values as a number of modules – especially Module E – are arguably relevant to HRQoL measures more generally.

The conceptual framework and selection of modules was based on the combined expert opinion of the authors, across our two research teams (QUOKKA and TORCH)[1] comprising researchers in Australia and the UK. Different ways of grouping and presenting the relevant checklist items would be possible and may be considered more useful in other contexts. Similarly, the process of initially generating the long list of checklist items, and its refinement leading to the checklist items reported in Table 1, reflects our individual and collective experiences and opinions as a group of researchers. There is inevitably a degree of subjectivity and researcher judgement involved in all such checklists. 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. Nevertheless, the checklist reported here could potentially be strengthened by undertaking wider consultation and feedback among the research community and this is planned as a next step.

Similarly, we are mindful of the challenge in striking the right 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

[1] Quality of Life in Kids: Key evidence for Decision Makers in Australia (QUOKKA) and Tools for Outcomes Research to Measure and Value Child Health (TORCH) are each funded by the Medical Research Future Fund (MRFF).

checklists (e.g. CREATE; Xie et al., 2015), 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 of the checklist (e.g., those designing clinical trials; or choosing between available value sets for a given instrument to use in economic evaluation) whose needs may be different and for whom the correct balance between depth and brevity may be different than for researchers reviewing others' (or reporting their own) valuation studies. In ongoing work, we are considering the potential for developing a concise version of this checklist aimed at those who 'demand' value sets and may need only a high-level overview of the methods and results; and 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 and most papers do a good job in reporting 'basic' aspects of this, such as the minimum value and the proportion of negative values in value sets. We additionally suggest (item E4) that authors to supply the distribution of values over all the states defined by the instrument, which is currently not commonly reported. An example of a figure summarising a distribution of 'theoretical' values for an (adult) HRQoL descriptive system can be found in Figure 1 of Pan et al (2022). In our ongoing work, we will produce these figures for the two value sets reported in Appendix 1 and will provide the code to enable others to produce similar figures for other HRQoL instruments.

The checklist we have developed focuses on papers reporting stated preference studies aimed at producing values for child HRQoL. The checklist is intended to be capable of being applied to *any* paper reporting efforts to directly elicit stated preferences for child HRQoL, whether that be to establish mean values for a handful of specific states described by vignettes or disease-specific instruments, or modelled values for all states defined by a generic pediatric HRQoL instrument. However, there are some other aspects of methods used to obtain values for child HRQoL states that are not covered by the checklist. For example, it is not designed to apply to studies that report mapping from a disease specific to a generic instrument as a means of assigning values to disease specific states. Similarly, the checklist does not address research efforts aimed at item reduction of a longer-form instrument to produce a short-form instrument capable of being directly valued. Both mapping and instrument reduction are therefore outside the scope of checklist but may contribute to variations in values available for child HRQoL.

A novel aspect of the checklist is the inclusion of items relating to the characteristics of reported values. Existing checklists 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 a starting point for the further dialogue needed to establish criteria for judging values. This work might include the legitimacy of the process used to generate them, and *ex ante* judgements about their empirical characteristics. Such discussion might also include the extent to which values (and methods used to obtain them) comply with any stated requirements of end users such as local decision-makers - which might perhaps be termed 'context validity'. We plan to consult the research community on Module E content and items as part of our ongoing work.

We welcome comments on any aspect of this paper. We particularly welcome feedback from plenary participants on the following:

1. Does the conceptual model/modular approach taken to the checklist provide a useful and appropriate foundation to this work?

2. With respect to the checklist items in each module: are there any important gaps/omissions? Given the trade-off between comprehensive coverage and the need to be concise for useability, what was your view about the level of detail in this checklist?

3. What further work (eg. consultation/ feedback; testing and application) do you consider is required before this checklist is suitable for publication?

4. Do you agree that checklists should address reporting the characteristics of values, and not just methods to produce them?

References

Bailey, C., Howell, M., Raghunandan, R., Salisbury, A., Chen, G., Coast, J., Craig, J., Devlin, N., Huynh, E., Lancsar, E., Mulhern, B., Norman, R., Petrou, S., Ratcliffe, J., Street, D. J., Howard, K., & Viney, R. (2022). Preference elicitation techniques used in valuing children's health-related quality-of-life: a systematic review. *PharmacoEconomics*, Accepted 11/04/2022.

Bailey, Cate, Dalziel, K., Cronin, P., Devlin, N., & Viney, R. (2021). How are Child-Specific Utility Instruments Used in Decision Making in Australia? A Review of Pharmaceutical Benefits Advisory Committee Public Summary Documents. *PharmacoEconomics*, *40*(2), 157–182. https://doi.org/10.1007/s40273-021-01107-5

Brazier, J., Deverill, M., Green, C., Harper, R., & Booth, A. (1999). A review of the use of health status measures in economic evaluation. *Health Technology Assessment*, *3*(9), 174–184. https://doi.org/10.3310/hta3090

Brazier, John, Ara, R., Azzabi, I., Busschbach, J., Chevrou-Séverac, H., Crawford, B., Cruz, L., Karnon, J., Lloyd, A., Paisley, S., & Pickard, A. S. (2019). Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value in Health*, *22*(3), 267–275. https://doi.org/10.1016/j.jval.2019.01.004

Devlin, N. J. (2022). Valuing Child Health Isn't Child's Play. *Value in Health*, *In press*.

Devlin, N., Pan, T., Stolk, E., Kreimeier, S., Vertsraete, J., Rand, K., & Herdman, M. (2022). Valuing EQ-5D-Y: the state of play. *Health and Quality of Life Outcomes*, *In press*.

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., Denton, M., & Boyle, M. (2002). Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. *Medical Care*, *40*(2), 113–128. https://doi.org/10.1097/00005650-200202000-00006

Kwon, J., Freijser, L., Huynh, E., Howell, M., Chen, G., Khan, K., Daher, S., Roberts, N., Harrison, C., Smith, S., Devlin, N., Howard, K., Lanscar, E., Bailey, C., Craig, J., Dalziel, K., Hayes, A., Mulhern, B., Wong,

G., … Petrou, S. (2022). Systematic Review of Conceptual, Age, Measurement and Valuation Considerations for Generic Multidimensional Childhood Patient-Reported Outcome Measures. In *PharmacoEconomics* (Issue 0123456789). Springer International Publishing. https://doi.org/10.1007/s40273-021-01128-0

Lancsar, E., & Swait, J. (2014). Reconceptualising the External Validity of Discrete Choice Experiments. *PharmacoEconomics*, *32*(10), 951–965. https://doi.org/10.1007/s40273-014-0181-7

Lloyd, A., Swinburn, P., Boye, K. S., Curtis, B., Sarpong, E., Goldsmith, K., Bode, B., & Aronoff, S. (2010). A valuation of infusion therapy to preserve islet function in type 1 diabetes. *Value in Health*, *13*(5), 636–642. https://doi.org/10.1111/j.1524-4733.2010.00705.x

Nerich, V., Saing, S., Gamper, E. M., Holzner, B., Pivot, X., Viney, R., & Kemmler, G. (2017). Critical appraisal of health-state utility values used in breast cancer-related cost–utility analyses. *Breast Cancer Research and Treatment*, *164*(3), 527–536. https://doi.org/10.1007/s10549-017-4283-8

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

Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., & Gray, A. (2015). Preferred Reporting Items for Studies Mapping onto Preference-Based Outcome Measures: The MAPS Statement. *PharmacoEconomics*, *33*(10), 985–991. https://doi.org/10.1007/s40273-015-0319-2

Pickles, K., Lancsar, E., Seymour, J., Parkin, D., Donaldson, C., & Carter, S. M. (2019). Accounts from developers of generic health state utility instruments explain why they produce different QALYs: A qualitative study. *Social Science and Medicine*, *240*(September), 112560. https://doi.org/10.1016/j.socscimed.2019.112560

Prevolnik Rupel, V., Ogorevc, M., Greiner, W., Kreimeier, S., Ludwig, K., & Ramos-Goni, J. M. (2021). EQ-5D-Y Value Set for Slovenia. *PharmacoEconomics*, *39*(4), 463–471. https://doi.org/10.1007/s40273-020-00994-4

Retzler, J., Grand, T. S., Domdey, A., Smith, A., & Romano Rodriguez, M. (2018). Utility elicitation in adults and children for allergic rhinoconjunctivitis and associated health states. *Quality of Life Research*, *27*(9), 2383–2391. https://doi.org/10.1007/s11136-018-1910-8

Rowen, D., Rivero-Arias, O., Devlin, N., & Ratcliffe, J. (2020). 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*, *38*(4), 325–340. https://doi.org/10.1007/s40273-019-00873-7

Stalmeier, P. F., Goldstein, M. K., Holmes, A. M., Lenert, L., Miyamoto J. Stiggelbout, A.M. Torrance, G. W., & Tsevat, J. (2001). What should be reported in a methods section on utility assessment? *Medical Decision Making*, *21*(3), 200–207.

Stevens, K. (2012). Valuation of the child health utility 9D index. *PharmacoEconomics*, *30*(8), 729–747. https://doi.org/10.2165/11599120-000000000-00000

Tosh, J. C., Longworth, L. J., & George, E. (2011). Utility values in National Institute for Health and Clinical Excellence (NICE) Technology Appraisals. *Value in Health*, *14*(1), 102–109.

https://doi.org/10.1016/j.jval.2010.10.015

Xie, F., Pickard, A. S., Krabbe, P. F. M., Revicki, D., Viney, R., Devlin, N., & Feeny, D. (2015). A Checklist for Reporting Valuation Studies of Multi-Attribute Utility-Based Instruments (CREATE). *PharmacoEconomics*, *33*(8), 867–877. https://doi.org/10.1007/s40273-015-0292-9

Zoratti, M. J., Pickard, A. S., Stalmeier, P. F. M., Ollendorf, D., Lloyd, A., Chan, K. K. W., Husereau, D., Brazier, J. E., Krahn, M., Levine, M., Thabane, L., & Xie, F. (2021). Evaluating the conduct and application of health utility studies: a review of critical appraisal tools and reporting checklists. *European Journal of Health Economics*, *22*(5), 723–733. https://doi.org/10.1007/s10198-021-01286-0

Appendices

Appendix 1a

Review of Prevolnik Rupel, 2021 (EQ-5D-Y value set) using the Bailey et al (2022) Checklist.

Paper title: *EQ-5D-Y Value Set for Slovenia*

Section A - Whose stated preferences were considered relevant to valuing child HRQoL?

A1 - Whose stated preferences were sought?

- ☐ [x] Adults: *go to A2*
- ☐ Children: *go to A3*
- ☐ Mixed adult and adolescent: *complete both A2 and A3*

A2 Adults' stated preferences

A2a Which adults were the focus of preference elicitation?

- ☐ [x] General population adults
- ☐ Parents
- ☐ Health care professionals
- ☐ Adult patients
- ☐ Other adults. {please specify} _____

A2b What was the rationale for the choice of whose stated preferences were sought? None specifically stated other than seeking a representative sample of adults in Slovenia. Authors state they were adhering to the International Valuation Protocol for the EQ-5D-Y-3L (Ramos-Gani et al., 2020) (reference 28)

A2c What perspective was the adult 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
- ☐ [x] A hypothetical child
- ☐ Their own health, when they were a child
- ☐ Their own health, as if they were a child now,
- ☐ Their own health, blinded to the states under consideration being specific to children

☐ A patient (e.g. a health professional asked to take the patients' perspective)

☐ Other, please specify _____

A2d Were adult respondents:

☐ explicitly asked to act as a proxy for an imagined child's preferences, or

☐ [x] to indicate their *own* preferences regarding a child's health? _____

☐ other_____

A2e Was the age of child, for whom respondents were asked to imagine health states to be valued, specified? *Yes~~/no/unclear/not applicable~~*

A2f If so, what was the age of child? 10 years

A2g What was the rationale for the choice of the age of child? Prior studies and following the EQ-5D-Y valuation protocol

A3 Children's stated preferences – note: not relevant to the value set reported by Prevolnik-Rupel (2021)

~~A3a Which children were the focus of preference elicitation?~~

~~☐ General population children~~

~~☐ Patients~~

~~☐ Other children [please specify]_____~~

~~A3b What perspective was the (child) 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~~

~~A3c If not 'themselves'/own perspective, was the age of child stated?~~

~~A3d If so, what was the age of child? _____~~

~~A3ee What was the rationale for age of child? _____~~

A4 Sample

A4a What was the population from which the sample was drawn? Slovenian adults

A4b How was the sample frame defined? (e.g., country, age, condition) DCE survey: Country (Slovenia) and representative of the general population (age, sex, statistical region).

For the cTTO interviews: a non-representative sample of adults recruited from one Slovenian region (Primorska).

A4c Did the sample frame match the intended population? The sample of adults in the DCE survey slightly under-represented women aged>70 years in east Slovenia and slightly over-represented men in the same age group residing in the west Slovenian region. All other groups were well represented. The sample of adults in the cTTO survey was not representative of the Slovenian population but was not designed to be.

A4d How was the sample recruited? (e.g., doorknocking, geographical, online panel, convenience sample)? Online panel for DCE and unclear for cTTO.

A4e Was the sampling method stated and justified? *Yes*/~~*No*~~ The authors state and justify their sampling methods, but whether the unrepresentative cTTO sample is justified in an objective sense depends on whether the International Valuation Protocol for the EQ-5D-Y-3L permits the use of unrepresentative samples for the derivation of the anchor.

A4f What was the target sample size? Stated as 1276 for the DCE and 200 for the cTTO. No justification was provided although this study was following the protocol for valuation of EQ-5D-

A4g What was the justification for the sample size (or sample sizes if by block – e.g. number of tasks per block (e.g. DCE) or health state (e.g. TTO))? The authors state early on that they adhered to the recommendation of the International Valuation Protocol for the EQ-5D-Y-3L; they don't repeat this when discussing sample sizes.

A4h What sample size was achieved? 1074 for the DCE and 202 for the cTTO. Not all data met the quality control criteria

A4i In what year were the data collected? Nov 2019 to Feb 2020

A4j Were any of the stated preference data excluded from analysis/ value set modelling? Yes~~/no/unclear.~~ *If yes, was a rationale for exclusion provided?* "The participants were not included in the analysis if they failed two or more of the three QC tasks. Additionally, participants were excluded if the minimum amount of time spent on all the DCE tasks was less than 150 s" QC checks took the form of rationality questions where one health state was logically dominant.

A4k Were there any missing data (non-completion, withdrawals)? ~~*Yes/No/*~~Unclear

89% completion for DCE and 96% for TTO after excluding per data quality.

A4l If so, describe how missing data were handled ((e.g.: imputation, complete case analysis). No specific handling of missing data.

A4m Characteristics of respondents included in the analysis were described (following any exclusions of data points) *Yes*~~*/No*~~ – however no data on those who did not meet the quality criteria

A4n Were the background characteristics of the respondents whose data were used in analysis representative of the intended population? *Yes*~~*/No/not reported*~~ This is shown graphically and described in the text – some deviation

Section B - What child HRQoL states were valued?

B1 Did the values reported in this paper comprise:

☐ [x] a value set? *Go to B2, or*

☐ values for a limited number of health states (vignette or condition-specific)? *Go to B3*

B2 Value sets

B2a Which HRQoL instrument was valued? *[please state]*_EQ-5D-Y-3L

B2b Were the dimensions of the instrument described? Yes/~~*No*~~

B2c Was the number of levels in each dimension (domain) of the instrument described? Yes~~*/No*~~

B2d What was the subset of states for which preferences were elicited? randomly selected 150 pairs that maximised the Fisher information matrix

B3 Specific health states - note: not relevant to the value set reported by Prevolnik-Rupel (2021)

~~B3a How were the health states described?~~

~~☐ Disease specific vignettes~~

~~☐ By a disease-specific HRQoL instrument~~

~~☐ Other [please specify]_____~~

~~_____~~

~~B3b For how many health states were preferences elicited _____~~

~~B3c What was the rationale for the selection of these specific states? [please specify]_____~~

~~B3d Method(s) used in assigning the health states to respondents were stated *Yes/No*~~

Section C - What methods were used to elicit stated preferences for child HRQoL?

C1 Which methods were used to elicit stated preferences? [If more than one method was used to obtain stated preferences, tick whichever boxes apply)

☐ [x] DCE

☐ [x] TTO

☐ SG

☐ BWS

☐ VAS

☐ Other [please specify] _____

C2 If more than one method was used, was the intention to:

☐ [x] use of data from *both* methods to establish values (e.g. via hybrid modelling/using TTO to anchor DCE) specifically cTTO used to anchor DCE to 0 to 1 *or*

☐ to allow values from each to be generated separately and compared, *or*

☐ other [please specify] _____

C3 Was a rationale for the choice of method provided? Yes~~/*No*~~

If so, what was the rationale?_ Complying with The International Valuation Protocol for the EQ-5D-Y-3L

C4. Was the duration of the states to be valued reported? Yes/~~no~~

C4a If so, was it fixed? Yes/~~No~~

C4b what duration(s) were used? __10 years_____

C5 How were the survival durations characterized (e.g 'x years in this state, followed by death') Has to be implied – BTD states for TTO the choice was 10 years full health then death. The WTD choices is between x years in full health and fixed life of 10 years in full health followed by 10 years in the health state.

C6 Did the methods allow values to be elicited that were < 0 ('worse than dead')? Yes/~~no~~

[If 'no', go to C8]

C6a If yes, how were these values elicited? Using a lead time TTO, which is part of the composite TTO approach (cTTO)

C6b What was the minimum value possible? -1

C6c If TTO/SG, was the task description adequate? (E.g. What determined how the task was terminated?) Yes (the tasks were not actually described in the paper, but rather referenced the EQ protocol)

C7 How were values anchored at 0 = dead? Using cTTO; All variable dummy coded and DCE coefficients divided by the overall utility range and re-scaled to the value of the pits state (33333) obtained from cTTO.

C8 What experimental design approach was used to choose the health states (combination of dimension levels) to be valued? Followed EQ protocol for valuing EQ-5D-Y. Health states were selected at random to maximise the Fisher matrix.

C8a. Were interactions included? *Yes/~~no~~ The* model specification incorporates all 40 two-way interaction parameters.

C9 What was the mode of administration for the stated preference tasks?

- ☐ [x] Online self-completion by the respondent (Note: DCE only)
- ☐ Self-completion of mailed questionnaires
- ☐ Online computer assisted personal interview (CAPI)
- ☐ In person CAPI
- ☐ [x] In person interview (note: cTTO only)
- ☐ Other

C10 How was the quality of stated preference data assessed? DCE Included 3 rationality questions ~~for the DCE~~. Four criteria were identified for cTTO QC – with interview data discarded if one was met. These questions included: 1. No explanation of the 'worse than dead' task. 2 Not enough time spent on wheelchair example. 3 Inconsistency - 33333 not the lowest and at least 0.5 higher than state with lowest value. 4. Not enough time spent on the cTTO task.

C11 Were any exclusions made to the preference data (eg used to represent average preferences)? Yes~~/no/unclear~~ _____

C11a If yes, were reasons for excluding any respondents or observation provided? Yes~~/no/unclear~~ See C10

C12 Was the order in which the health states were presented to respondents in valuation tasks randomized? ~~Yes/No/~~unclear

Ethics and Conflict of interest

C12 Was ethics consent for the study obtained? Yes~~/no/not stated~~

C13 Were sources of funding and non-monetary support and the role of the funder(s) in the design described? *Yes*~~/No~~

C14 Was conflict of interest reported? Yes~~/no~~

Section D – Analysis

D1 Did the values reported in this paper comprise:

- ☐ [x] a value set? *Go to D2, or*

☐ values for a limited number of health states (vignette or condition-specific)? *Go to D3*

D2 - Econometric modelling of value sets for HRQoL instruments

D2a What was the theoretical model? Random utility model – Linear additive utility with all variables dummy coded

D2b What were the main assumptions of this model? (e.g. assumptions about preference homogeneity/heterogeneity) Assumed that mixed logit model was associated with a better fit than MNL based on previous studies

D2c How was the constant term treated (if included)? The authors state that for the cTTO exercise, they included only the constant as the regressor on the data for the pits state.

D2d Were interaction terms included? No

D2e Were non-linear specifications considered? No

D2f Were covariates/confounders included? No

D2g Was more than one model described? ~~*Yes/*~~No

If yes,

D2h Were goodness-of-fit statistics for each model reported? Only one model reported but statistics were reported in Table 2

D2i Was the preferred model clearly stated? Not relevant only one

D2j How was the final model selected? e.g. what criteria were used to make this choice? Not applicable

D2k Do the preference parameters for the health states follow a logical order (monotonic)? ~~*Yes/*~~no

If no:

D2l Was any post estimation undertaken to force monotonicity (e.g. collapsing levels) ~~*Yes/no/*~~ *not stated*

D2m How were insignificant differences between adjacent levels managed (collapsed/ forced to be different)? Not clear as differences between coefficients not presented would need to calculate from Table 2 using the SEs

D2n What robustness checks were conducted All coefficients were significant

D2o How is uncertainty around the values reported? Standard errors

D3 – Analysis of values for specific HRQoL states note: not relevant to the value set reported by Prevolnik-Rupel (2021)

~~D3a Have the statistical methods been described? *Yes/no*~~

~~D3b Have statistical methods been justified? *Yes/no*~~

~~D3d Have subgroup analyses and interactions been undertaken? Y *Yes/no*~~

~~*If yes*~~

~~D3e Were the statistical methods described? *Yes/no* and justified *Yes/no*?~~

~~D3f Were sub-groups and interaction variable chosen for assessment justified? *Yes/no*~~

~~D3f Were sensitivity analyses undertaken *Yes/no*? and described *Yes/no/not applicable*?~~

Section E Characteristics and validity of values

E1 Was there evidence the respondents engaged with and understood the valuation tasks? Yes~~*/no*~~
The evidence comes from the QC tasks

E2 Were the mean values plausible, compared to values for the same or other relevant states? (face validity) Yes with reference to other value sets

E3 Where a value was reported, were the values generated by the final model logically consistent? Yes

E4 Did authors report the distribution of values over all states defined by the HRQoL instrument? No, but they provided plots of predictive ability.

E5 What were the key characteristics of the values?. *For example,*

- E5a what % values were < 0? 50 health states – 20.6%
- E5b What was the maximum value < 1? 0.962
- E5c Where in the descriptive system did the biggest changes in values occur, when shifting between adjacent states? Unclear, but possibly the shift from 33333 to 32333.

E6 As levels worsen in any one item, holding all other items constant, did values decrease in a linear or non-linear manner? Hard to tell from the data without plotting or additional analysis

E7 What was the order of dimension (domain) importance suggested by the value set? The most important health dimension was pain/discomfort, followed by anxiety/depression, usual activities, mobility and self-care (page 467).

E8 Was uncertainty around the estimated values reported? Yes~~/no~~

E9 Do the values meet any stated 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. Not stated

Appendix 1b

Review of *Stevens 2012 (CHU9D value set)* using the Bailey et al (2022) Checklist.

Paper title: *Values for child health related quality of life; a guideline for studies reporting the elicitation of stated preferences.*

Section A - Whose stated preferences were considered relevant to valuing child HRQoL?

A1 - Whose stated preferences were sought?

- ☐ [x] Adults: *go to A2*
- ☐ Children: *go to A3*
- ☐ Mixed adult and adolescent: *complete both A2 and A3*

A2 Adults' stated preferences

A2a Which adults were the focus of preference elicitation?

- ☐ [x] General population adults
- ☐ Parents
- ☐ Health care professionals
- ☐ Adult patients
- ☐ Other adults. {please specify}

A2b What was the rationale for the choice of whose stated preferences were sought? As per NICE recommendations

A2c What perspective was the adult 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, when they were a child
- ☐ Their own health, as if they were a child now,

- ☐ [x] Their own health, blinded to the states under consideration being specific to children. "The perspective was chosen to be simple and the respondent was asked to imagine themselves in this health state for the rest of their lives."
- ☐ A patient (e.g. a health professional asked to take the patients' perspective)
- ☐ Other, please specify

A2d Were adult respondents:

- ☐ explicitly asked to act as a proxy for an imagined child's preferences, or
- ☐ to indicate their *own* preferences regarding a child's health? _____
- ☐ [x] other See A2c

A2e Was the age of child, for whom respondents were asked to imagine health states to be valued, specified? ~~*Yes/no/not clear*~~*/not applicable*

A2f If so, what was the age of child?

A2g What was the rationale for the choice of the age of child?

A3 Children's stated preferences note: not relevant to the value set reported by Stevens et al (2012)

~~A3a Which children were the focus of preference elicitation?~~

- ~~☐ General population children~~
- ~~☐ Patients~~
- ~~☐ Other children [please specify]_____~~

~~A3b What perspective was the (child) 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~~

~~A3c If not 'themselves'/own perspective, was the age of child stated?~~

~~A3d If so, what was the age of child? _____~~

~~A3ee What was the rationale for age of child? _____~~
~~_____~~

A4 Sample

A4a What was the population from which the sample was drawn? General public (adults) UK (Sheffield and Huddersfield).

A4b How was the sample frame defined? (e.g., country, age, condition) See A4a Random sample (street)

A4c Did the sample frame match the intended population? Yes – although no comparison with general population provided except for affluence level.

A4d How was the sample recruited? (e.g., doorknocking, geographical, online panel, convenience sample)? Software used to randomly select street addresses – then posted invitation followed by door knocking at the sampled addresses.

A4e Was the sampling method stated and justified? *Yes*/~~*No*~~ Yes based on practical limitations and prior evidence for geographic variation.

A4f What was the target sample size? 300.

A4g What was the justification for the sample size (or sample sizes if by block – e.g. number of tasks per block (e.g. DCE) or health state (e.g. TTO))? Based on what was achievable with the resources available.

A4h What sample size was achieved? 300 (from 1245 addresses) 282 used in final analysis

A4i In what year were the data collected? Not stated

A4j Were any of the stated preference data excluded from analysis/ value set modelling? Yes*/no/unclear. If yes, was a rationale for exclusion provided?*

A4k Were there any missing data (non-completion, withdrawals)? *Yes*/~~*No*/Unclear~~

A4l If so, describe how missing data were handled ((e.g.: imputation, complete case analysis). Complete case analysis

A4m Characteristics of respondents included in the analysis were described (following any exclusions of data points) *Yes/~~No~~* , also provides comparison with excluded.

A4n Were the background characteristics of the respondents whose data were used in analysis representative of the intended population? ~~*Yes/No/*~~ not reported, although the sample was a random selection from a defined area

Section B - What child HRQoL states were valued?

B1 Did the values reported in this paper comprise:

☐ [x] a value set? *Go to B2, or*

☐ values for a limited number of health states (vignette or condition-specific)? *Go to B3*

B2 Value sets

B2a Which HRQoL instrument was valued? *[please state]* CHU9D

B2b Were the dimensions of the instrument described? Yes/~~*No*~~

B2c Was the number of levels in each dimension (domain) of the instrument described? Yes~~*/No*~~

B2d What was the subset of states for which preferences were elicited? 64 out of 1,953,125

B3 Specific health states note: not relevant to the value set reported by Stevens et al (2012)

~~B3a How were the health states described?~~

~~☐ Disease specific vignettes~~

~~☐ By a disease-specific HRQoL instrument~~

~~☐ Other [please specify]~~ ~~_____~~

~~_____~~

~~B3b For how many health states were preferences elicited _____~~

~~B3c What was the rationale for the selection of these specific states? [please specify]_____~~

~~B3d Method(s) used in assigning the health states to respondents were stated *Yes/No*~~

Section C - What methods were used to elicit stated preferences for child HRQoL?

C1 Which methods were used to elicit stated preferences? [If more than one method was used to obtain stated preferences, tick whichever boxes apply)

- ☐ DCE
- ☐ TTO
- ☐ [x] SG
- ☐ BWS
- ☐ VAS
- ☐ Other [please specify] _____

C2 If more than one method was used, was the intention to: n/a

- ☐ use of data from *both* methods to establish values (e.g. via hybrid modelling/using TTO to anchor DCE) *or*
- ☐ to allow values from each to be generated separately and compared, *or*
- ☐ other [please specify] _____

C3 Was a rationale for the choice of method provided? Yes~~/*No*~~

If so, what was the rationale?_ Based on prior valuations for NICE

C4. Was the duration of the states to be valued reported? Yes/~~no~~

C4a If so, was it fixed? Yes~~/*No*~~

C4b what duration(s) "Rest of their lives" – so strictly speaking could be considered not fixed

C5 How were the survival durations characterized (e.g 'x years in this state, followed by death') See C4b

C6 Did the methods allow values to be elicited that were < 0 ('worse than dead')? Yes~~/no~~

[If 'no', go to C8]

C6a If yes, how were these values elicited? Ranking of nine health states against dead. A different SG task was used, "worse than dead form of SG", for states ranked below dead in the warm-up task. This warm-up task asked participants to rank the set of health states in the SG tasks against dead.

C6b What was the minimum value possible? -1

C6c If TTO/SG, was the task description adequate? (E.g. What determined how the task was terminated?) Not clear – it was by interview until point of indifference

C7 How were values anchored at 0 = dead? Using the values from the SG task that are automatically on the 1-0 scale where 0=dead

C8 What experimental design approach was used to choose the health states (combination of dimension levels) to be valued? Orthogonal array with minimum number required to predict all health states (found to be 64) but this included two duplicate states and best state that cannot be valued in the SG task with the upper anchor as state 111111111. Therefore two 'best' states were included with 8 of 9 dimensions at 1 (i.e. no problems) to retain the number of (64) states.

C8a. Were interactions included? ~~*Yes*~~*/no.* Interactions were not included in the design; in D2d they were included in the modelling, but not reported

C9 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
- ☐ [x] In person interview
- ☐ Other

C10 How was the quality of stated preference data assessed? Excluded on basis of 'unusable' and if valued all health states the same.

C11 Were any exclusions made to the preference data (eg used to represent average preferences)? Yes/~~*no/unclear*~~ _____

C10a If yes, were reasons for excluding any respondents or observation provided? Yes/~~*no/unclear*~~ See C10

C12 Was the order in which the health states were presented to respondents in valuation tasks randomized? ~~*Yes/No*~~/unclear

Ethics and Conflict of interest

C12 Was ethics consent for the study obtained? Yes/~~*no/not stated*~~

C13 Were sources of funding and non-monetary support and the role of the funder(s) in the design described? *Yes/*~~No~~

C14 Was conflict of interest reported? ~~*Yes/no/*~~*unclear*

Section D – Analysis

D1 Did the values reported in this paper comprise:

☐ [x] a value set? *Go to D2, or*

☐ values for a limited number of health states (vignette or condition-specific)? *Go to D3*

D2 - Econometric modelling of value sets for HRQoL instruments

D2a What was the theoretical model? Additive model $U_{ij} = g(\beta x_{ij}) + \varepsilon_{ij}$

D2b What were the main assumptions of this model? (e.g. assumptions about preference homogeneity/heterogeneity) OLS, RE and FE if individual effects considered important i.e. g is a linear function (the warm-up rank data was modelled separately)

D2c How was the constant term treated (if included)? Fixed at 1 to give disutility

D2d Were interaction terms included? Yes – 'MOST' value of 1 if a health state had any level 1 and 'LEAST' value of 1 if any had a value of 5 – however not reported as they did not improve the modelling and were not included in the value set

D2e Were non-linear specifications considered? No

D2f Were covariates/confounders included? No (stated that sample size was too small)

D2g Was more than one model described? *Yes*/~~No~~

If yes,

D2h Were goodness-of-fit statistics for each model reported? Yes

D2i Was the preferred model clearly stated? Yes

D2j How was the final model was selected? e.g. what criteria were used to make this choice? MAE

D2k Do the preference parameters for the health states follow a logical order (monotonic)? ~~*Yes*~~/no

If no:

D2l Was any post estimation undertaken to force monotonicity (e.g. collapsing levels) *Yes/no/not stated*

D2m How were insignificant differences between adjacent levels managed (collapsed/ forced to be different)? Adjacent inconsistent levels were collapsed and for levels insignificant at p<0.1. These were undertaken using the general-to-specific approach

D2n What robustness checks were conducted Mean absolute error and root mean square error

D2o How is uncertainty around the values reported? Standard errors

D3 – Analysis of values for specific HRQoL states **note: not relevant to the value set reported by Stevens et al (2012)**

~~D3a Have the statistical methods been described? *Yes/no*~~

~~D3b Have statistical methods been justified? *Yes/no*~~

~~D3d Have subgroup analyses and interactions been undertaken? Y *Yes/no*~~

~~*If yes*~~

~~D3e Were the statistical methods described? *Yes/no* and justified *Yes/no*?~~

~~D3f Were sub-groups and interaction variable chosen for assessment justified? *Yes/no*~~

~~D3f Were sensitivity analyses undertaken *Yes/no*? and described *Yes/no/not applicable*?~~

Section E Characteristics and validity of values

E1 Was there evidence the respondents engaged with and understood the valuation tasks? Yes~~*/no*~~ Reported in Table 1.

E2 Were the mean values plausible, compared to values for the same or other relevant states? (face validity) Yes with reference to rank model (based on ranking exercise)

E3 Where a value was reported, were the values generated by the final model logically consistent? Yes, the final model was logically consistent. In initial models there were inconsistencies requiring additional parsimonious models

E4 Did authors report the distribution of values over all states defined by the HRQoL instrument? No – however they provide plots of predicted utility values for the parsimonious models and OLS models and observed values for the 64 health states.

E5 What were the key characteristics of the values?. *For example,*

E5a what % values were < 0? 23 (0.93%)

E5b What was the maximum value < 1? 0.993 (Table 2)

E5c Where in the descriptive system did the biggest changes in values occur, when shifting between adjacent states? Unclear

E6 As levels worsen in any one item, holding all other items constant, did values decrease in a linear or non-linear manner? Not able to tell from the data without additional analysis

E7 What was the order of dimension (domain) importance suggested by the value set? Not discussed, however greatest disutility was for pain 5 (0.1461) and smallest for Worry 2345 (0.0251) and Sleep 23 (0.028)

E8 Was uncertainty around the estimated values reported? Yes*/no*

E9 Do the values meet any stated 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. Not stated

Appendix 2

Checklist items with comments

No.	Item	Comments
Section A – Whose stated preferences were considered relevant to valuing child HRQoL		
A1	Whose stated preferences were sought?	It needs to be clear if children, adults or both were included as they require different considerations when eliciting preferences.
A2	Adult stated preferences	
A2a	Which adults were the focus of preference elicitation?	This may include the general population or a select group such as parents, patients or health care professionals all of whom may have different preferences, reference points and experience.
A2b	What was the rationale for the choice of whose stated preferences were sought?	Clear statement of rationale is required to ensure that the generalizability or applicability of values are clear.
A2c	What perspective was the adult asked to take in considering the child states to be valued?	This could include their own health as an adult or a child, their child or a hypothetical child etc. The perspective needs to be clear as it may influence stated preferences.
A2d	Were adults asked to act as a proxy for a child?	It should be clear whether adults were asked to act as a proxy for a child or to indicate their own preferences as this would be expected to influence preferences.
A2e	Was the age of child, for whom respondents were asked to imagine health states to be valued, specified?	The age of the child may have a strong influence on values and should be stated and reasons given for the choice.
A3	Children's stated preferences	
A3a	Which children were the focus of preference elicitation?	Answers could include the general population, school children, patients or another specified group all of whom may have different preferences, reference points and experience.
A3b	What perspective was the (child) respondent asked to take?	Children could be asked to consider themselves, another child they know or a hypothetical child. If considering themselves then preferences may be influenced by whether they are patients or from the general population.
A3c	If not 'themselves'/own perspective, was the age of child stated?	The age of the child may have a strong influence on stated preferences and should be clearly described and reasons given for the choice.
A4	Sample	
A4a	What was the population from which the sample was drawn?	Was the population from a single country or region, or setting (e.g. school)? This is critical to understanding applicability of value sets.
A4b	How was the sample frame defined?	The sample could be defined on the basis of geographic region, age, condition or other defining population characteristic. There

		should be a clear rationale and justification for inclusion if it is a convenience sample.
A4c	Did the sample frame match the intended population?	It should be clear that the approach taken would result in a representative sample of the intended population.
A4d	How was the sample recruited?	The recruitment method needs to be clearly stated to enable understanding of possible selection bias or unrepresentative samples. For example, random selection, door knocking across defined area, online panel, convenience samples etc.
A4e	Was the sampling method stated and justified?	As for recruitment a clear statement and justification for selection of the sample in particular exclusion criteria, is needed to understand selection bias and generalizability.
A4f	What was the target sample size?	The sample size needs to be stated as it is important to understanding overall missingness.
A4g	What was the justification for the sample size	Sample size justification needs to be clear if the sample size is related to the valuation method (i.e. minimum sample number, number of tasks required to be completed), the sampling strategy or for pragmatic reasons.
A4h	What sample size was achieved?	This question should include reasons for not achieving the target sample size.
A4i	In what year were the data collected?	This question is needed to ensure there has not been an excessive time between valuation and publication.
A4j	Were any of the stated preference data excluded from analysis/ value set modelling?	This answer needs to be clearly stated and include the basis for exclusion. For example, unrealistically short time to completion, answers to dominance questions etc.
A4k	Were there any missing data (non-completion, withdrawals)?	Missing data should be appropriately categorized, for example partial or non-completions.
A4l	How were missing data addressed?	If complete-case analysis has not been used, then the basis for imputation or use of partial data should be described.
A4m	Were characteristics of respondents included in the analysis described?	There should be sufficient detail on characteristics to enable assessment as per A4n
A4n	Were the background characteristics of the respondents whose data were used in analysis representative of the intended population?	Background characteristics are important when considering generalizability and application of value sets produced.
Section B – What child HRQoL states were valued?		
B1	Is there a value set or values for a limited number of health states?	The distinction here is between studies that have developed a value set for a HRQoL instrument primarily for defining utility value in economic evaluations or similar, versus those defining a value for a specified health condition or health state(s).
B2	Value sets	

B2a	Which HRQoL instrument was valued?	References to development of the instrument should be included.
B2b	Were the dimensions of the instrument described?	Dimensions should be clear without the need to refer back to development studies.
B2c	Was the number of levels in each dimension (domain) of the instrument described?	Levels should be clear without the need to refer back to development studies.
B2d	What was the subset of states for which preferences were elicited?	In most cases it will not be possible to value every health state. Thus, the rationale for selection of the subset should be clear.
B3	Specific health states	
B3a	How were the health states described?	For example, was a vignette used (disease specific or generalized conditions) or selected health states from a HRQoL instrument?
B3b	For how many health states were preferences elicited?	Clearly reported with reasons.
B3c	What was the rationale for the selection of these specific states?	Selection may be limited by the preference elicitation method or the research question and objectives or for pragmatic reasons. Nonetheless it should be linked back to the objectives of the study.
B3d	Were the method(s) used in assigning the health states to respondents stated?	Methods may be random or defined as part of the study design, however they should be clearly described.
Section C – What methods were used to elicit stated preferences for child HRQoL?		
C1	Which methods were used to elicit stated preferences?	Identify all methods used (e.g. DCE, TTO etc.).
C2	If more than one method was used, what was the intention?	The reasons for multiple methods could include hybrid modelling, anchoring, scaling for values worse than death or for comparative purposes.
C3	Was a rationale for the choice of method provided?	Method selection may relate to factors such as the target population (e.g. age of respondents), the number and complexity of the health states, for ethical reasons (avoiding reference to death), or to meet policy requirements.
C4	Was the duration of the states to be valued reported?	The duration may or may not be fixed and should be clearly stated. This is of particular importance in the context of the perspective respondents are asked to take (items A2c and A3bA)
C5	How were the survival durations characterized?	For example, number of years in a health state followed by death.
C6	Did the methods allow values to be elicited that were < 0 ('worse than dead')?	Needs to be clearly stated as this is key to understanding limitations of the values.
C6a	How are these values elicited?	There are multiple approaches that can be taken, such as a ranking exercise to identify health states worse than death followed by alternate elicitation methods.
C6b	What was the minimum value possible?	The minimum value possible will vary with the method used and should be clearly stated.

C6c	If TTO/SG, was the task description adequate?	In particular, what determined how the task was terminated in cases where it proves difficult to reach indifference.
C7	How were values anchored at 0 = dead?	There are many approaches to anchoring including using select responses from adult respondents where children are involved, and valuing select health states using methods such as TTO or SG where DECEs or BWS are used. Or ranking exercises.
C8	What experimental design approach was used to choose the health states (combination of dimension levels) to be valued?	The purpose of C8 and C8a is primarily related to value sets for HRQoL instruments where the large number of health states will require modelling to predict all values.
C8a	Were interactions included?	Interactions such as health states that include one or more 'most' or 'least' scale values may have been considered.
C9	What was the mode of administration for the stated preference tasks?	Mode of administration is of particular importance with more difficult elicitation tasks such as TTO and SG. Irrespective it should be clearly stated and justified.
C10	How was the quality of stated preference data assessed?	Criteria for assessing quality and exclusions should be clearly defined.
C11	Were any exclusions made to the preference data?	For example, to represent average preferences.
C11a	Were reasons for excluding any respondents or observation provided?	This should reference back to the criteria detailed in response to Item C9.
C11	Was the order in which the health states were presented to respondents in valuation tasks randomized?	The potential for bias should be addressed where tasks were unable to be randomized.
C12	Ethics and Conflict of interest	
C12a	Was ethics consent for the study obtained?	If not obtained, then a clear statement of why ethics was not required should be included.
C12b	Were sources of funding and non-monetary support and the role of the funder(s) in the design described?	Check author statements
C12c	Were there potential conflicts of interest of those conducting the valuation research?	Check author statements
Section D – Analysis		
D1	Value set or values for a limited number of health states?	Analytical requirements may vary depending on whether the study objective is to produce a value set for an HRQoL instrument or single or limited number of value sets.
D2	Section D2 - Econometric modelling of value sets for HRQoL instruments	Values sets require application of econometric modelling methods.
D2a	What was the theoretical model?	The basis for modelling preferences should be described and supported by rationale and simple explanation.

D2b	What were the main assumptions of this model?	For example, what assumptions have been included about preference heterogeneity? How was the model estimated?
D2c	How was the constant term treated?	For example the constant may be set at 1 for a disutility model.
D2d	Were interaction terms included?	Interaction terms may be included to explore influence of 'most' and 'least' dimension scores in developing value sets.
D2e	Were non-linear specifications considered?	If a non-linear functional form was considered, then the specifications evaluated should be described.
D2f	Were covariates/confounders included?	When included the effects on value sets should be reported and described.
D2g	Was more than one model described?	Rationale for each model should be given and include criteria for identifying the preferred model.
D2h	Were goodness-of-fit statistics for each model reported?	Goodness-of-fit statistics should be reported for all models and reference back to criteria for model selection.
D2i	Was the preferred model clearly stated?	This question is important for understanding of which model formed the basis of the value set.
D2j	How was the final model was selected?	Was it based solely on goodness of fit criteria, or modelled versus observed or a combination?
D2k	Do the preference parameters for the health states follow a logical order (monotonic)?	Inconsistencies should be clearly described including insignificant parameters.
D2l	Was any post estimation undertaken to force monotonicity?	The collapsing or omitting of levels within dimensions needs to be clearly reported. Where multiple approaches have been taken, the process for selecting the final combination for the value set should be included.
D2m	How were insignificant differences between adjacent levels managed?	As per D2l
D2n	What robustness checks were conducted?	Should be described in methods section and reported in appropriate detail.
D2o	How is uncertainty around the values reported?	Should be described in methods section and reported in appropriate detail in the results section.
D3	Analysis of values for specific HRQoL states	The analytical approach for studies valuing a single or a selection of health states from a HRQoL will vary according to the research question and objective of the study.
D3a	Have the statistical methods been described?	D3a to D3f need to be addressed in order to understand the approaches taken and the limitations of the analyses.
D3b	Have statistical methods been justified?	This needs to be relevant to the type of data and planned analyses.
D3c	Have subgroup analyses or interactions been undertaken?	Sub-group analyses may be undertaken to evaluate differences in preferences/values. Interactions may be relevant where multiple health states are included.

D3e	Were sub-groups or interaction variables chosen for assessment justified?	The reasons for selecting sub-groups and interaction variables needs to be stated. Where sub-group analyses have been undertaken, it should be stated whether these were defined in advance or exploratory.
D3f	Were sensitivity analyses undertaken and described?	Sensitivity analyses may or may not be warranted depending on the objective of the study for example to address confounding or selection bias. If included they should be adequately described and justified.
Section E Characteristics and validity of values		
E1	Was there evidence the respondents engaged with and understood the valuation tasks?	Evidence may include qualitative data from interview or think aloud as part of pilot testing, missingness, time to complete, specific questions aimed assessing the level of understanding, responses to dominant scenarios, and illogical ranking.
E2	Were the mean values plausible, compared to values for the same or other relevant states?	Do the values have face validity based on a priori hypotheses or comparison with reported values for different populations or specific conditions.
E3	Where a value was reported, were the values generated by the final model logically consistent?	Inconsistencies would suggest that the final model is not appropriate for deriving the value set. This needs to be discussed.
E4	Did authors report the distribution of values over all states defined by the HRQoL instrument?	The values across all health states estimated from modelling based on a subset of health states may indicate bimodal or otherwise unexpected/unusual distributions compared to other HRQoL instruments or alternate value sets for the same instrument. Expected that a histogram would be generated to show this
E5	What were the key characteristics of the values?	Where the distribution of values have not been provided, the following may provide an indication of the validity of the value sets. However, this will also be determined by the way in which data have been reported.
E5a	What percentage of values were < 0?	This is in addition to the average values and provides an indication of variability/uncertainty in preferences for values worse than dead.
E5b	What was the maximum value <1?	This is particularly relevant to elicitation methods that cannot value the full health state and rely on a close to full health state.
E5c	Where in the descriptive system did the biggest changes in values occur, when shifting between adjacent states?	This is relevant to understanding the distribution of values and inconsistencies.
E6	As levels worsen in any one item, holding all other items constant, did values decrease in a linear or non-linear manner?	Whether linear or not this is relevant to demonstrating the extent to which a change in a level (from 1 to 2, then 2 to 3, and so on) within an item results in the same proportional drop in utility (for that item) compared to other items.

E7	What was the order of dimension (domain) importance suggested by the value set?	Does this reflect an expectation of the order of importance based on similar domains from other HRQoL or other value sets.
E8	Was uncertainty around the estimated values reported?	How was it estimated?
E9	Do the values meet any stated 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.	This item is intended as an alert that the values need to be suitable to the requirements of the relevant HTA bodies.

Appendix 3

Number of items per section on the Checklist

	NUMBER OF ITEMS	Subgroups	Number per subgroup
A	**23**		
		Pre-question	1
		A2 – Adults'	5
		A3 – Children's	3
		A4 - Sample	14
B	**9**		
		Pre-question	1
		B2 -value sets	4
		B3- Specific health states	4
C	**18**		
		12 question plus 3 sub-questions	15
		Ethics and Conflict of Interest	3
D	**20**		
		Pre-question	1
		D2 – value sets	15
		D3- specific states	4
E	**12**		
		9 question plus 3 sub-questions	12
Total	**82**		