BMJ Open

Developing a preference-based measure for weight-specific health-related quality of life in adolescence: the WAItE UK valuation study protocol

Tomos Robinson ![ORCID], Sarah Hill, Yemi Oluboyede

Population Health Sciences Institute, Newcastle University Faculty of Medical Sciences, Newcastle upon Tyne, UK

Correspondence to
Dr Tomos Robinson;
tom.robinson@ncl.ac.uk

ABSTRACT

Introduction Adolescent obesity is a public health problem in the UK. The Weight-Specific Adolescent Instrument for Economic Evaluation (WAItE) has been developed as the first weight-specific health-related quality of life measure appropriate for economic evaluation, but currently cannot be used to generate quality-adjusted life years (QALYs), which are the basis of cost-utility analysis. Generic measures (such as the EQ-5D-Y or CHU-9D) may be insensitive to small but important health changes in overweight or obese adolescents. This study aims to generate a preference-based scoring algorithm for the WAItE.

Methods and analysis A discrete choice experiment (DCE) will be administered to value health states described by the WAItE classification system. These health states will be presented to members of the adult general population of the UK via an online survey. A range of regression models will be used to produce the utility algorithm for the WAItE. The DCE-visual analogue scale and time trade-off (TTO) anchoring methods will be used anchor the value set on to the 0–1 QALY scale.

Ethics and dissemination The Newcastle University Medical School Ethics Committee approved the study (references 4772/2020 (DCE) and 9978/2020 (TTO)). The developed algorithm can be applied to future economic evaluations of weight management interventions and treatments for adolescents.

Strengths and limitations of this study

► We have developed an extensive protocol to undertake a valuation of the Weight-Specific Adolescent Instrument for Economic Evaluation using an online discrete choice experiment (DCE).

► The DCE-visual analogue scale and time trade-off (TTO) anchoring methods will be used to anchor the generated preference value set on to the 0–1 quality-adjusted life year (QALY) scale, and the differences in the value sets generated using these different methods will also be evaluated.

► The generated utility values can be used to estimate QALYs for cost-utility analysis of new treatments or interventions that are aimed at adolescents in the context of obesity.

► There are several potential sources of bias for the DCE survey, including the hypothetical nature of the health states that will be valued and online data collection methods.

► There are also several potential sources of bias for the TTO study, including the cognitive difficulty of the exercise and framing issues.

that are seen to be affected by specific health conditions and are therefore in theory more sensitive to disease-specific improvements in HRQoL.

To estimate quality-adjusted life years (QALYs), a key metric used in cost-utility analysis (CUA), health utility weights anchored at 0 (a state equivalent to being dead) and 1 (full health) are needed. These health utility weights can be generated from PROMs that have an associated utility 'value set' that provides a value for each of the health state states described by the PROM's classification system.[3]

There are substantial levels of child and adolescent obesity in the UK,[4] which is likely to track into adulthood[5] and have immediate and long-term implications for health and well-being.[6] Given the lack of a weight-specific adolescent measure of HRQoL

INTRODUCTION

Healthcare interventions, treatments and policies impact patients' health-related quality of life (HRQoL). Information regarding a patient's HRQoL is usually obtained using a patient-reported outcome measure (PROM), which can be further classified as either 'generic' or 'condition-specific'.[1] Generic PROMs, such as the EQ-5D,[2] are designed to cover a general spectrum of health-related problems, and therefore are in theory applicable for any health condition. However, these generic measures may not capture important aspects of HRQoL specific to certain health conditions. In contrast, condition-specific PROMs focus on the domains of HRQoL

measure suitable for economic evaluation, the **W**eight-Specific **A**dolescent **I**nstrumen**t** for **E**conomic Evaluation (WAItE) was developed.[7] The WAItE is a brief, seven-item PROM, which incorporated the views and experiences of adolescents aged 11–18 years in its design. Crucially, the WAItE contains several weight-specific aspects of HRQoL that are not usually captured in generic measures, such as 'embarrassment' and 'being treated differently'. The WAItE has been shown to have good psychometric properties,[8] as well as good levels of concurrent validity and test–retest reliability.[9]

However, although specifically designed to be preference based (and therefore applicable for use in a CUA), a value set for the WAItE classification system has yet to be developed. A statistical algorithm has been developed to calculate QALYs indirectly by 'mapping' the WAItE to the generic, preference-based Child Health Utility-9 Dimension (CHU-9D),[10] however, this is generally seen as being a 'second best' solution compared with applying direct preference weights through the use of valuation study.

Given this background, the aim of this study is to generate a preference algorithm which can then be used to generate a value set for the WAItE. This value set will allow utility values to be calculated from the responses to the WAItE questionnaire, enabling the derivation of QALYs and thus enabling the use of the WAItE in the context of CUA. This will expand the use of CUA in the economic evaluation of new treatments or interventions aimed at adolescents in the context of obesity.

It should be noted that relatively few valuation studies for condition-specific, preference-based measures in paediatric populations have previously been undertaken,[11–14] with the majority of valuation studies for this population having been conducted in relation to generic preference-based measures. The remainder of this paper describes the methods that will be used in this study to generate a preference algorithm for the WAItE, paying particular attention to the methodological considerations outlined in the comprehensive recent review by Rowan *et al*.[15]

Aims
1. To value health states generated from the WAItE classification system with a general population sample
2. To compare the WAItE value sets generated using the discrete choice experiment-visual analogue scale (DCE-VAS) and time trade-off (TTO) anchoring methods

METHODS AND ANALYSIS
Methodological considerations
Whose preferences and what perspective
In developing a value set for child preference-based measures, preferences may be gathered from several different groups of individuals, including: adults, children and 'informed' samples such as healthcare professionals. Furthermore, recent research has shown that preferences gathered from different populations may differ.[16] In this

study, we will collect data from a representative sample of the UK adult general population. It can be argued that adults will have a greater understanding of preference elicitation tasks as compared with children and may be better informed about the impact of ill health. Furthermore, as adult preferences are usually used to generate value sets for preference-based measures for adults, using adult preferences for child and adolescent health states provides a level of comparability in methodology across all age groups. Adult general population preferences have used to generate value sets for several generic preference-based measures of HRQoL for children and adolescents.[17–23]

When using an adult general population sample, there are several 'perspectives' that an adult can answer from, including their own health as an adult, the health state for themselves as a child and the health state in the context of a child at a specified age. In this study, the adult general population sample will be answering from their own perspective. It can be argued that when answering from this perspective, the adult respondent is under a 'veil of ignorance', meaning that their views should not be influenced by any views they may have about child health. Furthermore, using the adult's own perspective again provides a level of comparability with the methods usually used generate value sets for adult preference-based methods. Adult population preferences elicited using their own perspective have been used in several previous studies.[17 19 21 22] Furthermore, this perspective was selected as the clearest and easiest to understand by the current study's patient and public involvement (PPI) group during the piloting process.

It should be noted that the choice of whose preferences to use and which perspective to take when valuing child preference-based measures is a matter of normative debate, and that the different preferences and perspectives each have advantages and disadvantages. For example, critics of using adult preferences in this situation may argue that adults are not fully informed, do not necessarily have experience of the health states in question and are not potential users of healthcare services for children. As noted by Rowen *et al*,[15] there is currently limited guidance regarding the most appropriate methods to use, and further research is needed in several areas to inform best practice guidelines.

Elicitation technique and mode of administration
Several different methods can be used to elicit utilities for health states and generate a value set. From a theoretical standpoint, the standard gamble (SG) and TTO methods are regarded to be the most appropriate method of eliciting utility under uncertainty.[24] The SG method involves an individual being asked to choose between an assured, but impaired, health state option and a risky option in which the individual faces a chance of living in full health for a fixed amount of time or a chance of dying immediately. The TTO method involves asking the respondent to consider living in an impaired health state for a given

time period, compared with living in full health for a time period equal to, or less than, that of the impaired health state. The time in full health is adjusted until this option is equally as desirable as the fixed time-period in the impaired health state. For both the SG and TTO, a utility value for a health state can be calculated using information about the point of indifference between the two options.

However, both the SG and TTO have been criticised on several grounds. For example, the SG method relies on a respondent's ability to interpret relatively complex probabilities, while answers to TTO tasks may be biased because individuals are likely to have different discount rates (the value of health in future years compared with the same health at the current time). Aside from general problems related to the framing and cognitive burden of the SG and TTO tasks, there are several other weaknesses for using these methods to generate a value set for HRQoL tool. For example, conducting SG and TTO tasks face-to-face is extremely time and resource intensive, and it has been argued that these measures are particularly sensitive to the mode of administration.

Given the problems associated with the methods discussed above, in this study, a DCE will be used to apply utility weights to the dimensions and severity levels of the WAItE. DCEs describe a health state in terms of its characteristics (or 'attributes'), which can be dimensions of health and quality of life more generally (eg, pain, tiredness). The extent to which an individual values each health state is expected to depend on the 'level' of these characteristics, which can be seen as the relative severity or frequency level of the dimension (eg, never, sometimes and always). DCEs present individuals with two (or more) health state options and each health state is described in terms of a set of attributes and associated levels. Individuals then choose which option they prefer.

It is worth noting that the best-worst scaling (BWS) method has also been used to value preferences for some adolescent preference-based measures,[25][26] and a recent study has argued that BWS may be easier to understand for adolescents due to its relative cognitive simplicity.[27] However, previous research has found that adult respondents find the DCE tasks preferable and easier to complete than BWS tasks[28][29] and other studies have raised concerns about using BWS for valuation exercises.[30][31]

In terms of the mode of administration, studies using adult preferences have previously been conducted using both face-to-face interviews and using online surveys. For ease of recruitment, this sample will be recruited from an online panel managed by a survey and market research company. The online panel consists of individuals who have opted in to participate in online surveys in return for a small level of compensation. An online panel is a cost-effective way of recruiting respondents and has been widely used in general valuation studies.[32] Although online data collection may be more susceptible to data quality issues as compared with other methods, it has

been noted that other data collection methods may also have issues related to data quality.[33]

Anchoring

One disadvantage of methods such as DCEs in the context of health state valuation is that the results generated are on a latent scale, meaning that the resulting preferences are measured in terms of unanchored utility with no units.[34] Therefore, these results in isolation do not provide estimates of utility values that are ranked on the 0–1 QALY scale. A number of methods may be used to anchor the latent utility estimates generated from DCEs onto the 0–1 QALY scale, including using a TTO value for the worst health state as defined by the classification system,[35] including a 'dead' state as part of the DCE tasks,[36] conducting a DCE task with a duration attribute included[37] or a hybrid approach involving a combination of DCE and TTO tasks.[38]

In this study, we will use two different anchoring techniques. The first technique that we will use is the recently forwarded 'DCE-VAS' method.[34] This method involves the study participants rating three health states using the VAS: the best health state possible from the WAItE (1111111); the worst health state possible from the WAItE (5555555) and 'dead', and then using these responses to reweight the latent scale results from the DCE. These VAS states will be collected as part of the online survey.

The second technique we will use is a TTO anchoring study as described by Ratcliffe *et al*.[35] This method involves gathering a TTO value for the worst health state possible from the WAItE classification system (i.e, the PITS state, 5555555), and then using this value to reweight the latent coefficient estimates from the DCE. This method has previously been used to anchor latent utility estimates to the 0–1 QALY scale for the CHU-9D in Australia and China.[25][26] The TTO will not form part of the online survey, and will be conducted as a one-to-one interview via a video call. The TTO study is discussed in detail in 'TTO anchoring study' section.

Discrete choice experiment

Each of the seven attributes in the WAItE classification system consists of five frequency of occurrence levels (never, almost never, sometimes, often and always). The full WAItE questionnaire is shown in figure 1. The choice sets for the DCE survey will each contain two health states to be compared; each health state will comprise all seven attributes, each described using one of the five frequency levels. The health states will include a mixture of levels across the seven attributes.

DCE experimental design and sample size

If we were to present a full combination of attributes and levels (known as a 'Full Factorial' design), this would generate 78,125 (5^7) possible scenarios and 3,051,718,750 possible combinations of two-alternative choice questions ($5^7*(5^7-1)/2$). It is clearly infeasible to present all combinations of attributes to the participants, therefore,

INSTRUCTIONS:

The questions below ask about different things in your life. Read through each question carefully

and **TICK ONE RESPONSE THAT BEST DESCRIBES YOU AT THE MOMENT**

EXAMPLE:

If you 'sometimes' have body pain then you should tick this box:

	Never	Almost Never	Sometimes	Often	Always
I have body pain			✓		

Now please read through each of the 7 questions below and choose <u>ONE</u> answer that best

describes you at the moment

	Never	Almost Never	Sometimes	Often	Always
1. I get tired					
2. I struggle to keep up when I am walking around with others					
3. I avoid doing sports					
4. I struggle to concentrate on my studies/work					
5. I feel embarrassed shopping for clothes					
6. I feel unhappy because I am unable to do the same things as others					
7. People treat me differently when I go out					

Figure 1 WAItE questionnaire. WAItE, Weight-Specific Adolescent Instrument for Economic Evaluation.

a subset of the health states will be used in the DCE survey. To optimise the efficiency of the experimental design, we will use a Bayesian D-efficient design.[39] The Bayesian prior distributions used to inform the design will be taken from pilot work undertaken as part of one of the authors' PhD studies.[40] In this pilot study, 171 individuals recruited from an online survey company completed 10 DCE choice tasks with different WAItE health states. The design will be generated using the Ngene software package,[41] and the relative design efficiency (as measured by the D-efficiency statistic)[42] will be used to select the set of pairwise choices used in the final analysis. The target sample size for the DCE study is 1,000 respondents.

It has become standard practice to ask respondents to complete between eight and 12 choices sets (each made up of two different health-state options) when using a DCE to value health states generated from a multiattribute instrument. This range has been considered to maximise data collection per respondent without incurring significant responder bias which could undermine the quality of the data.[14] Given this guidance, 12 choice sets will be presented to each respondent as part of the main DCE survey. The respondent will also complete a practice DCE task at the beginning of the main survey, and an additional 'fixed pair' at the end of the main survey, which will present a logically dominant choice. This logically dominant choice will be used as a dominance test to examine data quality, as a large proportion of respondents failing to choose the dominant option could be considered a sign of poor data quality. The impact of excluding the individuals who fail the dominance test from the estimation sample will be explored as part of the data analysis.

DCE sample recruitment and survey

The respondents will be recruited from an existing online panel of residents of the UK administered by *SurveyEngine*, a survey company with expertise in designing and hosting online DCEs. The data collected will be anonymous as no personal identifiable information will be provided to the researchers.

The online survey will contain several sections. At the beginning of the questionnaire, respondents will be

initially introduced to the research study. A description of the rationale and objectives of the study will be provided. Consent will be obtained by respondents after they have read the conditions outlined on the consent page.

In the first section of the survey, the participant will complete the DCE questions. In the second section of the survey, the respondents will complete feedback questions related to how much they understood the DCE and how difficult they found the DCE questions. In the third section of the survey, the respondent will complete the WAItE and the EQ-5D-5L, as well as the VAS questions that will be used to anchor the latent DCE estimates to the 0–1 QALY using the EQ-VAS anchoring method. In the final section of the survey, the participant will be asked to answer several demographic questions, including age, gender, marital status, income level, employment status, education level and self-assessed health.

DCE pilot study
In the pilot phase, a soft-launch of the survey will recruit a 10% sample (n=100) of the target sample size. This phase will start in September 2021. If no changes are made, the data collected during the soft launch will be included in the final data set. The pilot data set will be collected to:
1. Check that the experimental design of the DCE is fit for purpose
2. Check respondents' perception of the difficulty, clarity and understanding of the survey. Particularly, responses to the questions included after the DCE survey (related to difficulty in answering and understanding the questions) will be analysed.
3. Assess the amount of time spent by the respondents on each part of the survey, particularly the DCE tasks and the EQ-VAS questions. This information will be used to assess the level of respondent burden.

TTO anchoring study
As discussed previously, in addition to the DCE-VAS anchoring method (where data will be collected during the same online survey as the DCE), we will also conduct a virtual, face-to-face anchoring study using the TTO method external to the online DCE survey.

TTO interview
Due to precautions imposed by the COVID-19 pandemic, the TTO interviews will be conducted virtually via a video teleconferencing medium. The sample for the TTO will be gathered by the survey company *Dynata*. In an online screening interview, respondents will complete a set of background demographic questions including age, gender, marital status, income level, employment, education level and whether they have any children. They will then consent to be contacted by email by one of the study team to organise a virtual TTO interview at a time convenient to them.

At the start of the interview, respondents will be introduced to the research study. A description of the rationale and objectives of the study will be provided. Either

written or verbal informed consent will be obtained from respondents via a consent form discussed with the participants at the beginning of the interview.

Prior to the TTO exercise, respondents will complete the WAItE questionnaire for themselves to introduce them to the wording and formatting of the WAItE classification system. A series of TTO tasks will then be conducted using an interviewer-administered mode of administration with a visual prop (ie, a virtual TTO board which will be presented via a screen-sharing function in the video teleconference) in the form of a sliding scale to represent life years. The respondents will be asked to value four WAItE health states in total, covering a variety of levels of HRQoL impairment ('high': 2212122, 'medium': 2234 442 and 'low': 4445555), and the PITS state (5555555). The valuation of health states other than the PITS state act as 'warm up' tasks to ensure that the respondent is comfortable with the format of the TTO task before valuing the state required for anchoring.

There are several issues with the conventional TTO in this context, including the use of this technique when health states may be considered worse than dead.[43] The lead time TTO and lag time TTO are two variants of the TTO that have previously been used in the literature to account for these difficulties.[44–46] The precise type of TTO that will be presented to the respondents when a state is considered worse than dead will be determined during the pilot period.

The respondents will complete a small number of postinterview questions related to how difficult they found the TTO. The respondents will then be thanked for completing the survey and given contact details where they can contact the research team if they have any questions about the study.

Inclusion criteria
Due to ethical concerns relating to the presentation of scenarios involving death and the cognitive complexities of the TTO task, like the DCE study, the target population for the TTO study will be a representative sample of the adult UK general population.

Sample size
Given the resource intensive nature of collecting face-to-face TTO data, the target sample for the TTO study will be substantially less than the online survey. When developing the value sets for the Australian and Chinese versions of the CHU-9D, Ratcliffe *et al*[25] and Chen *et al*[26] both used convenience samples of 38 undergraduate students for their external TTO survey. Given these previous sample sizes, our target sample size will be 40 respondents.

Pilot testing
Like the online DCE survey, the TTO survey will also be piloted. First, the questionnaire will be tested with members of the project PPI group, specifically focusing on the wording and ease of completion of the candidate TTO variants (ie, standard, lag time and lead time). Once

the wording of the materials has been refined, pilot TTO interviews will be conducted with both members of the PPI group and members of the Health Economics Group at Newcastle University. The data generated from these pilot interviews will be checked for internal consistency, to ensure that respondents fully understand the task. If necessary, the materials will be further refined and tested again prior to the definitive interviews with the study participants.

Patient and public involvement

The project PPI group have been consulted regarding several aspects of the project to date, including the refinement of the materials for the DCE study. The PPI group will continue to be consulted throughout the project, for example, to refine the materials and pilot the materials for the TTO study.

Analytical plan

To determine the coefficients for the main effects from the DCE, several regression models will be fit to the data collected from the DCE survey, including the multinomial logit model[47] and the mixed logit model.[48] Different choice models are obtained from different assumptions about the distribution of the random terms. The final model selection will be determined based on model fit, as measured by the log likelihood ratio or McFadden's pseudo R^2.[47]

As previously mentioned, to reweight the latent coefficients calculated from the DCE onto the 0–1 QALY scale, both the EQ-VAS anchoring technique and the TTO anchoring technique will be used. For the EQ-VAS anchoring technique, the latent coefficients from the DCE will be multiplied through by a constant term Ω :

$$\Omega = \frac{1 - \mathrm{Mean}\widetilde{\mathrm{VAS}}_{5555555}}{\sum_{j=1}^{7} \beta_{j5}} \tag{1}$$

β_{j5} represents the level 5 latent scale coefficients for each of the seven WAItE dimension. $\widetilde{\mathrm{VAS}}_{5555555}$ can be calculated using the equation:

$$\widetilde{\mathrm{VAS}}_{5555555} = \frac{\mathrm{VAS}_{5555555} - \mathrm{VAS}_{\mathrm{Dead}}}{\mathrm{VAS}_{1111111} - \mathrm{VAS}_{\mathrm{Dead}}} \tag{2}$$

$\mathrm{VAS}_{1111111}$, $\mathrm{VAS}_{5555555}$ and $\mathrm{VAS}_{\mathrm{Dead}}$ represent the individual rankings of the best healthstate possible in the WAItE classification system (1111111), the worst health state possiblein the WAItE classification system (5555555) and the dead state using the VAS.

For the TTO anchoring technique, the latent coefficients will be recalled by dividing through by a constant term λ:

$$\lambda = \frac{(\mathrm{LatentDCE}_{5555555} - 1)}{(\mathrm{MeanTTO}_{5555555} - 1)} \tag{3}$$

where $\mathrm{LatentDCE}_{5555555}$ represents the latent value of the worst state possible in the WAItE classification system (5555555), and $\mathrm{MeanTTO}_{5555555}$ represents the mean value of the same health state calculated using the TTO.

The value sets generated from the two different anchoring methods will then be compared in terms of the value of the worst possible health state from the WAItE classification system (5555555), the proportion of health states valued worse than death and the value of the health state with the smallest utility decrement from the best possible health state from the WAItE classification system (1111111).

The data will be analysed in Stata and Microsoft Excel.[49 50] If there are any incomplete or missing data, methods such as multiple imputation will be considered.[51] The final decision regarding this will depend on the type and pattern of missing data.

Ethics and dissemination

There are no known health and safety risks associated with participants in the DCE or TTO studies. At the beginning of both the DCE and TTO studies, there will be an information sheet which will ensure that the participant is fully informed about the aims of the study, and a consent form which will allow participants to enter the survey voluntarily. For the DCE study, anonymised data will be sent to the study team by *SurveyEngine*. For the TTO study, the data from the online screening questionnaire will be sent to the study team by *Dynata*. This will then be linked to the information collected in the TTO interview itself using a personalised study identification number. Data on the names and email addresses of the TTO interview participants will subsequently be destroyed after the data have been linked. All study data will be stored in a password protected folder on a Newcastle University computer.

The algorithm developed from this study can be used to generate health state utility values for any study using the WAItE, enabling the estimation of QALYs for CUAs of new treatments or interventions aimed at adolescents in the context of obesity. The distribution of the results of this study will be via publication in peer-reviewed journals and presentations at conferences. A user guide will also be developed, which will include instructions explaining how the algorithm can be used, including examples of statistical code.

ORCID iD
Tomos Robinson http://orcid.org/0000-0001-8695-9738

REFERENCES

1 Longworth L, Yang Y, Young T. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technol Assess* 2014;18:1–224.

2 Devlin NJ, Brooks R. EQ-5D and the EuroQol group: past, present and future. *Appl Health Econ Health Policy* 2017;15:127–37.

3 Brazier JE, Rowen D, Mavranezouli I, *et al*. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess* 2012;16:1–14.

4 HM Government. Childhood obesity: a plan for action, 2016URL. Available: www.gov.uk/government/publications/childhood-obesity-a-plan-for-action [Accessed 28 Sep 2021].

5 Singh AS, Mulder C, Twisk JWR, *et al*. Tracking of childhood overweight into adulthood: a systematic review of the literature. *Obes Rev* 2008;9:474–88.

6 Søltoft F, Hammer M, Kragh N. The association of body mass index and health-related quality of life in the general population: data from the 2003 health survey of England. *Qual Life Res* 2009;18:1293–9.

7 Oluboyede Y, Hulme C, Hill A. Development and refinement of the WAItE: a new obesity-specific quality of life measure for adolescents. *Qual Life Res* 2017;26:2025–39.

8 Oluboyede Y, Smith AB, Hill A, *et al*. The weight-specific adolescent instrument for economic evaluation (WAItE): psychometric evaluation using a Rasch model approach. *Qual Life Res* 2019;28:969–77.

9 Oluboyede Y, Robinson T. Measuring weight-specific quality of life in adolescents: an examination of the concurrent validity and test-retest reliability of the WAItE. *Value Health* 2019;22:348–54.

10 Robinson T, Oluboyede Y. Estimating CHU-9D utility scores from the WAItE: a mapping algorithm for economic evaluation. *Value Health* 2019;22:239–46.

11 Chiou C-F, Weaver MR, Bell MA, *et al*. Development of the multi-attribute pediatric asthma health outcome measure (PAHOM). *Int J Qual Health Care* 2005;17:23–30.

12 Stevens KJ, Brazier JE, McKenna SP, *et al*. The development of a preference-based measure of health in children with atopic dermatitis. *Br J Dermatol* 2005;153:372–7.

13 Kerr C, Breheny K, Lloyd A, *et al*. Developing a utility index for the aberrant behavior checklist (ABC-C) for fragile X syndrome. *Qual Life Res* 2015;24:305–14.

14 Bahrampour M, Norman R, Byrnes J, *et al*. Utility values for the CP-6D, a cerebral palsy-specific multi-attribute utility instrument, using a discrete choice experiment. *Patient* 2021;14:129–38.

15 Rowen D, Rivero-Arias O, Devlin N, *et al*. Review of valuation methods of preference-based measures of health for economic evaluation in child and adolescent populations: where are we now and where are we going? *Pharmacoeconomics* 2020;38:325–40.

16 Mott DJ, Shah KK, Ramos-Goñi JM, *et al*. Valuing EQ-5D-Y-3L health states using a discrete choice experiment: do adult and adolescent preferences differ? *Med Decis Making* 2021;41:584–96.

17 Feeny D, Furlong W, Torrance GW, *et al*. Multiattribute and single-attribute utility functions for the health Utilities index mark 3 system. *Med Care* 2002;40:113–28.

18 McCabe C, Stevens K, Roberts J, *et al*. Health state values for the HUI 2 descriptive system: results from a UK survey. *Health Econ* 2005;14:231–44.

19 Seiber WJ, Groessl EJ, David KM. *Quality of well being self-administered (QWB-SA) scale*. San Diego: Health Services Research Center, University of California, 2008.

20 Beusterien KM, Yeung J-E, Pang F, *et al*. Development of the multi-attribute adolescent health utility measure (AHUM). *Health Qual Life Outcomes* 2012;10:102–9.

21 Stevens K. Valuation of the child health utility 9D index. *Pharmacoeconomics* 2012;30:729–47.

22 Rowen D, Mulhern B, Stevens K, *et al*. Estimating a Dutch value set for the pediatric preference-based CHU9D using a discrete choice experiment with duration. *Value Health* 2018;21:1234–42.

23 Craig BM, Greiner W, Brown DS, *et al*. Valuation of child health-related quality of life in the United States. *Health Econ* 2016;25:768–77.

24 Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ* 1986;5:1–30.

25 Ratcliffe J, Chen G, Stevens K, *et al*. Valuing child health utility 9D health states with young adults: insights from a time trade off study. *Appl Health Econ Health Policy* 2015;13:485–92.

26 Chen G, Xu F, Huynh E, *et al*. Scoring the child health utility 9D instrument: estimation of a Chinese child and adolescent-specific tariff. *Qual Life Res* 2019;28:163–76.

27 Rogers HJ, Marshman Z, Rodd H, *et al*. Discrete choice experiments or best-worst scaling? A qualitative study to determine the suitability of preference elicitation tasks in research with children and young people. *J Patient Rep Outcomes* 2021;5:1.

28 Whitty JA, Walker R, Golenko X, *et al*. A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. *PLoS One* 2014;9:e90635.

29 Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Rev Pharmacoecon Outcomes Res* 2010;10:259–67.

30 Krucien N, Watson V, Ryan M. Is best–worst scaling suitable for health state valuation? A comparison with discrete choice experiments. *Health Econ* 2017;26:e1–6.

31 Krucien N, Sicsic J, Ryan M. For better or worse? Investigating the validity of best-worst discrete choice experiments in health. *Health Econ* 2019;28:572–86.

32 Wittenberg E, Prosser LA. Ordering errors, objections and invariance in utility survey responses. *Appl Health Econ Health Policy* 2011;9:225–41.

33 Ratcliffe J, Flynn T, Terlich F, *et al*. Developing adolescent-specific health state values for economic evaluation. *Pharmacoeconomics* 2012;30:713–27.

34 Webb EJD, O'Dwyer J, Meads D, *et al*. Transforming discrete choice experiment latent scale values for EQ-5D-3L using the visual analogue scale. *Eur J Health Econ* 2020;21:787–800.

35 Ratcliffe J, Brazier J, Tsuchiya A, *et al*. Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. *Health Econ* 2009;18:1261–76.

36 Brazier J, Rowen D, Yang Y, *et al*. Comparison of health state utility values derived using time trade-off, RANK and discrete choice data anchored on the full health-dead scale. *Eur J Health Econ* 2012;13:575–87.

37 Bansback N, Brazier J, Tsuchiya A, *et al*. Using a discrete choice experiment to estimate health state utility values. *J Health Econ* 2012;31:306–18.

38 Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Med Decis Making* 2015;35:328–40.

39 Bliemer MCJ, Rose JM, Hess S. Approximation of Bayesian efficiency in experimental choice designs. *Journal of Choice Modelling* 2008;1:98–126.

40 Oluboyede Y. *Quality of life assessment in adolescent obesity: Development of a new instrument for economic evaluation [dissertation*. The University of Leeds: Leeds, 2013.

41 ChoiceMetrics. *User Manual & Reference Guide.Ngene [program*. Sydney, Australia: ChoiceMetrics, 2014.

42 Kuhfeld WF, Tobias RD, Garratt M. Efficient experimental design with marketing research applications. *Journal of Marketing Research* 1994;31:545–57.

43 Tilling C, Devlin N, Tsuchiya A, *et al*. Protocols for time tradeoff valuations of health states worse than dead: a literature review. *Med Decis Making* 2010;30:610–9.

44 Robinson A, Spencer A. Exploring challenges to TTO utilities: valuing states worse than dead. *Health Econ* 2006;15:393–402.

45 Devlin N, Buckingham K, Shah K, *et al*. A comparison of alternative variants of the lead and lag time TTO. *Health Econ* 2013;22:517–32.

46 Augustovski F, Rey-Ares L, Irazola V, *et al*. Lead versus lag-time trade-off variants: does it make any difference? *Eur J Health Econ* 2013;14 Suppl 1:25–31.

47 McFadden D. *Conditional logit analysis of qualitative choice behavior*, 1973.

48 McFadden D, Train K. Mixed MNL models for discrete response. *J Appl Econ* 2000;15:447–70.

49 StataCorp. *Stata statistical software: release 16*. College Station, TX: StataCorp LLC, 2019.

50 Microsoft Corporation. Microsoft Excel, 2018. Available: https://office.microsoft.com/excel.tw

51 Rubin DB. Inference and missing data. *Biometrika* 1976;63:581–92.