Monograph:
Bailey, C. orcid.org/0000-0001-5030-430X, Howell, M. orcid.org/0000-0001-9740-712X, Raghunandan,, R. orcid.org/0000-0001-7848-6109 et al. (9 more authors) (2023) The RETRIEVE checklist for studies reporting the elicitation of stated preferences for child health related quality of life. Report. ScHARR HEDS Discussion Papers . School of Health and Related Research, University of Sheffield

---

University of Sheffield | Sheffield Centre for Health & Related Research

HEALTH ECONOMICS & DECISION SCIENCE

Discussion Paper Series

HEDS Discussion Paper 23.01

Title: The RETRIEVE checklist for studies reporting the elicitation of stated preferences for child health related quality of life

Authors:

Bailey, Cate ORCID ID: 0000-0001-5030-430X
Howell, Martin ORCID ID: 0000-0001-9740-712X
Raghunandan, Rakhee ORCID ID: 0000-0001-7848-6109
Dalziel, Kim ORCID ID: 0000-0003-4972-8871
Howard, Kirsten ORCID ID: 0000-0002-0918-7540
Mulhern, Brendan ORCID ID: 0000-0003-3656-8063
Petrou, Stavros ORCID ID: 0000-0003-3121-6050
Rowen, Donna ORCID ID: 0000-0003-3018-5109
Salisbury, Amber ORCID ID: 0000-0003-1557-5918
Viney, Rosalie ORCID ID: 0000-0002-0039-9635
Lancsar, Emily ORCID ID: 0000-0003-2404-6735
Devlin, Nancy ORCID ID: 0000-0002-1561-5361

Corresponding author: Dr Cate Bailey, Health Economics Unit, Centre for Health Policy, Melbourne School of Population and Global Health, Level 4, 207 Bouverie St, Carlton, 3053, The University of Melbourne, Victoria, Australia. Tel: +61 409149737, Email: cate.bailey@unimelb.edu.au

Disclaimer:
This series is intended to promote discussion and to provide information about work in progress.
The views expressed in this series are those of the authors.
Comments are welcome, and should be sent to the corresponding author.

Title: **The RETRIEVE checklist for studies reporting the elicitation of stated preferences for child health related quality of life**

Running title: The RETRIEVE checklist

Bailey, Cate[1ψ]	ORCID ID: 0000-0001-5030-430X
Howell, Martin[2,3ψ]	ORCID ID: 0000-0001-9740-712X
Raghunandan, Rakhee[2,3]	ORCID ID: 0000-0001-7848-6109
Dalziel, Kim[1]	ORCID ID: 0000-0003-4972-8871
Howard, Kirsten[2,3]	ORCIC ID: 0000-0002-0918-7540
Mulhern, Brendan[4]	ORCID ID: 0000-0003-3656-8063
Petrou, Stavros[5]	ORCID ID: 0000-0003-3121-6050
Rowen, Donna[6]	ORCID ID: 0000-0003-3018-5109
Salisbury, Amber[2,3]	ORCID ID: 0000-0003-1557-5918
Viney, Rosalie[4]	ORCID ID: 0000-0002-0039-9635
Lancsar, Emily[7ξ]	ORCID ID: 0000-0003-2404-6735
Devlin, Nancy[1ξ]	ORCID ID: 0000-0002-1561-5361

[1] Health Economics Unit, Melbourne School of Population and Global Health, University of Melbourne, 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, Australia

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

[6] School of Health and Related Research, University of Sheffield, UK.

[7] Department of Health Services and Policy Research, Research School of Population Health, Australian National University, Australian Capital Territory, Australia

ψ joint first authors

ξ joint senior authors

Corresponding author - Dr Cate Bailey, cate.bailey@unimelb.edu.au, Health Economics Unit, Centre for Health Policy, Melbourne School of Population and Global Health, Level 4, 207 Bouverie St, Carlton, 3053, The University of Melbourne, Victoria, Australia; phone: +61 409149737

Acknowledgments We are grateful to Feng Xie, and other participants of the EuroQol Plenary meeting in Chicago 2022, for helpful suggestions made on an earlier version of this paper presented at that meeting. We are grateful to Tianxin Pan for sharing code developed to produce the theoretical distribution of values. We acknowledge with thanks the suggestions and feedback on this paper received from wider members of the QUOKKA and TORCH research programmes: Wendy Ungar, Janine Verstraete, Jonathan Craig, David Whitehurst, Mike Herdman and Karen Lee.

Funding Details of funding sources that supported the work; The QUOKKA MRFF1200816 and TORCH APP1199902 research programmes are funded by the Australian Medical Research Future Fund, Preventive and Public Health Research Initiative - 2019, Targeted Health System and Community Organisation Research Grants. SP receives support as a UK National Institute of Health Research (NIHR) Senior Investigator (NF-SI-0616-10103) and from the NIHR Applied Research Collaboration Oxford and Thames Valley.

Conflict of interest statement No authors have a conflict of interest. We note that ND, BM, RV and DR are members of the EuroQol Group, ND is Chair of the Board of the EuroQol Research Foundation and RV is Chair of the Scientific Executive of the EuroQol Group.

10 key words: utilities, HRQoL, QALYs, paediatric, children, health state values, stated preferences, generic, vignettes.

Abstract

Recent systematic reviews show varying methods for eliciting, modelling, and reporting preference-based values for child health-related quality-of-life (HRQoL) outcomes, producing value-sets with different characteristics. Checklists can improve standards of reporting; however, existing checklists do not address methodological issues for valuing child HRQoL. 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. 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. 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.

(200/200 words)

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 (Drummond et al., 2015; Neumann et al., 2016). When considering interventions targeted at children and young people, the development of health-related quality of life (HRQoL) instruments specifical to the measurement of child health (Kwon et al., 2022; Kwon et al., 2023), and the valuation of child HRQoL anchored on a 0-1 scale required for estimation of quality adjusted life years (QALYs) (Bailey et al., 2021), are key elements. There is a lack of consensus, though, about fundamental aspects of these research methods used in valuing child HRQoL (Devlin, 2022; Rowen et al., 2020).

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, recent systematic reviews of measurement and valuation of child HRQoL (Kwon et al., 2022), the psychometric performance of generic childhood multi-attribute utility instruments (Kwon et al., 2023), 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 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 (Xie et al., 2015), and SpRUCE (Brazier et al., 2019). 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 ((Kwon et al., 2022). 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 have been used in cost-effectiveness models considered by decision-making groups such as the Australian Pharmaceutical Benefits Advisory Committee (PBAC) (Bailey et al., 2021) 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 (Bailey et al., 2022). Child and young person (or for brevity, child) is used here to describe a person under 18 years of age.

2. Methods

2.1 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 they be 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 single comprehensive checklist. An initial conceptual framework was developed by the authors to identify relevant modules, informed by existing checklists for adult HRQoL values (Zoratti et al., 2021) and reviews of methods for valuing child HRQoL (Bailey et al., 2022; Kwon et al., 2022). This was refined through checklist item development and testing, using an iterative process (expanded on below). 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. 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.

Our methodology for the development of the checklist has been adapted from the EQUATOR-network guidelines for developing reporting checklists (EQUATOR network, 2018), 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) (Bailey et al., 2022; Kwon et al., 2022; Zoratti et al., 2021). The reporting checklist was then developed following the EQUATOR toolkit, including generating a list of items and conducting a series of meetings and workshops (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 discussed in 4.2 below.

2.2 Establishing potential items for each module

A review of items for reporting values for adult HRQoL (Zoratti et al., 2021) was used to identify items common to both adult and child HRQoL. Two sets of checklists were included in Zoratti et al: 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. (Zoratti et al., 2021). Items from the latter were considered for our checklist, with potential items also identified from 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 include Table 9 – MAPS (Petrou et al., 2015), 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 (a) methods issues relating to valuation of child HRQoL as identified by Rowen et al., (2020) and (b) information from two systematic reviews (Bailey et al., 2022; Kwon et al., 2022) 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.

2.3 Creating an initial list of items for each module (long version)

A series of five workshops were held with a sub-set 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. 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 5 modules.

The process of eliminating redundant items and checking relevance yielded a first draft 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. The commentary was compiled, and the checklist items were edited accordingly (MH, CB).

2.4 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 (Husereau et al., 2013) where, instead of questions, users are asked to indicate where the relevant information is located in the manuscript by page number.

2.5 Testing the checklists: an application to four studies of child HRQoL 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 (Bailey et al., 2022). 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 (Prevolnik Rupel et al., 2021) and the CHU9D (Stevens, 2012). Two papers used vignettes (Lloyd et al., 2010; Retzler et al., 2018). 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.

2.6 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. 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. From the invitations sent, 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 2.5.

3. Results

3.1 Conceptual model

The conceptual model for the RETRIEVE (**RE**porting inven**T**o**R**y ch**I**ld h**E**alth **V**alu**E**s) checklists is shown in Figure 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 to 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 (B2) or values for specific states or vignettes (B3). Modules C and A4 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 a 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 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).

3.2 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 (Husereau et al., 2013). Table S1 contains a formatted version of the long RETRIEVE. Table S2 contains examples of the use of the long and short RETRIEVE checklists. 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 Supplementary Material.

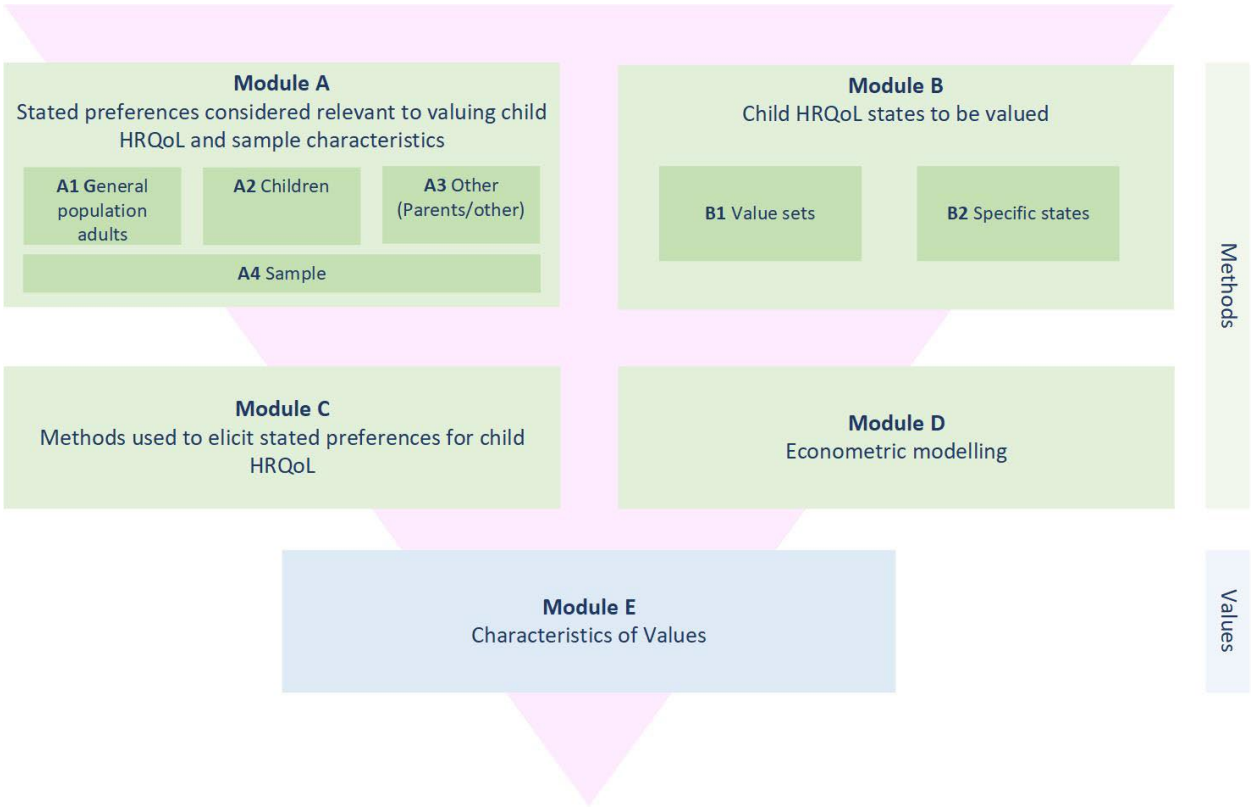

Figure 1. A conceptual framework for the RETRIEVE modular checklists for reporting values for child HRQoL.

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 preference were sought? (Yes/No)

A2 Adults' stated preferences

A2a -Which adults were the focus of preference elicitation? (General population/ Parent or caregiver of child/ Health care 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/Person 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., doorknocking, location, online panel, convenience sample)? (Yes/Partial/No)

A4c - If data were collected online, were efforts made to avoid on-line panel fraud? (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)

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 (eg 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 sub-groups and interaction variable 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 there qualitative or quantitative evidence reported that demonstrates the extent to which respondents engaged with and understood the valuation tasks? (Yes/No)

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

E3 - Did authors report the distribution of values over all states defined by the HRQoL instrument (e.g. as per Figure 1 (see Figure 1 in (Yes/No)

E4 - Key characteristics of the values

E4a - How many percentage values less than zero were possible? ____

E4b - What was the maximum possible value less than one? ____

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)

Table 2- RETRIEVE Short Checklist

Please note that this checklist is modular, and not all sections/questions will apply to all papers**.**

		Location
MODULE A	**Stated preferences considered relevant to valuing child HRQoL and sample characteristics**	
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**	
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 a HRQoL instrument, there was sufficient information to enable readers to estimate utility scores for all health states described by the instrument	

4. 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, 17 value-sets for the EQ-5D-Y-3L instrument have commenced or been completed since 2020 (Devlin et al., 2022). However, the methods being used to value the EQ-5D-Y-3L and other childhood HRQoL instruments vary widely (Bailey et al., 2022; Kwon et al., 2022), 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. While our objective was to develop a reporting checklist for childhood HRQoL values to fill a key gap, we also expect the checklist will prove useful for those developing or using adult HRQoL values, as the modules – especially Module E – are relevant to HRQoL measures more generally.

The conceptual framework and selection of modules was based on the combined expert views of the authors across our two research teams, QUOKKA and TORCH comprising researchers in Australia, the UK and North America.. 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; (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 (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 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. 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 (Bailey et al., 2022; Kwon et al., 2022). We additionally suggest in item E4 in the RETRIEVE long 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 reported in Appendix 1; see Figures 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 (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" (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 (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 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' (National Institute for Health and Care Excellence., 2022). 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; however, through the process of workshopping the items, this item was removed as being overly challenging for users of the checklists and we were concerned about reviewer burden.

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 as a means of assigning values to the disease specific states rather than directly using stated preference methods.

The EQUATOR guidelines note that production of a checklist, on its own, will not necessarily result in its use (section 5) (EQUATOR network, 2018). In respect to the suggestions included in the guidelines, we plan to make the checklist available 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). 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 (section 6).

5. 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’.

6. References

Bailey, C., 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

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.

Brazier, J., 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

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

Devlin, N. J. (2022). Valuing Child Health Isn't Child's Play. *Value in Health*, *25*(7), 1087–1089. https://doi.org/10.1016/j.jval.2022.05.009

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*.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes* (4th ed.). Oxford University Press. https://global.oup.com/academic/product/methods-for-the-economic-evaluation-of-health-care-programmes-9780199665884?cc=nz&lang=en&

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

Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., Augustovski, F., Briggs, A. H., Mauskopf, J., & Loder, E. (2013). Consolidated Health Economic Evaluation Reporting Standards (CHEERS)—Statement. *European Journal of Health Economics*, *14*(3), 367–372. https://doi.org/10.1007/s10198-013-0471-6

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

Kwon, J., Smith, S., Raghunandan, R., Howell, M., Huynh, E., Kim, S., Bentley, T., Roberts, N., Lancsar, E., Howard, K., Wong, G., Craig, J., & Petrou, S. (2023). Systematic Review of the Psychometric Performance of Generic Childhood Multi-attribute Utility Instruments. *Applied Health Economics and Health Policy*. https://doi.org/10.1007/s40258-023-00806-8

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

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

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

Neumann, P. J., Ganiats, T. G., Russell, L. B., Sanders, G. D., & Siegel, J. E. (Eds.). (2016). *Cost-effectiveness in health and medicine* (Second edition). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190492939.002.0007

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

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

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

Supplementary files for RETRIEVE

The RETRIEVE checklist for studies reporting the elicitation of stated preferences for child health related quality of life. Bailey, Howell et al.

Title The RETRIEVE checklist for studies reporting the elicitation of stated preferences for child health related quality of life

Table S1 - The RETRIEVE long checklist

This checklist is modular, not all sections 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 — *Go to A2* ☐ Children and young people (CYP) <18 years — *Go to A3* ☐ Mixed adults and CYP — *Complete A2 and A3*
A1b	Did the authors provide a rationale for whose preference were sought? ☐ Yes ☐ No
A2 Adults' stated preferences	
A2a	Which adults were the focus of preference elicitation? ☐ General population ☐ Parent or caregiver of child ☐ Health care 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 ☐ Person 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 — Go to A2d ☐ No — Go to A4 ☐ Not applicable — Go to A4
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 children, 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) Go to A4 ☐ It was applicable but not stated Go to A3f ☐ Yes Go to A3e
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., doorknocking, location, online panel, convenience sample)? ☐ Yes ☐ Partial ☐ No
A4c	If data were collected online, were efforts made to avoid on-line panel fraud? ☐ 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 Go to A4g
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 Go to A4i

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? *Go to B2* ☐ Values for a limited number of health states (e.g. vignette)? *Go to B3*
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	
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	Go to C4
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	Go to C5
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 (eg used to represent average preferences)?** ☐ Yes ☐ No ☐ Unclear	 Go to C9 Go to C9
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? Go to D2 ☐ 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? 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 *If no, go to D2h*
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 *If no, go to D2m*
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 *If yes, go to D2p* ☐ No
D2n	Was any post estimation undertaken to force monotonicity (e.g. collapsing levels)? ☐ Yes ☐ No ☐ Unclear/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 *If no, go to D3c*
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 *If no, go to D3h*

D3e	Were sub-groups and interaction variable chosen for assessment justified? ☐ Yes ☐ No
D3f	Were sensitivity analyses undertaken? ☐ Yes *If no, go to Section E* ☐ No
D3g	Were sensitivity analyses described? ☐ Yes ☐ No

Section E - Characteristics of values	
E1	**Was there qualitative or quantitative evidence reported that demonstrates the extent to which respondents engaged with and understood the valuation tasks?** ☐ Yes ☐ No
E2	**Where a value was reported, were the values generated by the final model logically consistent?** ☐ Yes ☐ No ☐ Unclear
E3	**Did authors report the distribution of values over all states defined by the HRQoL instrument (e.g. as per Figure 1 from Pan et al 2022, showen below)** ☐ Yes ☐ No
E4	**Key characteristics of the values**
E4a	How many percentage values less than zero were possible? _____
E4b	What was the maximum possible value less than one? _____
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

Figure 1 of Pan et al. (2022)

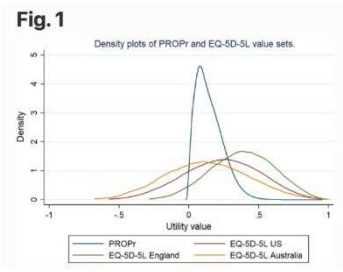

Table S2 – Examples of the use of long and short forms

Table S2a

Review of Prevolnik Rupel, 2021 (EQ-5D-Y value set) using the RETRIEVE Checklist (Short form and long form).

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

SHORT FORM:

MODULE A		**Stated preferences considered relevant to valuing child HRQoL and sample characteristics**	Location
	1	Whose preferences were sought was stated	page 464 (Sampling sub-section & Online DCE and face to face composite TTO survey sub-section, Methods)
	2	Whose perspective was used was stated	page 464 (EQ-5D-Y sub-section & Online DCE and face to face composite TTO survey sub-section, Methods)
	3	If the perspective was as a child, the child's age was stated	page 464 (EQ-5D-Y sub-section & Online DCE and face to face composite TTO survey sub-section, Methods)
	4	The population from which the sample was drawn was described and justified	page 464 (Introduction & Methods sections)
	5	The target sample size was provided and achieved	page 464 (Methods section & Sampling sub-section)
MODULE B		**Child HRQoL states to be valued**	
	6	The HRQoL instrument or health states being valued were described	page 464 (EQ-5D-Y sub-section, Methods)
	7	The choice of health states being valued was stated and justified	page 464-465 (Online DCE and face to face composite TTO survey sub-section, Methods)
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.)	page 464 (Introduction), page 464-465 (Online DCE and face to face composite TTO survey sub-section, Methods)
	9	The mode of administration for the valuation tasks was stated (e.g. face-to-face, online, in person etc.)	page 464-465 (Online DCE and face to face composite TTO survey sub-section, Methods)
	10	How values were anchored at 1 = full health and 0 = dead was stated	page 464-465 (Methods section, Sampling sub-section & Online DCE and face to face composite TTO survey sub-section, Methods)
MODULE D		**Econometric modelling and statistical methods**	
	11	The modelling and statistical methods applied to the data were stated and justified	page 465-466 (Data analysis sub-section, Methods)
	12	The basis for choosing the final model and any post-model decisions were clearly stated and justified	page 465-466 (Data analysis sub-section, Methods) & page 466-467 (Results section)
MODULE E		**Characteristics of values**	
	13	The characteristics and distributions of values for all health states relevant to the study are reported	Not reported
	14	If a value set was derived for a HRQoL instrument, there was sufficient information to enable readers to estimate utility scores for all health states described by the instrument	page 468 (Table 2)

LONG FORM:

This checklist is modular, not all sections 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 [x] *Go to A2* ☐ Children and young people (CYP) <18 years *Go to A3* ☐ Mixed adults and CYP *Complete A2 and A3*
A1b	Did the authors provide a rationale for whose preference were sought? ☐ Yes ☐ No [x] 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)
A2 Adults' stated preferences	
A2a	Which adults were the focus of preference elicitation? ☐ General population [x] ☐ Parent or caregiver of child ☐ Health care 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 [x] ☐ 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 ☐ Person 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 [x] *Go to A2d* ☐ No *Go to A4* ☐ Not applicable *Go to A4*
A2d	If yes, what was the age of the child? 10 years
A2e	Was the rationale for the choice of the age of child provided? ☐ Yes [x] Prior studies and following the EQ-5D-Y valuation protocol ☐ No
A3 Children and young people's stated preferences **Section A3 is not relevant to the value set reported by Prevolnik-Rupel (2021)**	
A3a	From which child/young person were preferences elicited? [N/A]

	☐ General population ☐ Person with a health condition ☐ Other children, please specify: _____
A3b	What perspective was the (child/young person) respondent asked to take? e.g. thinking about the health states as experienced by: [N/A] ☐ 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? [N/A] ☐ Not applicable (i.e. own perspective/themselves) *Go to A4* ☐ It was applicable but not stated *Go to A3e* ☐ Yes *Go to A3d*
A3d	If the age was specified, what was the age? [N/A]
A3e	Was the rationale for the choice of the age of child/young person provided? [N/A] ☐ Yes ☐ No
A4 Sample	
A4a	Was the population or sample frame defined from which the sample was drawn? (e.g., country, age, condition) ☐ Yes [x] Slovenian adults 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). ☐ No
A4b	Is information provided on how the sample was recruited (e.g., doorknocking, location, online panel, convenience sample)? ☐ Yes ☐ Partial [x] Online panel for DCE and unclear for cTTO. ☐ No
A4c	If data were collected online, were efforts made to avoid on-line panel fraud? ☐ Yes ☐ No [x] ☐ 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 [x] Stated as 1276 for the DCE and 200 for the cTTO. ☐ No *Go to A4g*
A4e	Was the target sample justified? ☐ Yes ☐ No [x] No justification was provided although this study was following the protocol for valuation of EQ-5D. 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.

A4f	Was the target sample achieved? ☐ Yes ☐ No [x] 1074 for the DCE and 202 for the cTTO. Not all data met the quality control criteria ☐ Unclear
A4g	Were the characteristics of the final sample described? ☐ Yes [x] ☐ No *Go to A4i*
A4h	Did the sample characteristics match the intended population? ☐ Yes ☐ No [x] ? 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. ☐ Unclear
A4i	Was the year the data collected stated? ☐ Yes – what year(s) were the data collected? [x] Nov 2019 to Feb 2020 ☐ No
A4j	Was information provided on missing data? (non-completion, withdrawals)? ☐ Yes [x] Unclear. 89% completion for DCE and 96% for TTO after excluding per data quality. ☐ 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? [x] *Go to B2* ☐ Values for a limited number of health states (e.g. vignette)? *Go to B3*
B2 Value Sets	
B2a	Which HRQoL instrument was valued? EQ-5D-Y-3L
B2b	Were the domains and response options of the instrument clearly described? ☐ Yes [x] ☐ No
B2c	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. The experimental design for the DCE utilised a D-efficient design with main effects, all two way interactions, a minimal number of unrealistic health states, overlapping of health states in two dimensions levels, and the right level and utility balance. The DCE design then randomly selected 150 pairs of health states that maximised the Fisher information matrix. The randomly selected 150 pairs of health states were divided into 10 blocks of 15 DCE tasks.
B2d	How were the health states assigned to respondents? Each respondent was asked to complete 1 of the 10 blocks of 15 DCE tasks. Each of the 15 DCE tasks presented 2 health states and the

	respondent was asked to choose their preferred state (i.e. a forced choice). No information was given regarding how respondents were assigned to complete 1 out of the 10 blocks of 15 DCE tasks.
B3 Specific health states Section B3 is not relevant to the value set reported by Prevolnik-Rupel (2021)	
B3a	How were the health states described? [N/A] ☐ Disease specific vignettes ☐ From a disease-specific HRQoL instrument ☐ Other, please specify _____
B3b	How many health states were preferences elicited for? [N/A]
B3c	Was the rationale for the selection of these health states specified? [N/A] ☐ 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 [x] ☐ TTO [x] ☐ SG ☐ BWS ☐ VAS ☐ Other, please specify_____
C2	**Was a rationale for the choice of method(s) provided?** ☐ Yes [x] ☐ No
C2a	If yes, what was the rationale? Complying with The International Valuation Protocol for the EQ-5D-Y-3L. Specifically the cTTO used to anchor the DCE to 0 to 1.
C3	**Was the duration of the states to be valued reported (e.g 'x years in this state, followed by death')?** ☐ Yes [x] ☐ No *Go to C4*
C3a	Was the duration fixed? ☐ Yes [x] ☐ No
C3b	What duration(s) was used? 10 years
C4	**Did the method(s) allow values to be elicited that were < 0 ('worse than dead')?** ☐ Yes [x] ☐ No *Go to C5*
C4a	How were values < 0 elicited? Using a lead time TTO, which is part of the composite TTO approach (cTTO).
C4b	What was the minimum value possible? (may vary according to the method used so should be clearly stated) -1
C4c	What determined how the task was terminated? The tasks were not actually described in the paper, but rather referenced the EQ-5D protocol.

C5	**How were the values anchored on a utility scale?** 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.	
C6	**What was the mode of administration for the stated preference tasks?** ☐ Online self-completion by the respondent [x] DCE only ☐ Self-completion of mailed questionnaires ☐ Online computer assisted personal interview (CAPI) ☐ In person CAPI ☐ In person interview [x] cTTO only ☐ Other, please specify _____	
C7	**How was the quality of stated preference data assessed?** The DCE included 3 rationality questions i.e. 3 fixed dominant pairs where only 1 health state considered logically dominant in each pair. 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.	
C8	**Were any exclusions made to the preference data (eg used to represent average preferences)?** ☐ Yes [x] ☐ No ☐ Unclear	 *Go to C9* *Go to C9*
C8a	**Were reasons for the exclusions provided?** ☐ Yes [x] ☐ No ☐ Unclear	
C9	**Were the health states randomly assigned?** ☐ Yes ☐ No ☐ Unclear [x]	
C10	**Was ethics approval for the study obtained from an appropriate research ethics committee?** ☐ Yes [x] ☐ 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 [x] ☐ No	

Section D – Econometric modelling and statistical methods		
D1 – Did the values reported comprise:	☐ A value set? [x] ☐ values for a limited number of health states (vignette or condition-specific)?	*Go to D2* *Go to D3*
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. Random utility model – Linear additive utility with all variables dummy coded	

D2b	Were the main assumptions of the model stated? (e.g. assumptions about preference homogeneity/heterogeneity) ☐ Yes [x] ☐ No ☐ Unclear
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	How were missing data handled (e.g.: imputation, complete case analysis) No details provided for handling of missing data.
D2e	Were subgroup analyses completed? ☐ Yes ☐ No ☐ Not applicable [x]
D2f	Were interaction terms included? ☐ Yes ☐ No [x] *If no, go to D2h*
D2g	Were details of the interactions provided? ☐ Yes ☐ No ☐ Not applicable [x]
D2h	Were non-linear specifications considered? ☐ Yes ☐ No [x]
D2i	Was more than one model described? ☐ Yes ☐ No [x] *If no, go to D2m*
D2j	Were goodness-of-fit statistics for each model reported? ☐ Yes [x] ☐ No
D2k	Was the preferred model clearly stated? N/A ☐ Yes ☐ No
D2l	Were the criteria used to select the preferred model described? N/A ☐ Yes ☐ No
D2m	Do the preference parameters for the health states follow a logical order (monotonic)? ☐ Yes *If yes, go to D2p* ☐ No [x]
D2n	Was any post estimation undertaken to force monotonicity (e.g. collapsing levels)? ☐ Yes ☐ No ☐ Unclear/not stated [x]

D2o	How were insignificant differences between adjacent levels managed (e.g. collapsed/ forced to be different)? Not clear as differences between coefficients not presented would need to calculate from Table 2 using the SEs.
D2p	Were robustness checks conducted? ☐ Yes [x] ☐ No
D2q	Was uncertainty around values reported? ☐ Yes [x] ☐ No
D3 Analysis of values for specific HRQoL states Not relevant to the value set reported by Prevolnik-Rupel (2021).	
D3a	Have the statistical methods been described? [N/A] ☐ Yes ☐ No *If no, go to D3c*
D3b	Have the statistical methods been justified? [N/A] ☐ Yes ☐ No
D3c	How were missing data handled (e.g.: imputation, complete case analysis)? [N/A]
D3d	Have subgroup analyses and interactions been undertaken? [N/A] ☐ Yes ☐ No *If no, go to D3h*
D3e	Were sub-groups and interaction variable chosen for assessment justified? [N/A] ☐ Yes ☐ No
D3f	Were sensitivity analyses undertaken? [N/A] ☐ Yes *If no, go to Section E* ☐ No
D3g	Were sensitivity analyses described? [N/A] ☐ Yes ☐ No

Section E - Characteristics of values	
E1	**Was there qualitative or quantitative evidence reported that demonstrates the extent to which respondents engaged with and understood the valuation tasks?** ☐ Yes [x] ☐ No
E2	**Where a value was reported, were the values generated by the final model logically consistent?** ☐ Yes [x] ☐ No ☐ Unclear
E3	**Did authors report the distribution of values over all states defined by the HRQoL instrument (e.g. as per Figure 1** ☐ Yes

	☐ No [x]
E4	**Key characteristics of the values**
E4a	How many percentage values less than zero were possible? 50 health states – 20.6%
E4b	What was the maximum possible value less than one? 0.962
E4c	Where in the descriptive system does the biggest change in values occur, when shifting between adjacent states? Unclear, but possibly the shift from 33333 to 32333.
E5	**Was the order of importance of dimensions (domains) suggested by the value set discussed?** ☐ Yes [x] ☐ No

Table S2b

Review of *Stevens 2012 (CHU9D value set)* using the RETRIEVE checklist (Short form and long form)

Paper title: *Valuation of the Child Health Utility 9D Index*

SHORT FORM:

MODULE A		**Stated preferences considered relevant to valuing child HRQoL and sample characteristics**	Location
	1	Whose preferences were sought were stated	page 729 (Abstract methods) & page 730-731 (Valuation technique and perspective sub-section, Methods)
	2	Whose perspective was used was stated	page 730-731 (Valuation technique and perspective sub-section, Methods)
	3	If the perspective is as a child, the child's age is stated	Not applicable – see page 731 (Valuation technique and perspective sub-section, Methods)
	4	The population from which the sample is drawn is described and justified	page 730-731 (Valuation technique and perspective sub-section, Methods) & page 731 (Sample sub-section, Methods)
	5	The target sample size is provided and achieved	page 731 (Sample sub-section, Methods) & page 735-736 (Sample sub-section, Results)
MODULE B		**Child HRQoL states to be valued**	
	6	The HRQoL instrument or health states being valued are described	page 730 (Introduction) & page 731 (Selection of health states sub-section, Methods)
	7	The choice of health states being valued is stated and justified	page 731 (Selection of health states sub-section, Methods)
MODULE C		**Methods used to elicit stated preferences for child HRQoL**	
	8	The valuation methods used to value health states are described and justified (e.g. cTTO, DCE etc.)	page 730-731 (Valuation technique and perspective sub-section, Methods) & page 732 (Valuation interviews sub-section, Methods)
	9	The mode of administration for the valuation tasks is stated (e.g. face-to-face, online, in person etc.)	page 731 (Sample sub-section, Methods)
	10	How values are anchored at 1 = full health and 0 = dead is stated	page 732 (Selection of health states sub-section, Methods)
MODULE D		**Econometric modelling and statistical methods**	
	11	The modelling and statistical methods applied to the data is stated and justified	page 734-735 (Modelling section, including all sub-sections)
	12	The basis for choosing the final model and any post-model decisions are clearly stated and justified	page 735 (Assessment of the models sub-section, Methods), page 737 (Further modelling sub-section, Results), page 739 (Discussion) & page 745 (Conclusion)
MODULE E		**Characteristics of values**	
	13	The characteristics and distributions of values for all health states relevant to the study are reported	Not reported
	14	If a value set is derived for a HRQoL instrument, there is sufficient information to enable readers to estimate utility scores for all health states described by the instrument	page 743 (Table 7)

LONG FORM:

This checklist is modular, not all sections 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 [x] *Go to A2* ☐ Children and young people (CYP) <18 years *Go to A3* ☐ Mixed adults and CYP *Complete A2 and A3*
A1b	Did the authors provide a rationale for whose preference were sought? ☐ Yes [x] As per NICE recommendations ☐ No
A2 Adults' stated preferences	
A2a	Which adults were the focus of preference elicitation? ☐ General population [x] ☐ Parent or caregiver of child ☐ Health care 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 [x] "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." ☐ Person 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 *Go to A2d* ☐ No *Go to A4*

	☐ Not applicable [x] *Go to A4*
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 Not relevant to the value set reported by Stevens et al (2012)	
A3a	From which child/young person were preferences elicited? [N/A] ☐ General population ☐ Person with a health condition ☐ Other children, please specify: _____
A3b	What perspective was the (child/young person) respondent asked to take? e.g. thinking about the health states as experienced by: [N/A] ☐ 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? [N/A] ☐ Not applicable (i.e. own perspective/themselves) *Go to A4* ☐ It was applicable but not stated *Go to A3e* ☐ Yes *Go to A3d*
A3d	If the age was specified, what was the age? [N/A]
A3e	Was the rationale for the choice of the age of child/young person provided? [N/A] ☐ Yes ☐ No
A4 Sample	
A4a	Was the population or sample frame defined from which the sample was drawn? (e.g., country, age, condition) ☐ Yes [x] Random sample (street) from general public (adults) UK (Sheffield and Huddersfield). ☐ No
A4b	Is information provided on how the sample was recruited (e.g., doorknocking, location, online panel, convenience sample)? ☐ Yes x] Random sample (street) from general public (adults) UK (Sheffield and Huddersfield) i.e. software used to randomly select street addresses – then posted invitation followed by door knocking at the sampled addresses. ☐ Partial ☐ No
A4c	If data were collected online, were efforts made to avoid on-line panel fraud? ☐ Yes ☐ No ☐ Not applicable [x]

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 [x] 300 ☐ No *Go to A4g*
A4e	Was the target sample justified? ☐ Yes [x] Based on what was achievable with the resources available. ☐ No
A4f	Was the target sample achieved? ☐ Yes [x] 300 (from 1245 addresses) but 282 used in final analysis ☐ No ☐ Unclear
A4g	Were the characteristics of the final sample described? ☐ Yes [x] ☐ No *Go to A4i*
A4h	Did the sample characteristics match the intended population? ☐ Yes ☐ No ☐ Unclear [x] This was only described in terms of affluence level, and it did not match with the UK general population. No other sample characteristics were compared against the general population. Although the sample was a random selection from a defined area.
A4i	Was the year the data collected stated? ☐ Yes – what year(s) were the data collected? _____ ☐ No [x]
A4j	Was information provided on missing data? (non-completion, withdrawals)? ☐ Yes [x] ☐ 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? [x] *Go to B2* ☐ Values for a limited number of health states (e.g. vignette)? *Go to B3*
B2 Value Sets	
B2a	Which HRQoL instrument was valued? CHU9D
B2b	Were the domains and response options of the instrument clearly described? ☐ Yes [x] ☐ No
B2c	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.

B2d	How were the health states assigned to respondents? "The 64 states were divided into eight sets of eight, trying to balance the severity of states in each set (by looking at the levels on each dimension) and making sure the two duplicate states were separated. The worst health state (called 'PITS', which is the state with the lowest level on each dimension, i.e. state 555555555) was added to each set, giving a total of nine health states in each set. Each interviewer used all eight sets and rotated round the sets using a different set for each interview so that each state got an equal number of observations and each respondent only had nine SG valuation tasks to do."
B3 Specific health states Not relevant to the value set reported by Stevens et al (2012)	
B3a	How were the health states described? [N/A] ☐ Disease specific vignettes ☐ From a disease-specific HRQoL instrument ☐ Other, please specify _____
B3b	How many health states were preferences elicited for? [N/A]
B3c	Was the rationale for the selection of these health states specified? [N/A] ☐ 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 [x] ☐ BWS ☐ VAS ☐ Other, please specify_____
C2	**Was a rationale for the choice of method(s) provided?** ☐ Yes [x] ☐ No
C2a	If yes, what was the rationale? Based on prior valuations for NICE
C3	**Was the duration of the states to be valued reported (e.g 'x years in this state, followed by death')?** ☐ Yes [x] ☐ No *Go to C4*
C3a	Was the duration fixed? ☐ Yes [x] ☐ No
C3b	What duration(s) was used? "Rest of their lives" – so strictly speaking could be considered not fixed
C4	**Did the method(s) allow values to be elicited that were < 0 ('worse than dead')?** ☐ Yes [x] ☐ No *Go to C5*
C4a	How were values < 0 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.

C4b	What was the minimum value possible? (may vary according to the method used so should be clearly stated) -1	
C4c	What determined how the task was terminated? Not clear. Implication is that during interview the SG task was terminated at point of indifference which is the point at which the utility value is assigned	
C5	**How were the values anchored on a utility scale?** Using the values from the SG task that are automatically on the 1-0 scale where 0=dead.	
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 [x] ☐ Other, please specify _____	
C7	**How was the quality of stated preference data assessed?** Other than exclusions, no other detail was provided regarding assessing data quality. Data was excluded on basis of 'unusable' and if respondents valued all health states the same.	
C8	**Were any exclusions made to the preference data (e.g. used to represent average preferences)?** ☐ Yes [x] ☐ No ☐ Unclear	 *Go to C9* *Go to C9*
C8a	**Were reasons for the exclusions provided?** ☐ Yes [x] Data was excluded on basis of 'unusable' and if respondents valued all health states the same. ☐ No ☐ Unclear	
C9	**Were the health states randomly assigned?** ☐ Yes ☐ No ☐ Unclear [x]	
C10	**Was ethics approval for the study obtained from an appropriate research ethics committee?** ☐ Yes [x] ☐ 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 [x] ☐ No	

Section D – Econometric modelling and statistical methods		
D1 – Did the values reported comprise:		
	☐ A value set? [x]	*Go to D2*
	☐ 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? OR What models were estimated? e.g. OLS, Tobit etc. Additive model Uij = g(βxij) + εij	

	OLS, RE and FE if individual effects considered important i.e. g is a linear function (the warm-up rank data was modelled separately)
D2b	Were the main assumptions of the model stated? (e.g. assumptions about preference homogeneity/heterogeneity) ☐ Yes [x] ☐ No ☐ Unclear
D2c	How was the constant term treated (if included)? Fixed at 1 to give disutility
D2d	How were missing data handled (e.g.: imputation, complete case analysis) Complete case analysis
D2e	Were subgroup analyses completed? ☐ Yes ☐ No ☐ Not applicable [x] ☐
D2f	Were interaction terms included? ☐ Yes [x] ☐ No *If no, go to D2h*
D2g	Were details of the interactions provided? ☐ Yes [x] '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. ☐ No ☐ Not applicable
D2h	Were non-linear specifications considered? ☐ Yes ☐ No [x]
D2i	Was more than one model described? ☐ Yes [x] ☐ No *If no, go to D2m*
D2j	Were goodness-of-fit statistics for each model reported? ☐ Yes [x] ☐ No
D2k	Was the preferred model clearly stated? ☐ Yes [x] Conclusion states that "The model recommended for use in assigning preference weights for the health states defined by the CHU9Dis the OLS parsimonious model (model 5)." ☐ No
D2l	Were the criteria used to select the preferred model described? ☐ Yes [x] ☐ No

D2m	Do the preference parameters for the health states follow a logical order (monotonic)? ☐ Yes *If yes, go to D2p* ☐ No [x]
D2n	Was any post estimation undertaken to force monotonicity (e.g. collapsing levels)? ☐ Yes [x] ☐ No ☐ Unclear/not stated
D2o	How were insignificant differences between adjacent levels managed (e.g. 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
D2p	Were robustness checks conducted? ☐ Yes [x] Mean absolute error and root mean square error ☐ No
D2q	Was uncertainty around values reported? ☐ Yes [x] Standard errors ☐ No
D3 Analysis of values for specific HRQoL states **Not relevant to the value set reported by Stevens et al (2012)**	
D3a	Have the statistical methods been described? [N/A] ☐ Yes ☐ No *If no, go to D3c*
D3b	Have the statistical methods been justified? [N/A] ☐ Yes ☐ No
D3c	How were missing data handled (e.g.: imputation, complete case analysis)? [N/A]
D3d	Have subgroup analyses and interactions been undertaken? [N/A] ☐ Yes ☐ No *If no, go to D3h*
D3e	Were sub-groups and interaction variable chosen for assessment justified? [N/A] ☐ Yes ☐ No
D3f	Were sensitivity analyses undertaken? [N/A] ☐ Yes *If no, go to Section E* ☐ No
D3g	Were sensitivity analyses described? [N/A] ☐ Yes ☐ No

Section E - Characteristics of values	
E1	**Was there qualitative or quantitative evidence reported that demonstrates the extent to which respondents engaged with and understood the valuation tasks?** ☐ Yes [x] Reported in Table 1 of the paper.

	☐ No
E2	**Where a value was reported, were the values generated by the final model logically consistent?** ☐ Yes [x] The final model was logically consistent. In initial models there were inconsistencies requiring additional parsimonious models ☐ No ☐ Unclear
E3	**Did authors report the distribution of values over all states defined by the HRQoL instrument (e.g. as per Figure 1** ☐ Yes ☐ No [x]
E4	**Key characteristics of the values**
E4a	How many percentage values less than zero were possible? 23 (0.93%)
E4b	What was the maximum possible value less than one? 0.993 (stated in Table 2 of the paper).
E4c	Where in the descriptive system does the biggest change in values occur, when shifting between adjacent states? Unclear
E5	**Was the order of importance of dimensions (domains) suggested by the value set discussed?** ☐ Yes ☐ No [x] Not discussed, however greatest disutility was for pain 5 (0.1461) and smallest for Worry 2345 (0.0251) and Sleep 23 (0.028).

Table S2c

Review of *Lloyd 2010* using the RETRIEVE checklist (Short form and long form)

Paper title: *A Valuation of Infusion Therapy to Preserve Islet Function in Type 1 Diabetes.*

SHORT FORM:

MODULE A		**Stated preferences considered relevant to valuing child HRQoL and sample characteristics**	Location
	1	Whose preferences are sought was stated	page 636 (Methods), page 637-638 (Valuation study sub-section, Methods) & page 641 (Discussion)
	2	Whose perspective was used was stated	page 638 (Valuation study sub-section, Methods) & page 641 (Discussion)
	3	If the perspective was as a child, the child's age was stated	page 638 (Valuation study sub-section, Methods) & page 641 (Discussion)
	4	The population from which the sample was drawn was described and justified	page 637 (Valuation study sub-section, Methods)
	5	The target sample size was provided and achieved	page 641 (Discussion)
MODULE B		**Child HRQoL states to be valued**	
	6	The HRQoL instrument or health states being valued were described	page 637 (Health state development and piloting sub-section, Methods)
	7	The choice of health states being valued was stated and justified	page 637 (Health state development and piloting sub-section, Methods)
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.)	page 637-638 (Valuation study sub-section, Methods)
	9	The mode of administration for the valuation tasks was stated (e.g. face-to-face, online, in person etc.)	page 637 (Valuation study sub-section, Methods)
	10	How values are anchored at 1 = full health and 0 = dead was stated	page 638 (Valuation sub-section & Statistical analysis sub-section, Methods)
MODULE D		**Econometric modelling and statistical methods**	
	11	The modelling and statistical methods applied to the data was stated and justified	page 638 (Statistical analysis sub-section, Methods)
	12	The basis for choosing the final model and any post-model decisions ware clearly stated and justified	page 638 (Statistical analysis sub-section, Methods)
MODULE E		**Characteristics of values**	
	13	The characteristics and distributions of values for all health states relevant to the study were reported	Not reported
	14	If a value set was derived for a HRQoL instrument, there was sufficient information to enable readers to estimate utility scores for all health states described by the instrument	Not applicable

LONG FORM:

This checklist is modular, not all sections 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 [x] Parents of children & adolescents (<18 years old) with type 1 diabetes mellitus (T1DM) were selected and asked to assess the child & adolescent states using the EQ-5D proxy version; adult patients (18-35 years old) with T1DM & the general population were selected and asked to assess adult T1DM states, which is not applicable to this checklist. Therefore the remainder of the checklist will be applied to just the parents of children & adolescents (<18 years old) with type 1 diabetes mellitus (T1DM) were selected and asked to assess the child & adolescent states using the EQ-5D proxy version. *Go to A2* ☐ Children and young people (CYP) <18 years ☐ Mixed adults and CYP *Go to A3* *Complete A2 and A3*
A1b	Did the authors provide a rationale for whose preference were sought? ☐ Yes [x] Adult patients (18-35 years old) with type 1 diabetes mellitus (T1DM) or parents of children & adolescents (<18 years old) with T1DM were selected based on their direct experience with condition. No rationale/justification for use of general population. Note only the parents of children & adolescents (<18 years old) with T1DM are relevant to the checklist as they were asked to assess the child & adolescent states using the EQ-5D proxy version ☐ No
A2 Adults' stated preferences	
A2a	Which adults were the focus of preference elicitation? ☐ General population ☐ Parent or caregiver of child [x] Only the parents of children & adolescents (<18 years old) with T1DM were asked to assess the child & adolescent states using the EQ-5D proxy version. The adult patients with T1DM and the general population were asked to assess adult T1DM states, which is not applicable to this checklist. ☐ Health care 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 [x] "Parents were asked to complete the VAS and SG exercises as if they were a child of X years of age (where X was the age of their own child with T1DM) ☐ Their own health, but blinded to the states under consideration being specific to children

	☐ Person 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 [x] *Go to A2d* ☐ No *Go to A4* ☐ Not applicable *Go to A4*
A2d	If yes, what was the age of the child? When completing the VAS and SG tasks, the parents of the children with T1DM were asked to imagine themselves as a child who is the same age as their own child
A2e	Was the rationale for the choice of the age of child provided? ☐ Yes [x] Implied rationale is that to be the same as their child who has T1DM enables lived experience to be reflected ☐ No
A3 Children and young people's stated preferences Not relevant to Lloyd et al (2010)	
A3a	From which child/young person were preferences elicited? [N/A] ☐ General population ☐ Person with a health condition ☐ Other children, please specify: _____
A3b	What perspective was the (child/young person) respondent asked to take? e.g. thinking about the health states as experienced by: [N/A] ☐ 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? [N/A] ☐ Not applicable (i.e. own perspective/themselves) *Go to A4* ☐ It was applicable but not stated *Go to A3e* ☐ Yes *Go to A3d*
A3d	If the age was specified, what was the age? [N/A]
A3e	Was the rationale for the choice of the age of child/young person provided? [N/A] ☐ Yes ☐ No
A4 Sample	
A4a	Was the population or sample frame defined from which the sample was drawn? (e.g., country, age, condition) ☐ Yes [x] Sample was drawn from sample frame of parents of children & adolescents (<18 years old) with T1DM in England and Scotland (area/s not specified). ☐ No
A4b	Is information provided on how the sample was recruited (e.g., doorknocking, location, online panel, convenience sample)?

	☐ Yes [x] Parents of children & adolescents (<18 years old) with T1DM were recruited from England and Scotland, through a specialist patient recruitment agency. ☐ Partial ☐ No
A4c	If data were collected online, were efforts made to avoid on-line panel fraud? ☐ Yes ☐ No ☐ Not applicable [x] No information to indicate that data was collected online, and implication seems to be in-person data collection using trained interviewers.
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 [x] 50 parents of children & adolescents (<18 years old) with T1DM ☐ No *Go to A4g*
A4e	Was the target sample justified? ☐ Yes ☐ No [x]
A4f	Was the target sample achieved? ☐ Yes ☐ No [x] 44 instead of 50 parents of children & adolescents (<18 years old) with T1DM were able to be recruited. ☐ Unclear
A4g	Were the characteristics of the final sample described? ☐ Yes [x] Demographics of parents of children & adolescents (<18 years old) with T1DM included in Table 2 ☐ No *Go to A4i*
A4h	Did the sample characteristics match the intended population? ☐ Yes ☐ No [x] The manuscript only discussed the societal adult sample in terms of matching the UK general population (Table 1). However, information from Table 2 allows us to determine that the parent sample did not match the UK general population ☐ Unclear
A4i	Was the year the data collected stated? ☐ Yes – what year(s) were the data collected? _____ ☐ No [x]
A4j	Was information provided on missing data? (non-completion, withdrawals)? ☐ Yes ☐ Partial ☐ No [x]

Section B - Child HRQoL states to be valued	
B1 Type of study	
B1	Did the values reported in this paper comprise: ☐ A value set? *Go to B2*

	☐ Values for a limited number of health states (e.g. vignette)? [x] *Go to B3*
B2 Value Sets Not relevant to Lloyd et al (2010)	
B2a	Which HRQoL instrument was valued? [N/A]
B2b	Were the domains and response options of the instrument clearly described? [N/A] ☐ Yes ☐ No
B2c	What experimental design approach was used to choose the health states (combination of dimension levels) to be valued? [N/A]
B2d	How were the health states assigned to respondents? [N/A]
B3 Specific health states	
B3a	How were the health states described? ☐ Disease specific vignettes ☐ From a disease-specific HRQoL instrument ☐ Other, please specify [x] Short vignette descriptions of health and HRQL were produced. While the adult health state vignettes were T1DM specific, the parallel health states describing adolescents (13-17 years old) and children (8-12 years old) did not make explicit reference to T1DM.
B3b	How many health states were preferences elicited for? 5 health states for each adult participant
B3c	Was the rationale for the selection of these health states specified? ☐ Yes – What was the rationale? [x] "Short vignette descriptions of health and HRQL were produced based on the interviews and literature review" ☐ 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 [x] ☐ BWS ☐ VAS [x] ☐ Other, please specify_____
C2	**Was a rationale for the choice of method(s) provided?** ☐ Yes ☐ No [x]
C2a	If yes, what was the rationale? [N/A]
C3	**Was the duration of the states to be valued reported (e.g. 'x years in this state, followed by death')?** ☐ Yes ☐ No [x] *Go to C4*
C3a	Was the duration fixed? [N/A] ☐ Yes ☐ No

C3b	What duration(s) was used? [N/A]	
C4	**Did the method(s) allow values to be elicited that were < 0 ('worse than dead')?** ☐ Yes ☐ No [x] This was unclear as statistical analysis section indicated that "SG data were rescaled against dead so that all utilities were on a 0-1.0 scale"	*Go to C5*
C4a	How were values < 0 elicited? [N/A]	
C4b	What was the minimum value possible? (may vary according to the method used so should be clearly stated) [N/A]	
C4c	What determined how the task was terminated? [N/A]	
C5	**How were the values anchored on a utility scale?** The worst health state was compared with either full health or death, which then allowed the study to rescale responses to the other health states onto a 0 (dead) to 1.0 (full health) 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 [x] ☐ Other, please specify _____	
C7	**How was the quality of stated preference data assessed?** Unclear as no detail included in manuscript.	
C8	**Were any exclusions made to the preference data (e.g. used to represent average preferences)?** ☐ Yes ☐ No ☐ Unclear [x] No detail included in the manuscript	 *Go to C9* *Go to C9*
C8a	**Were reasons for the exclusions provided?** ☐ Yes ☐ No ☐ Unclear	
C9	**Were the health states randomly assigned?** ☐ Yes ☐ No ☐ Unclear [x] No detail included in the manuscript	
C10	**Was ethics approval for the study obtained from an appropriate research ethics committee?** ☐ Yes ☐ No ☐ Unclear ☐ Not stated [x]	
C11	**Were sources of funding and non-monetary support and the role of the funder(s) in the design described?** ☐ Yes [x] ☐ No	

Section D – Econometric modelling and statistical methods	
D1 – Did the values reported comprise:	
	☐ A value set? *Go to D2* ☐ values for a limited number of health states (vignette or condition-specific)? [x] *Go to D3*
D2 Econometric modelling of value sets for HRQoL instruments **Not relevant to Lloyd et al (2010)**	
D2a	What was the theoretical model? OR What models were estimated? e.g. OLS, Tobit etc. [N/A]
D2b	Were the main assumptions of the model stated? (e.g. assumptions about preference homogeneity/heterogeneity) [N/A] ☐ Yes ☐ No ☐ Unclear
D2c	How was the constant term treated (if included)? [N/A]
D2d	How were missing data handled (e.g.: imputation, complete case analysis)? [N/A]
D2e	Were subgroup analyses completed? [N/A] ☐ Yes ☐ No ☐ Not applicable
D2f	Were interaction terms included? [N/A] ☐ Yes ☐ No *If no, go to D2h*
D2g	Were details of the interactions provided? [N/A] ☐ Yes ☐ No ☐ Not applicable
D2h	Were non-linear specifications considered? [N/A] ☐ Yes ☐ No
D2i	Was more than one model described? [N/A] ☐ Yes ☐ No *If no, go to D2m*
D2j	Were goodness-of-fit statistics for each model reported? [N/A] ☐ Yes ☐ No
D2k	Was the preferred model clearly stated? [N/A] ☐ Yes ☐ No
D2l	Were the criteria used to select the preferred model described? [N/A] ☐ Yes ☐ No

D2m	Do the preference parameters for the health states follow a logical order (monotonic)? [N/A] ☐ Yes *If yes, go to D2p* ☐ No
D2n	Was any post estimation undertaken to force monotonicity (e.g. collapsing levels)? [N/A] ☐ Yes ☐ No ☐ Unclear/not stated
D2o	How were insignificant differences between adjacent levels managed (e.g. collapsed/ forced to be different)? [N/A]
D2p	Were robustness checks conducted? [N/A] ☐ Yes ☐ No
D2q	Was uncertainty around values reported? [N/A] ☐ Yes ☐ No
D3 Analysis of values for specific HRQoL states	
D3a	Have the statistical methods been described? ☐ Yes [x] This was implied by the general approach rather than specifically stated for the subgroup analysis *If no, go to D3c* ☐ No
D3b	Have the statistical methods been justified? ☐ Yes ☐ No [x]
D3c	How were missing data handled (e.g.: imputation, complete case analysis)? No detail given regarding missing data or how it was handled.
D3d	Have subgroup analyses and interactions been undertaken? ☐ Yes [x] Mean utility estimates by English and Scottish participants were compared for similarities and were presented separately in Table 6. *If no, go to D3h* ☐ No
D3e	Were sub-groups and interaction variable chosen for assessment justified? ☐ Yes ☐ No [x]
D3f	Were sensitivity analyses undertaken? ☐ Yes ☐ No [x] *If no, go to Section E*
D3g	Were sensitivity analyses described? [N/A] ☐ Yes ☐ No

Section E - Characteristics of values

E1	**Was there qualitative or quantitative evidence reported that demonstrates the extent to which respondents engaged with and understood the valuation tasks?** ☐ Yes [x] However this was a bit unclear how it applied to all the study samples. The manuscript indicated that visual aids were used for all participants, but no further information was provided. They manuscript did indicate that the health states, VAS, and SG tasks were piloted with the general population with cognitive debriefing interviews afterwards to ascertain their ability to rate the health states; and that no issues were identified from the interviews. ☐ No
E2	**Where a value was reported, were the values generated by the final model logically consistent?** ☐ Yes [x] ☐ No ☐ Unclear
E3	**Did authors report the distribution of values over all states defined by the HRQoL instrument (e.g. as per Figure 1)** ☐ Yes ☐ No [x]
E4	**Key characteristics of the values**
E4a	How many percentage values less than zero were possible? No information reported in manuscript.
E4b	What was the maximum possible value less than one? Unsure if there was a maximum value less that one as manuscript states that SG data were rescaled on to 0-1.0 utility scale.
E4c	Where in the descriptive system does the biggest change in values occur, when shifting between adjacent states? No information reported in manuscript.
E5	**Was the order of importance of dimensions (domains) suggested by the value set discussed?** ☐ Yes ☐ No [x] Not reported in manuscript. Note that this study did not report a value set.

Table S2d

Review of *Retzler 2018* using the RETRIEVE checklist (Short form and long form)

Paper title: *Utility elicitation in adults and children for allergic rhinoconjunctivitis and associated health states.*

SHORT FORM:

MODULE A	**Stated preferences considered relevant to valuing child HRQoL and sample characteristics**	Location
1	Whose preferences were sought was stated	page 2384-2385 (Elicitation methods sub-section, Methods)
2	Whose perspective was used was stated	page 2384-2385 (Elicitation methods sub-section, Methods)
3	If the perspective was as a child, the child's age was stated	Not reported
4	The population from which the sample is drawn was described and justified	page 2386 (Survey sub-section, Methods)
5	The target sample size was provided and achieved	page x
MODULE B	**Child HRQoL states to be valued**	
6	The HRQoL instrument or health states being valued were described	page 2384 (Health states sub-section, Methods)
7	The choice of health states being valued was stated and justified	page 2385 (Survey sub-section, Methods)
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.)	page 2384 (Elicitation methods sub-section, Methods)
9	The mode of administration for the valuation tasks was stated (e.g. face-to-face, online, in person etc.)	page 2385 (Survey sub-section, Methods)
10	How values were anchored at 1 = full health and 0 = dead was stated	page 2385 (Elicitation methods sub-section, Methods)
MODULE D	**Econometric modelling and statistical methods**	
11	The modelling and statistical methods applied to the data was stated and justified	page 2386 (Statistical analysis sub-section, Methods)
12	The basis for choosing the final model and any post-model decisions were clearly stated and justified	Not applicable
MODULE E	**Characteristics of values**	
13	The characteristics and distributions of values for all health states relevant to the study were reported	page 2387 (Table 3)
14	If a value set was derived for a HRQoL instrument, there was sufficient information to enable readers to estimate utility scores for all health states described by the instrument	Not applicable

LONG FORM:

This checklist is modular, not all sections 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 — *Go to A2* ☐ Children and young people (CYP) <18 years [x] Only the CYP sample (aged 8-11 years old) were asked to assess/value the equivalent child health states (suing VAS method), which means that only this aspect of the study is relevant to assess using RETRIEVE checklist. The adult sample was asked to assess/value adult health states (using SG method), which means that this aspect of the study is not relevant to the RETRIEVE paediatric checklist. — *Go to A3* / *Complete A2 and A3* ☐ Mixed adults and CYP
A1b	Did the authors provide a rationale for whose preference were sought? ☐ Yes ☐ No [x]
A2 Adults' stated preferences **Not relevant to Retzler 2018 study as adult stated preferences were for adult health states not child health states.**	
A2a	Which adults were the focus of preference elicitation? [N/A] ☐ General population ☐ Parent or caregiver of child ☐ Health care 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: [N/A] ☐ 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 ☐ Person 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? [N/A] ☐ Yes — *Go to A2d* ☐ No — *Go to A4* ☐ Not applicable — *Go to A4*
A2d	If yes, what was the age of the child? [N/A]
A2e	Was the rationale for the choice of the age of child provided? [N/A] ☐ Yes ☐ No

A3 Children and young people's stated preferences	
A3a	From which child/young person were preferences elicited? ☐ General population [x] This was implied rather than specifically stated for the child sample ☐ Person with a health condition ☐ Other children, 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 [x] ☐ 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) *Go to A4* ☐ It was applicable but not stated [x] *Go to A3e* ☐ Yes *Go to A3d*
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 [x] Children aged 8-11 years old residing in the UK, France, Germany or Slovakia were eligible. ☐ No
A4b	Is information provided on how the sample was recruited (e.g., doorknocking, location, online panel, convenience sample)? ☐ Yes [x] Online panel respondents recruited by third party (i.e. Qualtrics) ☐ Partial ☐ No
A4c	If data were collected online, were efforts made to avoid on-line panel fraud? ☐ Yes ☐ No [x] ☐ 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 [x] Child sample: 260 complete responses for each of the 4 countries, ensuring at least 150 responses per health state. Manuscript indicates that 14 health states were developed, and each respondent completed 8 out of the available 14 health states. ☐ No *Go to A4g*
A4e	Was the target sample justified? ☐ Yes [x] Manuscript indicated that for the child sample: 260 complete responses for each of the 4 countries, ensuring at least 150 responses

	per health state. However, other than stating that smaller target samples were selected for the child sample than the adult sample, as they were more difficult to reach and recruit, no further justification was provided for these target sample sizes per country and per health state. ☐ No
A4f	Was the target sample achieved? ☐ Yes ☐ No ☐ Unclear [x] Not reported whether target sample was achieved. Table 3 and Table SB1 (Supplementary material) indicate that the base case analysis sample met the target sample size of 150 for each of the 14 health states
A4g	Were the characteristics of the final sample described? ☐ Yes [x] Only compared by gender for child sample ☐ No *Go to A4i*
A4h	Did the sample characteristics match the intended population? ☐ Yes [x] Only compared by gender split for child sample. The manuscript did not include the gender split for the child general population for each of the 4 countries, it commented that the gender split (Table SA6) for the child sample was in line with the general population in each country. ☐ No ☐ Unclear
A4i	Was the year the data collected stated? ☐ Yes – what year(s) were the data collected? _____ ☐ No [x]
A4j	Was information provided on missing data? (non-completion, withdrawals)? ☐ Yes ☐ Partial ☐ No [x] For the child sample, the recruiting continued until the target sample size of 260 complete responses were received per country and ensuring at least 150 responses for each of the 14 health states. No information was reported on incomplete responses and/or withdrawals from respondents.

Section B - Child HRQoL states to be valued	
B1 Type of study	
B1	Did the values reported in this paper comprise: ☐ A value set? *Go to B2* ☐ Values for a limited number of health states (e.g. vignette)? [x] *Go to B3*
B2 Value Sets **Not relevant to Retzler 2018 study**	
B2a	Which HRQoL instrument was valued? [N/A]
B2b	Were the domains and response options of the instrument clearly described? [N/A] ☐ Yes ☐ No
B2c	What experimental design approach was used to choose the health states (combination of dimension levels) to be valued? [N/A]

B2d	How were the health states assigned to respondents? [N/A]
B3 Specific health states	
B3a	How were the health states described? ☐ Disease specific vignettes [x] ☐ From a disease-specific HRQoL instrument ☐ Other, please specify _____
B3b	How many health states were preferences elicited for? 14 health states
B3c	Was the rationale for the selection of these health states specified? ☐ Yes – What was the rationale? [x] The vignettes describing the 14 health states were developed using the relevant condition-specific clinical guidelines, then revised after input from 2 expert clinicians (1 paediatric specialist). The revised vignettes were piloted with 8 patients (did not state whether adult or child) and final vignettes for the 14 health states were developed incorporating all feedback. ☐ 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 [x] The child sample used VAS to assess the 14 child health states and only this sample in the study applied to the RETRIEVE checklist. The adult sample (which used SG methods) did not apply to the RETRIEVE checklist as they assessed 14 adult health states. ☐ Other, please specify_____
C2	**Was a rationale for the choice of method(s) provided?** ☐ Yes [x] ☐ No
C2a	☐ If yes, what was the rationale? The manuscript indicated that the SG method was not appropriate for children due to comprehension issues and the use of the death comparator. However, no specific rationale was reported for the use of the VAS method for the child sample.
C3	**Was the duration of the states to be valued reported (e.g 'x years in this state, followed by death')?** ☐ Yes ☐ No [x] N/A *Go to C4*
C3a	Was the duration fixed? ☐ Yes ☐ No [x] N/A
C3b	What duration(s) was used? [N/A]
C4	**Did the method(s) allow values to be elicited that were < 0 ('worse than dead')?** ☐ Yes ☐ No [x] *Go to C5*
C4a	How were values < 0 elicited? [N/A]

C4b	What was the minimum value possible? (may vary according to the method used so should be clearly stated) 0	
C4c	What determined how the task was terminated? [N/A]	
C5	**How were the values anchored on a utility scale?** [N/A]	
C6	**What was the mode of administration for the stated preference tasks?** ☐ Online self-completion by the respondent [x] ☐ 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?** Manuscript indicates that data were assess for extreme values and lack of face validity, and exclusion criteria were applied. Supplementary material indicated what the exclusion criteria were: i.e. that data were excluded if they were inconsistent (i.e. theory driven rules removed responses from participants who generated a 0 utility value for any of the health states (i.e. an extreme value) and any participants who rated mild health states with a lower utility value than severe health states), or if they were implausible (base case threshold was utility value <0.3 with stricter cut-off threshold <0.5). It also indicated which exclusion criteria were applied to each analysis in the sensitivity analysis.	
C8	**Were any exclusions made to the preference data (eg used to represent average preferences)?** ☐ Yes [x] ☐ No ☐ Unclear	 *Go to C9* *Go to C9*
C8a	**Were reasons for the exclusions provided?** ☐ Yes [x] see response to question C7 above ☐ No ☐ Unclear	
C9	**Were the health states randomly assigned?** ☐ Yes [x] the Qualtrics algorithm randomly assigned (with even presentation) 8 out the 14 available child health states to each child participant. It also randomised the order of the 8 health states presented in each to reduce order effects. ☐ No ☐ Unclear	
C10	**Was ethics approval for the study obtained from an appropriate research ethics committee?** ☐ Yes [x] ☐ 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 [x] ☐ No	

Section D – Econometric modelling and statistical methods
D1 – Did the values reported comprise:

	☐ A value set? *Go to D2* ☐ values for a limited number of health states (vignette or condition-specific)? [x] *Go to D3*
D2 Econometric modelling of value sets for HRQoL instruments **Not relevant to Retzler 2018 study**	
D2a	What was the theoretical model? OR What models were estimated? e.g. OLS, Tobit etc. [N/A]
D2b	Were the main assumptions of the model stated? (e.g. assumptions about preference homogeneity/heterogeneity) [N/A] ☐ Yes ☐ No ☐ Unclear
D2c	How was the constant term treated (if included)? [N/A]
D2d	How were missing data handled (e.g.: imputation, complete case analysis)? [N/A]
D2e	Were subgroup analyses completed? [N/A] ☐ Yes ☐ No ☐ Not applicable
D2f	Were interaction terms included? [N/A] ☐ Yes ☐ No *If no, go to D2h*
D2g	Were details of the interactions provided? [N/A] ☐ Yes ☐ No ☐ Not applicable
D2h	Were non-linear specifications considered? [N/A] ☐ Yes ☐ No
D2i	Was more than one model described? [N/A] ☐ Yes ☐ No *If no, go to D2m*
D2j	Were goodness-of-fit statistics for each model reported? [N/A] ☐ Yes ☐ No
D2k	Was the preferred model clearly stated? [N/A] ☐ Yes ☐ No
D2l	Were the criteria used to select the preferred model described? [N/A] ☐ Yes ☐ No
D2m	Do the preference parameters for the health states follow a logical order (monotonic)? [N/A] ☐ Yes *If yes, go to D2p* ☐ No
D2n	Was any post estimation undertaken to force monotonicity (e.g. collapsing levels)? [N/A]

	☐ Yes ☐ No ☐ Unclear/not stated
D2o	How were insignificant differences between adjacent levels managed (e.g. collapsed/ forced to be different)? [N/A]
D2p	Were robustness checks conducted? [N/A] ☐ Yes ☐ No
D2q	Was uncertainty around values reported? [N/A] ☐ Yes ☐ No
D3 Analysis of values for specific HRQoL states	
D3a	Have the statistical methods been described? ☐ Yes [x] ☐ No *If no, go to D3c*
D3b	Have the statistical methods been justified? ☐ Yes [x] ☐ No
D3c	How were missing data handled (e.g.: imputation, complete case analysis)? Study only included and used complete case analysis.
D3d	Have subgroup analyses and interactions been undertaken? ☐ Yes ☐ No [x] While there was subgroup analysis comparing the adult sample with the child sample in the study. Only the child sample in the study was relevant to using the RETRIEVE checklist, and there was no subgroup analysis within the child sample. *If no, go to D3h*
D3e	Were sub-groups and interaction variable chosen for assessment justified? [N/A] ☐ Yes ☐ No
D3f	Were sensitivity analyses undertaken? ☐ Yes [x] *If no, go to Section E* ☐ No
D3g	Were sensitivity analyses described? ☐ Yes [x] This was described and presented in the Supplementary material. ☐ No

Section E - Characteristics of values	
E1	**Was there qualitative or quantitative evidence reported that demonstrates the extent to which respondents engaged with and understood the valuation tasks?**

	☐ Yes [x] This is a little unclear. No information was obtained to demonstrate if the participants engaged with and understood the valuation tasks. However, the study did apply theory driven exclusion criteria to exclude inconsistent or implausible responses. Also, the revised vignettes were piloted with 8 patients (did not state whether adult or child) and final vignettes for the 14 health states were developed incorporating all feedback. ☐ No
E2	**Where a value was reported, were the values generated by the final model logically consistent?** ☐ Yes ☐ No ☐ Unclear [x] Not reported
E3	**Did authors report the distribution of values over all states defined by the HRQoL instrument (e.g. as per Figure 1)** ☐ Yes ☐ No [x] They did report the mean, std error, median, and IQR for each of the 14 child health states used in the study (Table 3).
E4	**Key characteristics of the values**
E4a	How many percentage values less than zero were possible? [N/A]
E4b	What was the maximum possible value less than one? [N/A]
E4c	Where in the descriptive system does the biggest change in values occur, when shifting between adjacent states? [N/A]
E5	**Was the order of importance of dimensions (domains) suggested by the value set discussed?** ☐ Yes ☐ No [x] [N/A]

Table S3 – Checklist items with descriptive comments

No.	Item	Comments
Section A – Whose stated preferences were considered relevant to valuing child HRQoL		
A1 (A1a and A1b)	Whose stated preferences were sought? Did the authors provide a rationale for whose preference were sought?	It needs to be clear if children, adults or both were included as they require different considerations when eliciting preferences. As preferences often may differ between children of different ages and adults, a justification needs to be provided.
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, adults with a specific condition or health care professionals, all of whom may have different preferences, reference points and experiences. These differences have been shown to influence stated preferences.
A2b	What perspective were adults 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 has been shown to influence stated preferences.
A2c, A2d, A2e	Was the age of the child, for whom respondents were asked to imagine health states to be valued, specified? If yes, what was the age of the child? Was the rationale for the choice of the age of child provided?	As 'child' or children can refer to anyone less than 18 years old, and as the age of the child is known to influence stated preferences it should be clearly described. This might be as an age range (e.g. 12 to 18 years) or a discrete age. The use of terms such as 'young child' or 'toddler' without definition of an age group leads to ambiguity. Given the influence on stated preferences the choice of age should be justified.
A3	Children's stated preferences	
A3a	From which child/young person were preferences elicited?	As with adults this could include the general population, school children, or children with a specified condition all of whom may have different preferences, reference points and experience. These differences have been shown to influence stated preferences.
A3b	What perspective was the (child/young person) respondent asked to take?	Children could be asked to consider themselves, another child they know or an unknown hypothetical child. If considering themselves then preferences may be influenced by whether they are patients or from the general population.
A3c, A3d and A3e	Was the age of the child/young person, for whom respondents were asked to imagine health states to be valued, specified? If the age was specified, what was the age? Was the rationale for the choice of the age of child/young person provided?	The age of the child may have a strong influence on stated preferences and should be clearly described and reasons given for the choice. The use of broad terms such as 'young child' or 'toddler' without definition is ambiguous. For children the age may defined as 'the same age as you' or similar. As all these may influence stated preferences, rationale should be provided.
A4	Sample	
A4a	Was the population or sample frame defined from which the sample was drawn? (e.g., country, age, condition)	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. This is critical to understanding applicability of value sets or preferences.
A4b	Is information provided on how the sample was recruited (e.g., door knocking, location, online panel, convenience sample)?	The approach to recruitment will influence selection bias and generalizability and should be described. 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. The extent to which the approach taken would result in a representative sample of the intended population should be understood.
A4c	If data were collected online, were efforts made to avoid on-line panel fraud?	The use of on-line panels can attract fraudulent or bogus answers for example to gain 'rewards' for partaking in a survey. Answers may be indicative of inattentive or lazy responders to dishonest answers. Indications can include unrealistically short completion times and incorrect responses to screening questions.
A4d, A4e and A4f	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))? Was the target sample justified? Was the target sample achieved?	The sample size needs to be stated as it is important to understanding overall missingness. 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. The reasons for not achieving the target sample size should also be provided as this may influence representativeness.

A4g and A4h	Were the characteristics of the final sample described? Did the sample characteristics match the intended population?	The final characteristics of the sample are important when considering generalizability and potential selection bias arising from recruitment.
A4i	Was the year the data collected stated?	This question is needed to ensure there has not been an excessive time between valuation and publication.
A4j	Was information provided on missing data (non-completion, withdrawals)?	Missing data should be appropriately categorized, for example partial or non-completions.
Section B – What child HRQoL states were valued?		
B1	Type of study.	
B1	Value set or values for a limited number of health states (e.g. vignette)?	The distinction here is between studies that have developed a value set for a HRQoL instrument primarily for defining utility values in economic evaluations or similar, versus those that define a value for a specified health condition or specific health state(s).
B2	Value sets	
B2a	Which HRQoL instrument was valued?	References to development of the instrument should be included so that the details of how the instrument was originally developed can be ascertained.
B2b	Were the domains and response options of the instrument clearly described?	Domains and response levels should be clearly described without the need to refer back to development studies.
B2c	What experimental design approach was used to choose the health states (combination of dimension levels) to be valued?	In most cases it will not be possible to value every health state. Thus, the rationale for selection of the subset should be clear.
B2d	How were the health states assigned to respondents?	For example participants may have seen the same health states, randomly assigned to a select number of health states, or randomly assigned to different blocks of health states.
B3	Specific health states	
B3a	How were the health states described?	For example, a vignette may be used to describe an individual affected by a particular condition or health states from a condition specific HRQoL instrument could be used.
B3b	How many health states were preferences elicited?	This should be clearly reported with reasons. For example, utility values may be developed for health states describing differing severity of a disability or condition.
B3c	Was the rationale for the selection of these health states described?	Selection may be limited by the preference elicitation method or the research question and objectives or for pragmatic reasons. It should be clearly linked to the objectives of the study.
Section C – What methods were used to elicit stated preferences for child HRQoL?		
C1	Which methods were used to elicit stated preferences?	All methods used should be identified. For example a DCE may have been used in combination with a TTO or SG to place preferences onto a utility scale, for values worse than death or for comparative purposes.
C2 and C2a	Was a rationale for the choice of method provided? If yes what was the rationale?	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.
C3, C3a and C3b	Was the duration of the states to be valued reported (e.g. x years in this state followed by death)? Was the duration fixed? What durations(s) were used?	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 A2b and A3b).
C4	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 value set.
C4a	How were values < 0 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. All of which may give different values for the worst health state.
C4b	What was the minimum value possible?	The minimum value possible will vary with the method used and should be clearly stated.
C4c	What determined how the task was terminated?	A description of how tasks were terminated particularly where it proved difficult to reach indifference.
C5	How were values anchored on a utility scale?	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.
C6	What was the mode of administration for the stated preference tasks?	There are a number of ways that the tasks could be administered ranging from fully self-completed to in person interviews. This is particularly relevant to the more difficult tasks such as TTO and SG and when participants are children.

C7	How was the quality of stated preference data assessed?	Criteria for assessing quality and exclusions should be clearly defined. This may or may not include consistency and dominance checks recognizing that these may not be considered appropriate for DCEs and BWS surveys. Refer also to item A4c for online data.
C8 and C8a	Were any exclusions made to the preference data? Were reasons for exclusions provided?	Exclusions may have been made to enable assessment of average preferences.
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.
C9	Were the health states randomly assigned?	See Item B2d. The potential for bias should be addressed where tasks were not randomized.
C10	Was ethics approval for the study obtained from an appropriate research ethics committee?	Given involvement and recruitment requires appropriately informed consent, involvement of children and the potential for distress arising from the tasks, ethics approval should be expected.
C11	Were sources of funding and non-monetary support and the role of the funder(s) in the design described?	Conflicts of interest are applicable to stated preference studies given the utility values generated may be used to support interventions in and decisions for public funding.
<u>Section D – Econometric and statistical methods</u>		
D1	Value set or values for a limited number of health states?	Analytical requirements will vary depending on whether the study objective is to produce a complete 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	
D2a	What was the theoretical model? OR What models were estimated? e.g. OLS, Tobit etc.	There are many theoretical approaches that can be taken to the modelling for development of values sets (i.e. there is no standardized approach). It is important that this is clearly stated and justified.
D2b	Were the main assumptions of the model stated? (e.g. assumptions about preference homogeneity/heterogeneity)	The assumptions underpinning the model are critical as they will have a significant impact on the value set.
D2c	How was the constant term treated?	The constant term may be handled differently for a disutility model (e.g. set at 1) than for a utility model.
D2d	How were missing data handled?	There should be a clear description of handling of missing data given there are a number of approaches that ca be used. Implications with respect to the final data set for analysis should be understood. For example, complete case analysis may affect representativeness.
D2e	Were subgroup analyses completed?	Subgroup analyses may be undertaken as part of an assessment of preference heterogeneity.
D2f and D2g	Were interaction terms included? Were details of the interactions provided?	Interaction terms may be included to explore influence of 'most' and 'least' dimension scores in developing value sets. If included there should be sufficient detail provided to understand what interactions were considered and how they were modelled.
D2h	Were non-linear specifications considered?	If a non-linear functional form was considered, then the specifications evaluated should be described.
D2i, D2j, Dk and Dl	Was more than one model described? Were goodness-of-fit statistics for each model reported? Was the preferred model clearly stated? Were the criteria used to select the preferred model described?	Rationale for each model should be given and include criteria for identifying the preferred model. Goodness-of-fit statistics should be reported for all models and reference back to criteria for model selection. It needs to be clear which model formed the basis of the value set. Was it based solely on goodness of fit criteria, or modelled versus observed or a combination? Model selection might also take into account other aspects such as prior qualitative studies in development of the instrument and the valuation study.
D2m, D2n, D2o	Do the preference parameters for the health states follow a logical order (monotonic)? Was any post estimation undertaken to force monotonicity? How were insignificant differences between adjacent levels managed?	Inconsistencies should be clearly described including insignificant parameters. 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.
D2p	Were robustness checks were conducted?	Should be described in methods section and reported in appropriate detail.
D2q	Was uncertainty around the values reported?	Uncertainty should be considered, 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 D3g 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	How were missing data handled?	There should be a clear description of handling of missing data given there are a number of approaches that ca be used. Implications with respect to the final data set for analysis should be understood. For example, complete case analysis may affect representativeness.
D3d and D3e	Have subgroup analyses and interactions been undertaken? Were sub-groups or interaction variables chosen for assessment justified?	Sub-group analyses may be undertaken to evaluate differences in preferences/values. Interactions may be relevant where multiple health states are included. 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.
<u>Section E Characteristics and validity of values</u>		
E1	Was there qualitative or quantitative evidence reported that demonstrates the extent to which 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	Where a value was reported, were the values generated by the final model logically consistent?	Inconsistencies would suggest that the final model may not be appropriate for deriving the value set. This needs to be discussed.
E3	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. This would be best demonstrated graphically.
E4	Key characteristics of the values	Where the distribution of values has not been provided, E5a to E5c 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.
E4a	How many values less than zero were possible?	This is in addition to the average values and provides an indication of variability/uncertainty in preferences for values worse than dead.
E4b	What was the maximum possible value less than one?	This is particularly relevant to elicitation methods that cannot value the full health state and rely on one that is close to full health.
E4c	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.
E5	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.

Reference for figure:

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

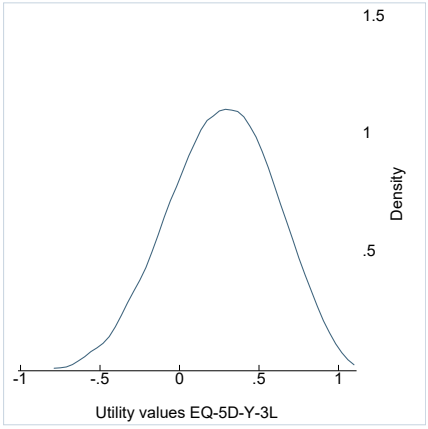

Figure S1 Density plot of theoretical values for EQ-5D-Y-3L value sets, where the utility value is on the x-axis and density on the y-axis

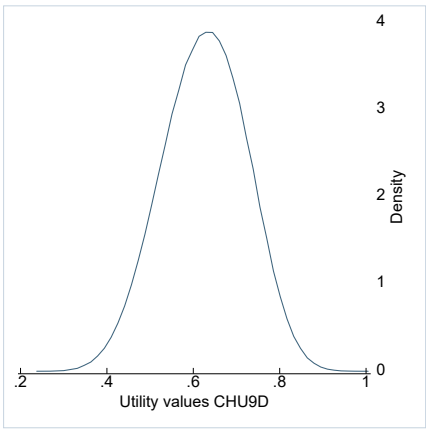

Figure S2 Density plot of theoretical values for CHU9D value sets value sets, where the utility value is on the x-axis and density on the y-axis

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

Formatted version of the long RETRIEVE

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

A1 – Stated preferences

A1a	Whose preferences were sought?		
	Adults	☐	*After A1b go to A2*
	Children and young people (CYP) <18 years	☐	*After A1b go to A3*
	Mixed adults and CYP	☐	*After A1b complete A2 and A3*
A1b	Did the authors provide a rationale for whose preference were sought?		
	Yes	☐	
	No	☐	

A2 Adults' stated preferences

A2a	Which adults were the focus of preference elicitation?		
	General population	☐	
	Parent or caregiver of child	☐	
	Health care professionals	☐	
	Adult with a health condition	☐	
	Other adults, please specify	☐	
A2b	What perspective were adults asked to take when 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	☐	
	Person 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	☐	Go to A2d
	No	☐	Go to A4
	Not applicable	☐	Go to A4
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 children, 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	☐	
	Another child they know	☐	
	A hypothetical child	☐	
	Other, please specify:	☐	
A3c	Was the age of the child, for whom respondents were asked to imagine health states to be valued, specified?		
	Not applicable (i.e. own perspective/themselves)	☐	Go to A4
	It was applicable but not stated	☐	Go to A4
	Yes	☐	Go to A3d
A3d	If yes, what was the age of the child?		
A3e	Was the rationale for the choice of the age of child 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., doorknocking, location, online panel, convenience sample)?		
	Yes	☐	
	Partial	☐	
	No	☐	
A4c	If data were collected online, were efforts made to avoid on-line panel fraud?		
	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	☐	Go to A4g
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	☐	Go to A4i
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?	☐	*Go to B2*
	Values for a limited number of health states (e.g. vignette)?	☐	*Go to B3*

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?	

B2 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?	
	No	☐
	Yes – What was the rationale?	☐

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	☐
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	☐ *Go to C4*
C3a	Was the duration fixed?	
	Yes	☐
	No	☐
C2a	What duration was used?	
C4	Did the method(s) allow values to be elicited that were < 0 ('worse than dead')?	
	Yes	☐ *Go to C5*
	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 (eg used to represent average preferences)?	
	Yes	☐
	No	☐ *Go to C9*
	Unclear	☐ *Go to C9*
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?	☐ *Go to D2*
	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? 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?	
D2e	Were subgroup analyses completed?	
	Yes	☐
	No	☐
	Not applicable	☐
D2f	Were interaction terms included?	
	Yes	☐
	No	☐ *Go to D2h*
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	☐ *Go to D2m*
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	☐ *Go to D2p*
	No	☐
D2n	Was any post estimation undertaken to force monotonicity (e.g. collapsing levels)?	
	Yes	☐
	No	☐
	Unclear/not stated	☐
D2o	How were insignificant differences between adjacent levels managed (e.g. collapsed/ forced to be different)?	
D2p	Wer 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	☐ *Go to D3c*
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	☐ *Go to D3f*
D3e	Were sub-groups and interaction variable chosen for assessment justified?	
	Yes	☐
	No	☐
D3e	Were sensitivity analysis undertaken?	
	Yes	☐
	No	☐ *Go to Module E*
D3f	Were sensitivity analyses described?	
	Yes	☐
	No	☐

Section E - Characteristics of values

E1	Was there qualitative or quantitative evidence reported that demonstrates the extent to which respondents engaged with and understood the valuation tasks?	
	Yes	
	No	
E2	Where a value was reported, were the values generated by the final model logically consistent?	
	Yes	
	No	
	Unclear	
E3	Did authors report the distribution of values over all states defined by the HRQoL instrument	
	Yes	
	No	

E5 Key characteristics of values

E4a	How many values less than zero were possible?	
E4b	What was the maximum possible value less than one?	
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	