
Article:
Bailey, C., Martin, H., Raghunandan, R. et al. (9 more authors) (2024) The RETRIEVE checklist for studies reporting the elicitation of stated preferences for child health-related quality of life. PharmacoEconomics, 42 (4). pp. 435-446. ISSN 1170-7690

https://doi.org/10.1007/s40273-023-01333-z

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 — *A1b then A2* ☐ Children and young people (CYP) <18 years — *A1b then A3* ☐ Mixed adults and CYP — *A1b then 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 A4 ☐ Yes
A3d	If the age was specified, what was the age?_____
A3e	Was the rationale for the choice of the age of child/young person provided? ☐ Yes ☐ No
A4 Sample	
A4a	Was the population or sample frame defined from which the sample was drawn? (e.g., country, age, condition) ☐ Yes ☐ No
A4b	Is information provided on how the sample was recruited (e.g., field-based recruitment, online panel, convenience sample)? ☐ Yes ☐ Partial ☐ No
A4c	If data were collected online, were efforts made to avoid on-line panel fraud? (eg, related to bots or automated software posing as participants and completing surveys) ☐ Yes ☐ No ☐ Not applicable
A4d	Was there a target sample size (or sample sizes if by block – e.g. number of tasks per block (e.g. DCE) or health state (e.g. TTO))? ☐ Yes ☐ No 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 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 ☐ 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? ☐ 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.	
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 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, shown below)** ☐ Yes ☐ No
E4	**Key characteristics of the values**
E4a	How many values less than zero were possible? _____
E4b	What was the maximum possible value less than 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

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

Figure 1 of Pan et al. (2022)

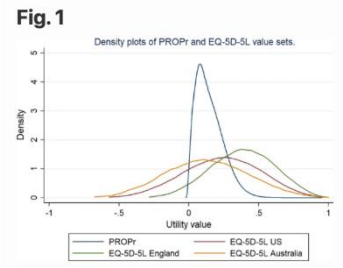

Table S2 – Examples of the use of long and short forms

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] *A1b then A2* ☐ Children and young people (CYP) <18 years *A1b then A3* ☐ Mixed adults and CYP *A1b then 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 A4* ☐ Yes
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., field-based recruitment, 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? (eg, related to bots or automated software posing as participants and completing surveys) ☐ 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]	*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. 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 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 in Pan et al 2022.)** ☐ 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] *A1b then A2* ☐ Children and young people (CYP) <18 years *A1b then A3* ☐ Mixed adults and CYP *A1b then 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 A4* ☐ Yes
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 field-based recruitment, 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? (eg, related to bots or automated software posing as participants and completing surveys) ☐ 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 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. *A1b then A2* ☐ Children and young people (CYP) <18 years *A1b then A3* ☐ Mixed adults and CYP *A1b then 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 A4* ☐ Yes
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., field-based recruitment, 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? (eg, related to bots or automated software posing as participants and completing surveys) ☐ 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 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:

			Location
MODULE A		**Stated preferences considered relevant to valuing child HRQoL and sample characteristics**	
	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 — *A1b then 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. — *A1b then A3* ☐ Mixed adults and CYP — *A1b then A2 and A3*
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 A4* ☐ 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 [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., field-based recruitment, 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 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., field-based recruitment, 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. We note that in TTO it is explicit that the respondent thinks the state is worse than dead; however, in DCE you can estimate values below 0 but the participant is not aware that they have made that choice.
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. What determined how the task was terminated (e.g. the decision rule for determining when a point of indifference has been reached in TTO and SG tasks).
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.
Section D – Econometric and statistical methods		
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.
Section E Characteristics and validity of values		
E1	Was 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.
E6	Did the authors report on specific requirements of users and decision makers about how such values are produced? e.g., as set out in the methods guides of local HTA bodies.	It is important that authors are cognizant of adapting the checklist for their local context by referring to relevant methods as per local health technology assessment guidelines.

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

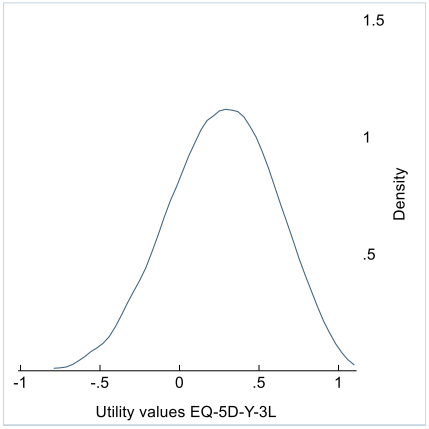

Figure S1 Density plot of theoretical values for EQ-5D-Y-3L value sets, where the utility value is on the x-axis and density on the y-axis

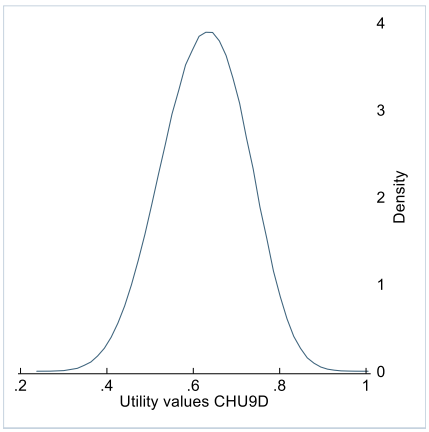

Figure S2 Density plot of theoretical values for CHU9D value sets value sets, where the utility value is on the x-axis and density on the y-axis