

BMJ Open Child health valuation protocol for a discrete choice experiment comparing paired comparison and kaizen tasks and estimating US EQ-5D-Y-3L values on an experience scale

Maksat Jumamyradov,¹ Benjamin Matthew Craig ,¹ Oliver Rivero-Arias ,² Michał Jakubczyk³

To cite: Jumamyradov M, Craig BM, Rivero-Arias O, *et al.* Child health valuation protocol for a discrete choice experiment comparing paired comparison and kaizen tasks and estimating US EQ-5D-Y-3L values on an experience scale. *BMJ Open* 2023;**13**:e077256. doi:10.1136/bmjopen-2023-077256

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2023-077256>).

Received 29 June 2023

Accepted 02 October 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Economics, University of South Florida, Tampa, Florida, USA

²Nuffield Department of Population Health, National Perinatal Epidemiology Unit, University of Oxford, Oxford, UK

³Division of Decision Analysis and Support, SGH Warsaw School of Economics, Warsaw, Poland

Correspondence to

Dr Benjamin Matthew Craig;
bccraig@usf.edu

ABSTRACT

Introduction A decade ago, the first national valuation study of the EQ-5D-Y-3L (Y-3L) involved a discrete choice experiment (DCE) that asked 4155 US adult respondents to complete 40 paired comparisons, choosing between two dying children. Instead of choosing between dying children, the respondents in this novel protocol are asked whether 'being in a coma' is better or worse than experiencing 'health problems' (ie, experience scale) and how they would relieve health problems (ie, kaizen tasks). Our aims are to compare the preference evidence of the paired comparison and kaizen tasks and to conduct a DCE for the valuation of Y-3L profiles on an experience scale.

Methods and analysis Under this protocol, we will conduct an online survey that collects preference evidence from 600 US adult respondents on the health of a 10-year-old child for a week. Across all scenarios, each child will be described as either being 'in a coma' or having 'health problems', namely five three-level attributes (Y-3L). In this DCE, each respondent will be randomly assigned to one of four D-efficient blocks, including five coma comparisons (ie, Y-3L vs coma), 10 paired comparisons (Y-3L vs Y-3L) and 10 kaizen tasks (preference paths). In addition to comparing evidence by task (aim 2), the analysis plan includes the estimation of main-effects conditional logit models to create a Y-3L value set on an 'experience scale' where positive (negative) experiences have positive (negative) values (0 is 'being in a coma' and 1 is full health).

Ethics and dissemination The institutional review board (IRB) (Advarra) determined that this project (Pro00072276) is exempt from IRB oversight based on DHHS 45 CFR 46.104(d)(2) and is not subject to requirements for continuing review. The results will be prepared for publication in peer-reviewed journals and presented at scientific meetings. The data and code will be made available on reasonable request.

INTRODUCTION

Discrete choice experiments (DCEs) are often conducted to elicit preferences from individuals on health-related objects.^{1 2} Economists

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Due to the challenges of asking adult respondents about their preference on dying children (ie, intrinsic to quality-adjusted life-year; QALY), no scenario in this protocol describes the death of children and values are expressed on an experience scale, where positive (negative) experiences have positive (negative) values.
- ⇒ Instead of choosing between dying children like the original US EQ-5D-Y-3L valuation study, respondents will express their preferences on the sequential relief of a child's health problems (ie, kaizen tasks), which may be less distressing and more informative than paired comparisons.
- ⇒ In the protocol, the survey will have more respondents and more tasks than what is needed for a child health valuation study alone (ie, overpowered) so that we can compare paired comparison and kaizen tasks and simulate the effects of smaller samples for future studies.
- ⇒ The protocol relies on recruitment from an online panel, which may not be generalisable to the US general population, particularly to those individuals who are less educated and do not participate in online surveys.
- ⇒ Although EQ-5D-Y-3L values on an experience scale (0 is 'being in a coma' and 1 is full health for a week) will be informative generally, we recognise that they may not be suitable for use in cost-utility analyses presently.

generally assume that given a choice set and a hypothetical scenario, each decision-maker selects the alternative that maximises their decisional utility.³ In economics, choices are behaviours that imply inequalities in utility (eg, A>B) that resolve the ambiguities between objects.⁴ The design of a DCE lists the type, number and diversity of alternatives in the choice set, assuring the breadth of preference evidence to meet the study objective.

In health valuation, DCEs are designed and conducted to elicit preferences on health-related outcomes, such as being 'in a coma' or having various problems in health-related quality of life (HRQoL) domains, and then estimate their value on an anchored scale.⁵

The **EQ-5D-Y3L** (Y3L) is a generic, child-friendly self-complete instrument measuring five HRQoL domains in children and adolescents aged 8–15 years.⁶ Expressing the best level of each domain as 1 and the worst level as 3, the Y3L descriptive system has five three-level attributes and describes 243 profiles (3^5), ranging from 11 111 to 33 333. For example, the Y3L profile of a child with moderate problems on all five attributes is 22 222. In 2013, Craig and colleagues conducted the first Y3L valuation study. Their DCE (4155 US adult respondents; 40 paired comparisons) produced values for all 243 Y3L profiles on a **quality-adjusted life year** (QALY) scale.⁷ According to respondent feedback from this study and subsequent child health valuation studies, choosing between dying children caused some to feel guilty for abandoning one child, question their role as accomplices in draconian policies and challenge the merits of the investigation.^{8–11}

All health valuation studies estimate values on a scale.² To do so, each study must collect preference evidence on the scale's anchors (ie, fixing the location of 0 and 1 on the scale). A QALY scale is anchored on two scenarios: 'dying immediately' as 0 and 'starting today, one year with no health problems then die' as 1. For example, the 2013 Y3L valuation study asked respondents to choose between two dying children, with variations in the child's age, health problems and their duration.⁷ Instead of eliciting preferences on dying children, this protocol replaces the lower QALY anchor, 'dying immediately', with 'being in a coma' for a week. The original Measurement and Valuation of Health protocol included 'unconscious' and 'dead' as EQ-5D-3L profiles^{12 13}; however, the Paris protocol replaced 'dead' with 'immediate death' and dropped 'unconscious' entirely.^{14 15} Based on feedback from clinical experts, 'unconscious' alone is not an appropriate description due to its heterogeneity. Instead, 'being in a coma' is a 'lengthy deep state of unconsciousness',¹⁶ which is a more specific profile and suitable to serve as the lower anchor of the experience scale.^{10 17–20}

To be clear, there are also no death descriptions in the hypothetical scenarios of this DCE: All children survive. If the evidence is promising, it is possible that experience scaling could replace QALY scaling in child health valuation and other forms of health preference research. On the experience scale, a positive value reflects an experience better than no experience (ie, being in a coma), and a negative value reflects an experience worse than no experience. However, we recognise that experience scaling does not convey the trade-off between quality and length of life and may not be suitable for cost-utility analyses.

Although paired comparisons (like those used in the original study) are commonplace in child health valuation,²¹ they are highly inefficient and burdensome to

respondents: Each choice implies only a single inequality (eg, $A > B$), and respondents are forced to choose between two children. In addition to 10 paired comparisons, the survey instrument in this protocol includes 10 kaizen tasks (online supplemental appendix 1). In a kaizen task, respondents express their preference for improving a single object in a choice set.⁷ **Kaizen** is a Japanese term that describes continuous improvement, which in this case, is the discrete evolution of an object over a sequence of choices. **Kaizen tasks elicit preference paths** (ie, each respondent's optimal sequence of improvements from an initial profile toward an idealised destination). In this protocol, respondents choose to relieve the health problems in a 10-year-old child, which is similar to the triage process in paediatric decision-making. Furthermore, the adaptive process involved in the kaizen task produces greater preference evidence (ie, multiple inequalities per task) than a paired comparison (a single inequality).

The kaizen task was originally developed for the valuation of the EQ-5D-5L⁹ and has been adapted to collect preference evidence on COVID-19 vaccinations in the USA⁸ and on child surgical outcomes in the UK.¹¹ Based on the findings of these three prior studies, we developed this protocol (the fourth implementation of the kaizen task). If the evidence here is positive, it is possible that kaizen tasks could replace paired comparisons in child health valuations and other forms of health preference research.

For example, a paired comparison under this protocol might ask a respondent to choose between two children with 'health problems' (eg, 13131<31311) or between having 'health problems' or being 'in a coma' (33331<coma). Alternatively, each kaizen task will introduce a single child with 'health problems' (eg, 33331) and ask the respondent to relieve the child's problems (33331<33311<33111<31111). Based on qualitative feedback in previous studies, respondents enjoyed the evolving format of the kaizen task because it gives them the ability to modify the child's health as they see fit.¹¹ This interactive experience is different from paired comparisons, wherein each respondent makes a single choice and does not see the implications, namely, how choices relieve health problems.

In summary, the protocol described in this paper was designed to implement and assess two innovations in child health valuation: (1) the use of experience scaling instead of QALY scaling and (2) the use of kaizen tasks instead of paired comparisons.

Aims

The overall objective of this protocol is to assess two potential innovations in child health valuation. Specifically, our aims are the following:

1. To develop and implement a DCE for the valuation of Y3L profiles on an **experience scale**.
2. To compare the preference evidence of the paired comparison and **kaizen tasks**.

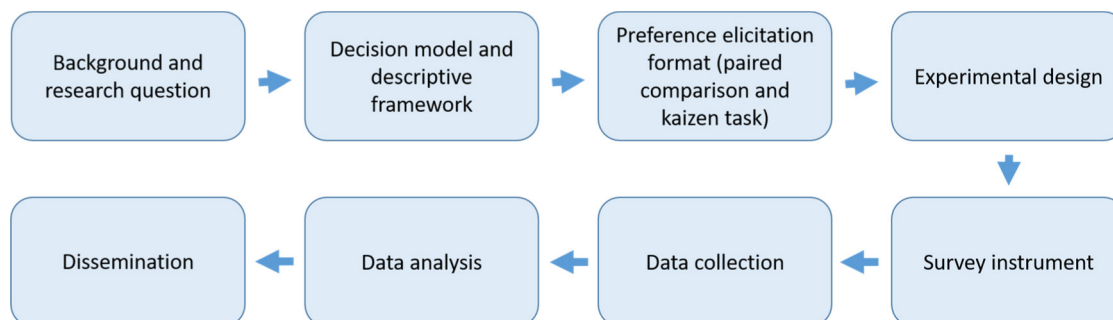


Figure 1 Phases of the study.

We hypothesise that the kaizen tasks will produce concordant responses but with more precise values (ie, smaller standard errors) than paired comparisons (allowing for smaller samples in future studies). The evidence from this study may also help us better understand the differences between tasks in terms of response behaviours and respondent feedback. Further, we hypothesise that the 10 main effects of the Y-3L profiles are statistically significant (p value <0.05) on an experience scale and that being ‘in a coma’ is significantly better than the worst Y-3L profile (33333) for a week. The findings will only reflect US Y-3L values on 1-week episodes for a generic 10-year-old child. Later on, we expect to adapt this protocol for other child ages, health problems, and durations, similar to the original study.⁷

METHODS AND ANALYSIS

Figure 1 presents the different phases of this study. In this protocol we describe the decision model and descriptive framework, preference elicitation format for the paired comparison and kaizen tasks, experimental design, survey instrument, data collection and analysis and dissemination.

Further details on the survey instrument, experimental design and analysis plan can be found in online supplemental appendices 1–3, respectively. For instance, in online supplemental appendix 1 we provide screenshots from the online survey instrument with detailed description of each page, as well as our responses to the comments we received from beta testing. In online supplemental appendix 2, we provide details about survey, block and subject design, as well as the variables that we submit to the LimeSurvey for the proper functioning of the online survey. Finally in online supplemental appendix 3, we provide a detailed list of descriptive, primary and secondary analyses.

Patient and public involvement

This is a national valuation study that will assess the health preferences of the general population using a DCE survey instrument. Members of the general population were involved in the beta-testing of the survey instrument and in the piloting of the survey instrument (see online supplemental appendix 1) and will be involved

as respondents to the survey. The primary results will be submitted as a manuscript to a peer-review journal.

Decision model and descriptive framework

In a child health valuation study, each respondent makes choices that express their preferences on HRQoL episodes for a generic child. Craig and colleagues found that the preferences of US adults on child HRQoL vary by child age (9 or 11 years old) and problem duration (1 or 2 years).⁷ Under this protocol, respondents will make choices regarding a 1-week episode for a 10-year-old child.

As in the original Y-3L valuation study, this protocol uses the Y-3L descriptive system to describe five HRQoL attributes: (1) mobility (walking around), (2) taking care of myself, (3) doing usual activities, (4) having pain or discomfort and (5) feeling worried, sad or unhappy. Each Y-3L attribute has three levels (table 1), and their definitions were taken directly from the US English version of the Y-3L instrument.²² In addition to these five Y-3L attributes, the 10-year-old child may be described as being ‘in a coma’.

To confirm understanding of the attributes, the background section of the survey instrument (see online supplemental appendix 1) includes the US version of the EQ-5D-5L instrument and a five-question **coma quiz** based on the published literature.^{17–20} The quiz questions were selected to address common misconceptions about comas in terms of (1) their causes, (2) self-care during comas, (3) duration of comas, (4) recovery time and (5) the outcomes of comas (see online supplemental appendix 1). In addition, the instrument provides a definition of coma for reference in the preference elicitation tasks.¹⁶

Prior to the coma comparisons (ie, Y-3L vs coma), the instrument clarifies that being ‘in a coma’ and having ‘health problems’ for a week has the same cause, recovery and survival. In addition, all respondents will confirm the following statement via a checklist: ‘In each scenario, the cause is the same, the recovery is the same, the rest of child’s life is the same, and regardless of the respondents’ choices, the child survives’. In other words, the preference evidence on being ‘in a coma’ for 1 week will pertain solely to that week and not to events before or after the week.

Table 1 Attributes and attribute levels for a 10-year-old child

Attributes	Attribute levels
Mobility (walking around)	No problems walking around
	Some problems walking around
	A lot of problems walking around
Looking after myself	No problems taking a bath or shower by myself or getting dressed by myself
	Some problems taking a bath or shower by myself or getting dressed by myself
	A lot of problems taking a bath or shower by myself or getting dressed by myself
Doing usual activities	No problems doing my usual activities
	Some problems doing my usual activities
	A lot of problems doing my usual activities
Having pain or discomfort	No pain or discomfort
	Some pain or discomfort
	A lot of pain or discomfort
Feeling worried, sad or unhappy	Not worried, sad or unhappy
	A bit worried, sad or unhappy
	Very worried, sad or unhappy
Being in a coma	Experiences health problems
	Is in a coma

The EuroQol Research Foundation understands that we are using the EQ-5D-Y-3L descriptive system for the purposes of these preference elicitation tasks. However, the Foundation asked us to provide the following trademark text: "© EuroQol Research Foundation. EQ-5D™ is a trade mark of the EuroQol Research Foundation. Reproduced by permission of EuroQol Research Foundation. Reproduction of is not allowed. For reproduction, use or modification of the EQ-5D (any version), please register your study by using the online EQ registration page: www.euroqol.org."

Preference elicitation

To estimate Y-3L values on an experience scale, this protocol includes a DCE with three types of preference elicitation tasks: 5 coma comparisons, 10 paired comparisons and 10 kaizen tasks. We assume the **Markov property** in that the likelihood of each choice is between 0 and 1 (McFadden's positivity assumption) and depends only on the choice set, not on any previous choices.³ Nevertheless, the survey instrument starts with the coma comparisons and randomises the order of the kaizen and paired comparison components. Furthermore, the sequence of sets is randomised within each component, and the object position in the paired comparisons is randomised as well (ie, left-right).

In the coma comparisons (figure 2A), each respondent will be presented with a single Y-3L profile and asked to choose between being 'in a coma' and having 'health problems'. We assume that being 'in a coma' relieves suffering in some scenarios.

In the paired comparisons (figure 2B), each respondent will be presented with a pair of Y-3L profiles (no

coma) and asked to choose between these profiles. This task is a commonplace in child health valuation.

In the kaizen tasks (figure 2C), each respondent will be presented with a single Y-3L profile and asked to relieve three health problems (ie, first, second, third improvement). The ranking of the four improvements creates a preference path from an origin profile (without the four improvements) to a destination profile (with the four improvements).

Comparing these three tasks, the coma and paired comparisons have a single response, and each will resolve one inequality (eg, 31311 vs 13131 and 33331 vs coma, respectively). In the kaizen task, there are four potential improvements between the origin and destination profiles (ie, origin–destination (OD) pairs; eg, 33331 to 11111); therefore, its three responses endorse one out of 24 preference paths and will resolve 11 inequalities. The first response indicates the preferred alternative out of four Y-3L profiles (eg, 13331, 31331, 33131, 33311), the second response indicates the preferred alternative out of six profiles (11331, 13131, 13311, 31131, 31311, 33111) and the third response indicates the preferred alternative out of four profiles (11131, 11311, 13111, 31111).

For the coma comparison and the kaizen task, a respondent will read and interpret a single Y-3L profile (ie, five adjectival statements) as well as being 'in a coma' or four improvements, respectively. For a paired comparison, a respondent will read and interpret two Y-3L profiles (ie, 10 adjectival statements), which is more cognitively burdensome. Unlike a coma or paired comparison, a kaizen task draws attention to the potential improvements rather than the profiles themselves, and a respondent must respond three times, not once.

Experimental design

To create the experimental design, we constructed four D-efficient blocks of 10 OD pairs and then generated the corresponding five Y-3L profiles for the coma comparisons and 10 Y-3L pairs for paired comparisons for each block. In all tasks, each pair includes one hold-out (ie, an attribute that the act of choosing will not change its level) to deter lexicographic preferences.²³ The full experimental design is presented in online supplemental appendix 2.

Selecting the OD pairs for the kaizen tasks by D-efficiency

As a temporary assumption, the hold-out in each OD pair was initially set to level 1 (no problems); therefore, the Y-3L descriptive system has only 405 candidate OD pairs (ie, five possible holdouts \times (three possible improvements per attribute (level 1–2, 2–3 and 1–3))^{four differential attributes} = 5×3^4). From this full factorial design, we eliminated those OD pairs with a predicted probability of less than 0.15 based on De Moivre's NP5 rule and McFadden's positivity assumption,³ reducing the number of candidate OD pairs to 321.²⁴

From these 321 candidate OD pairs, we randomly selected 10 OD pairs (a block) that satisfied three

Please select an alternative that you prefer.

Task 1 of 5:	A 10-year-old child has the following problems for 1 week	A 10-year-old child is in a coma for 1 week
1 Mobility (walking around)	A lot of problems walking around	
1 Taking care of myself	A lot of problems taking a bath or showering by myself or getting dressed by myself	
1 Doing usual activities	A lot of problems doing my usual activities	
1 Having pain or discomfort	Some pain or discomfort	
1 Feeling worried, sad or unhappy	Not worried, sad or unhappy	
Which do you prefer?	A 10-year-old child has the following problems for 1 week	A 10-year-old child is in a coma for 1 week

Next

A Example of a Coma Comparison

Please select an alternative that you prefer.

Task 1 of 10:	A 10-year-old child has the following problems for 1 week	A 10-year-old child has the following problems for 1 week
1 Mobility (walking around)	Some problems walking around	Some problems walking around
1 Taking care of myself	No problems taking a bath or showering by myself or getting dressed by myself	Some problems taking a bath or showering by myself or getting dressed by myself
1 Doing usual activities	A lot of problems doing my usual activities	No problems doing my usual activities
1 Having pain or discomfort	Some pain or discomfort	No pain or discomfort
1 Feeling worried, sad or unhappy	A bit worried, sad or unhappy	Very worried, sad or unhappy
Which do you prefer?	A 10-year-old child has the following problems for 1 week	A 10-year-old child has the following problems for 1 week

Next

B Example of a Paired Comparison

Please select three improvements.

Which problem do you prefer to relieve **first? Second? Third?**
Your choices directly relieve the health problems of the child.

Warm-up task	A 10-year-old child has the following problems for 1 week	Potential improvements
1 Mobility (walking around)	A lot of problems walking around	No problems walking around
1 Taking care of myself	No problems taking a bath or showering by myself or getting dressed by myself	3
1 Doing usual activities	No problems doing my usual activities	2
1 Having pain or discomfort	No pain or discomfort	1
1 Feeling worried, sad or unhappy	Not worried, sad or unhappy	
Which problem do you prefer to relieve third?	A 10-year-old child has the following problems for 1 week	Potential improvements

Clear

Next

C Example of a Kaizen task

Figure 2 (A) Example of a coma comparison. (B) Example of a paired comparison. (C) Example of a kaizen task.

conditions. First, each origin profile within a block can appear only once (ie, no duplicate origins). Second, each attribute is a hold-out exactly twice among the 10 OD pairs (ie, hold-out balance). Third, each of the incremental differences between attribute levels (ie, level 1–2 and level 2–3) appears at least three times within a block. We repeated this process to sample 500 blocks and selected the block with the highest D-efficiency. We then repeated this 500-block process four times to produce four D-efficient blocks of 10 OD pairs.

Lastly, we reassigned the levels of the hold-out attributes for each origin profile under three conditions. First, if an origin profile has four level-3 attributes, then the hold-out is always at level 1 (eg, 33331). Second, if an origin profile has four level-2 attributes, then the hold-out is always at level 3 (eg, 22223). These utility balance conditions reduced the range of origin values (eg, excluding 33333 and 22221). For the rest of the pairs, we assigned hold-out levels randomly such that all 40 origin profiles are unique (ie, unique OD pairs).

Generating the Y-3L profiles for the coma comparison and Y-3L pairs for the paired comparisons

For each of the four D-efficient blocks of 10 OD pairs, we generated five Y-3L profiles for the coma comparisons (Y-3L vs coma) and 10 Y-3L pairs for paired comparisons (Y-3L vs Y-3L).

For the coma comparisons, we identified the worst origin profiles from each block using the preference evidence from the original study. Next, we selected a set of five profiles that satisfied three conditions. First, the set of five Y-3L profiles must show level balance (ie, a level must not repeat more than three times for each attribute). Second, each set must be non-dominant (ie, include a profile better than coma and a profile worse than coma). Third, the probability of choosing to be ‘in a coma’ must always be within the range (0.05–0.95) (ie, De Moivre’s NP5 rule).

For the paired comparisons, we ranked the four improvements in each OD pair using the preference evidence from the original study. Next, we generated 10 Y-3L pairs such that one alternative has the first and fourth improvements and the other has the second and third improvements. This design implies that the evidence from a single kaizen task may predict the probability of each paired comparison.

Survey instrument

The survey instrument consists of six components (for screenshots, see online supplemental appendix 1). The consent and screener component (six questions) includes a consent form followed by demographic questions that are necessary for implementing the inclusion criteria and ensuring that quotas are met. The background component (seven questions) introduces the five domains in the Y-3L profiles and confirms respondent knowledge about comas using a quiz. For this purpose, respondents

completed the US English EQ-5D-5L and EQ-VAS questions and five true/false questions about comas.

The third, fourth and fifth components are different types of preference elicitation tasks. Specifically, the second component includes the coma comparisons (ie, coma pairs), which consist of a warm-up task (33333 vs coma), a confirmation checklist and five tasks. The third component includes paired comparisons consisting of 10 tasks. The fourth component is a kaizen task consisting of a warm-up task (33331 vs 11111) and 10 tasks.

The follow-up component consists of two debriefing questions, seven questions about socioeconomic status, work and family, and four questions on political affiliations. Debriefing questions ask the respondent about their preferences on two types of preference elicitation tasks (ie, paired comparison vs kaizen tasks). Specifically, the questions ask respondents which preference elicitation task they (1) found easier to complete, (2) preferred to complete and (3) found easier to understand. After the debriefing questions, we ask about parental status, residency, marital status, educational attainment (after age 25) and 2022 household income. The final survey question is an open-text response that allows respondents to provide feedback on the overall survey experience. All respondents should be able to complete the entire survey instrument in 20 min (about 20 s per question).

Data collection

Under this protocol, 600 adult members of the US general population will be recruited through a panel vendor to complete the survey instrument. The recruitment strategy will be similar to ones employed in prior US health preference studies.^{25–30} Specifically, potential participants will be invited to complete the survey. In the invitation, we will provide information about the survey, estimated completion time, compensation and a link to the survey instrument. Data monitoring will involve the regular download of spreadsheets from the survey platform to fulfil recruitment quotas. No further oversight will be required.

To be able to participate in a survey, individuals need a reliable internet connection, access to a computer (or other device), the ability to read English text and the capability to respond using their devices. There are no physical risks posed by this study, but there may be a risk of psychological distress resulting from questions that ask respondents to evaluate alternative scenarios. The screener component of the survey instrument will confirm that each respondent understands the risks and agrees to participate, resides in the 50 US states or Washington, DC, and belongs to one of the 18 demographic quotas.

To promote concordance with the 2021 American Community Survey,³¹ the 18 demographic quotas account for gender (female and other), age (18–34, 35–54, 55 and older) and race/ethnicity (Hispanic, non-Hispanic black and non-Hispanic other). The threshold of adulthood in 47 states and Washington, DC, is 18 years of age;

however, in Nebraska and Alabama, the threshold is 19, and in Mississippi, the threshold is 20 years of age. Instead of excluding interested participants who belong to filled quotas (ie, quotas at the maximum size), the recruitment strategy will target potential participants who belong to unfilled ones until sample size is achieved. Some respondents may be excluded after entry if they drop out prior to survey completion or exhaust their three attempts at the coma quiz.

The survey instrument (see online supplemental appendix 1) was programmed into LimeSurvey (Community Edition) V.5.6. Apart from the choice tasks, the survey instrument uses questions from Craig and colleagues,^{8,9} Rivero-Arias colleagues,¹¹ as well as a few novel questions (ie, quiz, perspective, confirmation, debrief). The preference elicitation tasks used question themes owned by Benjamin M Craig, which were originally programmed for the US vaccination study.⁸

Data analysis

The analysis plan includes descriptive analysis, primary analysis and sensitivity and secondary analysis. In the descriptive analysis, we will use various tests to analyse sample selection, experimental design, response behaviours and response frequencies. In the primary analysis, we will estimate the main-effects conditional logit model using the maximum likelihood estimator and compare the results of the paired comparisons and kaizen tasks. To assess the robustness of our primary analysis, we will conduct sensitivity analysis, including maximum simulated likelihood estimation of the mixed logit model. Finally, the secondary analyses will explore findings beyond the two primary aims of the study (see online supplemental appendix 3).

Ethics and dissemination

Ethics considerations

The independent review board (IRB) at Advarra determined that this research project (Pro00072276; 27 June 2023) is exempt from IRB oversight based on the Department of Health and Human Services regulations found at 45 CFR 46.104(d)(2). Furthermore, the IRB determined that the project is not subject to requirements for continuing review.

Dissemination

As part of the dissemination plan, the results of this study will be prepared for publication in peer-reviewed journals and presented at scientific meetings. The statistical code will be disseminated along with these results to facilitate their interpretation. The data has no direct identifiers and will be made available on reasonable request.

Acknowledgements The authors would like to thank the EuroQol Research Foundation for their support of Maksat Jumamyradov's dissertation (304-PHD), under the supervision of Drs Murat K Munkin and Benjamin M Craig, University of South Florida, Tampa, USA. We thank Juan M Ramos-Goñi at Maths in Health for his consultation, assistance and support with the survey instrument, including the question themes.

Contributors BMC developed the original idea for the study and secured funding. MJu and BMC were responsible for the study design, carried out the experimental design, implemented the online survey instrument and developed the data analysis strategy. MJu and BMC drafted the manuscript and appendices. MJu, OR-A, MJa and BMC revised and approved the final manuscript, assisted with the development of the analysis plan and experimental design and testing of the survey instrument.

Funding This work was supported by the EuroQol Research Foundation grant number 304-PHD (co-principal investigators: BMC and MJu).

Competing interests MJu received funding from the EuroQol Research Foundation. Apart from his appointment at University of South Florida, BMC is a member of the EuroQol Group, co-chair of the EuroQol PhD Network and a freelancer of the EuroQol Research Foundation. MJa is a member of the EuroQol Group and receives funding from the EuroQol Research Foundation.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Benjamin Matthew Craig <http://orcid.org/0000-0003-1121-1316>

Oliver Rivero-Arias <http://orcid.org/0000-0003-2233-6544>

REFERENCES

- 1 Craig BM, Lancsar E, Mühlbacher AC, *et al*. Health preference research: an overview. *Patient* 2017;10:507–10.
- 2 Stolk EA, Craig BM, Mulhern B, *et al*. Health valuation: demonstrating the value of health and LifeSpan. *Patient* 2017;10:515–7.
- 3 McFadden D. Conditional Logit analysis of qualitative choice behavior. In: Zarembka P, ed. *Frontiers in Econometrics*. New York: Academic Press, 1974: 105–42.
- 4 Coombs CH. *A theory of data*. Oxford, England: Wiley, 1964.
- 5 Thurstone LL. The measurement of values. *Psychol Rev* 1954;61:47–58.
- 6 Kreimeier S, Greiner W. EQ-5D-Y as a health-related quality of life instrument for children and adolescents: the instrument's characteristics, development, current use, and challenges of developing its value set. *Value Health* 2019;22:31–7.
- 7 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.
- 8 Craig BM. United States COVID-19 vaccination preferences (CVP): 2020 hindsight. *Patient* 2021;14:309–18.
- 9 Craig BM, Rand K, Hartman JD. Preference paths and their Kaizen tasks for small samples. *Patient* 2022;15:187–96.
- 10 Powell PA, Rowen D, Rivero-Arias O, *et al*. Valuing child and adolescent health: a qualitative study on different perspectives and priorities taken by the adult general public. *Health Qual Life Outcomes* 2021;19:222.
- 11 Rivero-Arias O, Buckell J, Allin B, *et al*. Using stated-preferences methods to develop a summary metric to determine successful treatment of children with a surgical condition: a study protocol. *BMJ Open* 2022;12:e062833.
- 12 Williams A. *The Measurement and Valuation of Health: A Chronicle*. Discussion Paper 136. York: Center for Health Economics, University of York, 1995.

- 13 Craig BM, Busschbach JJV. Revisiting United States valuation of EQ-5D states. *J Health Econ* 2011;30:1057–63.
- 14 Kind A. *A revised protocol for the valuation of health states defined by the EQ-5D-3L classification system: learning the lessons from the MVH study*. York: Centre for Health Economics, University of York, 2009.
- 15 Brooks RG. *28 Years of the EuroQol Group: An Overview*. EuroQol Working Paper Series Number 15003. 2015.
- 16 National Institute of Health. Coma. 2023. Available: <https://www.ninds.nih.gov/health-information/disorders/coma> [Accessed 23 May 2023].
- 17 Posner JB, Saper CB, Schiff N, et al. *Plum and posner's diagnosis of stupor and coma*. OUP USA, 2007.
- 18 Hux K, Schram CD, Goeken T. Misconceptions about brain injury: a survey replication study. *Brain Inj* 2006;20:547–53.
- 19 Young GB. Coma. *Ann N Y Acad Sci* 2009;1157:32–47.
- 20 Patel S, Hirsch N. Coma. continuing education in anaesthesia. *Critical Care & Pain* 2014;14:220–3.
- 21 Ramos-Goñi JM, Oppe M, Stolk E, et al. International valuation protocol for the EQ-5D-Y-3L. *Pharmacoeconomics* 2020;38:653–63.
- 22 Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res* 2010;19:875–86.
- 23 Lipman SA, Zhang L, Shah KK, et al. Time and lexicographic preferences in the valuation of EQ-5D-Y with time trade-off methodology. *Eur J Health Econ* 2023;24:293–305.
- 24 De Moivre A. *The doctrine of chances: or, A method of calculating the probabilities of events in play*. Chelsea Publishing Company Inc, 1756.
- 25 Craig BM, Reeve BB, Cella D, et al. Demographic differences in health preferences in the United States. *Med Care* 2014;52:307–13.
- 26 Craig BM, Reeve BB, Brown PM, et al. US valuation of health outcomes measured using the PROMIS-29. *Value Health* 2014;17:846–53.
- 27 Craig BM, Brown DS, Reeve BB. The value adults place on child health and functional status. *Value Health* 2015;18:449–56.
- 28 Craig BM, Brown DS, Reeve BB. Valuation of child behavioral problems from the perspective of US adults. *Med Decis Making* 2016;36:199–209.
- 29 Craig BM, Rand K. Choice defines Qalys: a US valuation of the EQ-5D-5L. *Value in Health* 2018;21:S12.
- 30 Craig BM, Rand K, Bailey H, et al. Quality-adjusted life-years without constant proportionality. *Value Health* 2018;21:1124–31.
- 31 United States Census Bureau. American community survey; 2021.