BMJ Open

Assessing the experimental EuroQol toddler and infant populations (EQ-TIPS) descriptive system: a protocol integrating discrete choice experiment (DCE) surveys in instrument development

Benjamin Matthew Craig ,[1] Oliver Rivero-Arias ,[2] Nancy Devlin,[3] Janine Verstraete[4]

To cite: Craig BM, Rivero-Arias O, Devlin N, *et al*. Assessing the experimental EuroQol toddler and infant populations (EQ-TIPS) descriptive system: a protocol integrating discrete choice experiment (DCE) surveys in instrument development. *BMJ Open* 2025;**15**:e100897. doi:10.1136/bmjopen-2025-100897

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2025-100897).

Received 19 February 2025
Accepted 18 June 2025

Check for updates

© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to
Dr Benjamin Matthew Craig; bcraig@usf.edu

ABSTRACT

Introduction The experimental EuroQol Toddler and Infant Populations (EQ-TIPS) instrument is currently under development as a health-related quality of life (HRQoL) measure for toddlers and infants aged 0–36 months. Using this protocol, researchers can conduct surveys with discrete choice experiments (DCEs) that examine the key properties of HRQoL instruments, specifically whether severity aligns with preferences across each attribute and the extent to which attributes influence choices. To demonstrate this protocol, we will conduct two waves of DCE surveys using different choice tasks and a common scenario (a 1-month health episode for a 1-year-old child).

Methods and analysis In the first wave (a general population sample of 400 Australian adults; 14 kaizen tasks each), respondents will view a single EQ-TIPS-five-level (v3.0) profile for each task and be asked to make a series of choices that sequentially alleviate the child's health problems. Using this exploratory evidence, we will assess whether EQ-TIPS severity aligns with preferences and the extent to which each attribute level influences choices (ie, main effects). In the second wave (1000 Australian adults; 28 paired comparisons), respondents will see two EQ-TIPS profiles for each task and be asked to choose between them. Using this confirmatory evidence, we will compare the main effects and their uncertainty by wave. For each DCE, we will estimate a main-effects conditional logit model and test for differences in effects using cluster bootstrap techniques. As sensitivity analyses, we will evaluate the effects of task sequence, attribute order and sample size on uncertainty in each wave.

Ethics and dissemination The independent review board at Includovate evaluated the application for ethical clearance and approved the study on 19 February 2025. To disseminate our findings, we will prepare multiple manuscripts for publication in peer-reviewed journals and present highlights at scientific meetings, such as the EuroQol Plenary Meeting and the International Academy of Health Preference Research.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ To compare preference evidence by task, wave-1 respondents will complete kaizen tasks by choosing between improvements (ie, 'goods'), while wave-2 respondents will complete paired comparisons by choosing between profiles of health problems (ie, 'bads'), which is more burdensome, especially when the choices affect the health of a 1 year-old child.

⇒ The statistical efficiency of kaizen tasks in gathering preference evidence, along with their reduced burden on respondents, better supports the development of other health-related quality of life instruments than paired comparisons.

⇒ The protocol relies on online surveys using a marketing panel, which may have quality issues and not be generalisable to the Australian general population, particularly those who are less educated or do not participate in such surveys or panels.

⇒ This paper provides the survey instrument, experimental design and analysis plan for the first wave only, because we will adapt wave-2 survey materials based on the wave-1 evidence.

INTRODUCTION

A wealth of instruments have been used to measure *health-related quality of life (HRQoL)* in adult and youth populations and health systems.[1] To inform the evaluation of health technologies, HRQoL instruments may be transformed into *descriptive systems* (ie, changing items into attributes) and incorporated into protocols for *discrete choice experiment (DCE)* surveys or other forms of health valuation (eg, interviews with time trade-off tasks).[2–5] Analyses of preference evidence can produce weights that facilitate the summary of HRQoL evidence, typically from a general population perspective.[6 7]

BMJ Open: first published as 10.1136/bmjopen-2025-100897 on 8 July 2025. Downloaded from http://bmjopen.bmj.com/ on July 14, 2025 at Bodleian Libraries of the University of Oxford . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Given a decision context (ie, setting, role and scenario underlying the decision) and descriptive framework (visual and verbal representation of objects, including the descriptive system of attributes), DCE surveys in health valuation commonly assume two key properties: each successive level is more detrimental along the dimension and less desirable (ie, arranged in descending rank) and that each incremental difference between attribute levels is decision-relevant (ie, main effect is non-zero). Yet, a paucity of available evidence compares alternative approaches to testing these properties. The objective of this protocol is to assess these two key properties using alternative choice tasks to inform the development of an HRQoL instrument for toddler and infant populations.

Experimental EuroQol Toddler and Infant Populations V.3.0 instrument

The EuroQol Toddler and Infant Populations (EQ-TIPS) instrument was developed as a generic measure of HRQoL for children aged 0–3 years for clinical research, population health studies and health technology assessment/economic analysis.[6–10] Thus, the instrument was developed with a limited number of items to be amenable to generating a concise descriptive system for future valuation studies.

The initial development and testing of the EQ-TIPS took place in South Africa and resulted in an instrument consisting of six items: movement, eating, pain, social interaction, play and communication.[11 12] All these items, except pain, were norm-referenced with reference to 'age-appropriate behaviour.' Although the items were developed ab initio, the layout, font, instructions, response options, Visual-Analogue Scale and recall period of other instruments in the EuroQol family of instruments were retained for consistency.

In line with the experimental nature of the EQ-TIPS, work is ongoing to refine the descriptive system and ensure the multinational validity of the instrument. As of January 2025, the instrument has been refined following engagement with multiple stakeholders and tested qualitatively in eight diverse countries (Argentina, Australia, Canada, Egypt, Germany, Singapore, South Africa and the UK) for content validity.[8] This work has resulted in a seven-item instrument with both a three-level (EQ-TIPS-3L) and a five-level (EQ-TIPS-5L) version (table 1). The five-level version offers greater opportunities for level inversion (successive levels along the dimension not in descending ranking) and non-zero main effects, which are part of this assessment.

Table 1 Experimental EQ-TIPS-5L V.3.0 descriptive system

Level 1	Level 2	Level 3	Level 4	Level 5
Movement (eg, brings hands to mouth, head control, reaching, sitting, crawling, walking, running)				
No problems moving	A little bit of a problem moving	Some problems moving	A lot of problems moving	Extreme problems moving
Eating or drinking (eg, sucking, swallowing, keeping food down)				
No problems eating or drinking	A little bit of a problem eating or drinking	Some problems eating or drinking	A lot of problems eating or drinking	Extreme problems eating or drinking
Sleep (eg, falling asleep, staying asleep, amount of sleep)				
No problems sleeping	A little bit of a problem sleeping	Some problems sleeping	A lot of problems sleeping	Extreme problems sleeping
Pain (a child in pain may cry non-stop, be restless or unusually still, make a face, moan or tell you)				
No pain	A little bit of pain	Some pain	A lot of pain	Extreme pain
Managing emotions (eg, the child can be comforted when upset)				
No problems managing emotions	A little bit of a problem managing emotions	Some problems managing emotions	A lot of problems managing emotions	Extreme problems managing emotions
Interacting with others (eg, eye contact, smiling, laughing, gesturing, talking)				
No problems interacting with others	A little bit of a problem interacting with others	Some problems interacting with others	A lot of problems interacting with others	Extreme problems interacting with others
Play (eg, explores objects or toys, plays games, pretend play)				
No problems playing	A little bit of a problem playing	Some problems playing	A lot of problems playing	Extreme problems playing

EQ-TIPS is a trademark of the EuroQol Research Foundation. Generic English V.3.0. Reproduction is not allowed. For the reproduction, use or translation of the EQ-TIPS (any version), please contact EuroQol's R&D Manager: rd@euroqol.org. EQ-TIPS has an experimental version status. It is not a final product: subject to change.
EQ-TIPS, EuroQol Toddler and Infant Populations; 5L, five level.

BMJ Open: first published as 10.1136/bmjopen-2025-100897 on 8 July 2025. Downloaded from http://bmjopen.bmj.com/ on July 14, 2025 at Bodleian Libraries of the University of Oxford . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Concordance between severity and preference

Psychometric evidence may demonstrate that the response options of an item concord with their severity as viewed by respondents. Likewise, a DCE survey may show that adjectival statements of an attribute concord with respondent preferences. Together, these two patterns imply concordance between severity and preference in descending rank along each dimension of a descriptive system.

Although unanimity in preferences may not be achievable, discordances between attribute levels (eg, the proportion who prefers level 5 to level 4) should be rare (eg, less than 20%).[13] For example, the EQ-five dimensions (5D)-5L descriptive system has five dimensions of HRQoL with five levels. Past paired comparison studies identified discordances in the EQ-5D-5L that led to changes in its translations[14]; however, these assessments were conducted using partial profiles (ie, only one attribute at a time). Currently, there is no consensus on what degree of discordance is acceptable. On this topic, we developed this protocol and will assess whether severity concords with preferences along each dimension of the EQ-TIPS-5L descriptive system (aim 1).

The influence of attribute levels on choices

HRQoL instruments tend to focus on measuring what matters, namely, outcomes that are decision relevant (ie, *non-zero main effects*). Sometimes an incremental difference in response levels concords with severity, but the corresponding difference in attribute levels does not influence choices (ie, its main effect is zero). When redundant response options/attribute levels are identified post launch, it is often too late to remove them from the HRQoL instrument/descriptive system.[6] This protocol was designed to identify redundant response options/attribute levels in the HRQoL instrument/descriptive system during its development. For this purpose, we examine the extent to which the EQ-TIPS-5L attribute levels influence choices (aim 2).

DCE with kaizen and pick-one tasks

A DCE is an experiment involving choice tasks, such as kaizen tasks (figure 1) and paired comparisons (figure 2). DCEs typically include pick-one tasks, like *paired comparisons*, where a respondent selects one out of two alternatives. *Kaizen* is a Japanese term describing continuous improvement (ie, kai~change; zen~good), and *kaizen tasks* elicit the discrete evolution following such an adaptive process of improvement (ie, preference paths). In a kaizen task, a respondent is assigned a single profile (origin) and makes a series of choices that sequentially improve the profile. Due to their adaptive approach, kaizen tasks elicit more preference evidence than pick-one tasks.

In the context of health valuation, a paired comparison involves choosing between profiles with health problems (ie, 'bads'), while a kaizen task entails selecting between improvements (ie, 'goods'); therefore, kaizen tasks are less burdensome on respondents than pick-one tasks, especially when the profiles represent toddlers or infants (ie, EQ-TIPS-5L). Furthermore, the process of sequentially alleviating health problems is more akin to clinical practice.

Please select four improvements.

Warm-up task	Starting today, a one-year-old child has health problems for 1 month then recovers	Potential changes
Sleep	Some problems sleeping	No problems sleeping
Managing Emotions	No problems managing emotions	
Movement	Some problems moving	No problems moving
Interacting with Others	No problems interacting with others	
Pain	Some pain	No pain
Play	No problems playing	Some problems playing
Eating or Drinking	Some problems eating or drinking	No problems eating or drinking
Which problem do you prefer to relieve first?	**Starting today, a one-year-old child has health problems for 1 month then recovers**	**Potential changes**

Figure 1 Warm-up kaizen task.

BMJ Open: first published as 10.1136/bmjopen-2025-100897 on 8 July 2025. Downloaded from http://bmjopen.bmj.com/ on July 14, 2025 at Bodleian Libraries of the University of Oxford . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Please select the alternative that you prefer.

Warm-up task	Starting today, a one-year-old child has health problems for 1 month then recovers	Starting today, a one-year-old child has health problems for 1 month then recovers
Sleep	Some problems sleeping	No problems sleeping
Managing Emotions	No problems managing emotions	No problems managing emotions
Movement	Some problems moving	No problems moving
Interacting with Others	No problems interacting with others	No problems interacting with others
Pain	No pain	Some pain
Play	Some problems playing	No problems playing
Eating or Drinking	Some problems eating or drinking	No problems eating or drinking
Which do you prefer?	**Starting today, a one-year-old child has health problems for 1 month then recovers**	**Starting today, a one-year-old child has health problems for 1 month then recovers**

Figure 2 Warm-up paired comparison task.

Since the original kaizen task study in 2021, three large studies have adapted the kaizen task for online surveys alongside other choice tasks.[15] Each study has a published protocol,[2 3 16] confirmed the performance of kaizen tasks alongside other choice tasks and led to key methodological advancements, which have been incorporated into this protocol, namely, the inclusion of multilevel improvements, randomisation of attribute order and lower numbers of respondents per block due to greater preference evidence per respondent. This protocol was designed to compare the performance of kaizen tasks and paired comparisons to test two key properties of the EQ-TIPS-5L descriptive system (aim 3).

Aims

The objective of this protocol is to assess two key properties of the EQ-TIPS-5L descriptive system using alternative choice tasks. Specifically, our aims are:

1. To assess whether severity concords with preferences along each EQ-TIPS-5L attribute.
2. To investigate to what extent the EQ-TIPS-5L attributes influence choices.
3. To compare the performance of kaizen tasks and paired comparisons for the EQ-TIPS-5L descriptive system.

Overall, we hypothesise that the frequency of each discordance will be nominal and will apply multiple thresholds (less than 10%, 20% and 30%; aim 1), that each main effect is non-zero (p value<0.05; aim 2), and that the main effects of the kaizen and paired comparison largely agree (aim 3).

To achieve this objective, we will conduct two waves of DCE surveys in the Australian general population. In parallel to these surveys, complementary studies will be testing the psychometric performance of EQ-TIPS-5L in multiple countries, including Australia, where the descriptive system has been tested in qualitative work. The combined evidence on preferences and psychometric performance in Australia may inform the development and refinement of the EQ-TIPS-5L instrument as well as other instruments on child HRQoL. For example, the level descriptors used in EQ-TIPS-5L are the same as those in some existing five-level instruments (eg, EQ-5D-Y-5L), and findings from these DCE surveys may be relevant for those instruments. If successful, this two-wave protocol may be incorporated into the development and refinement of future HRQoL instruments.

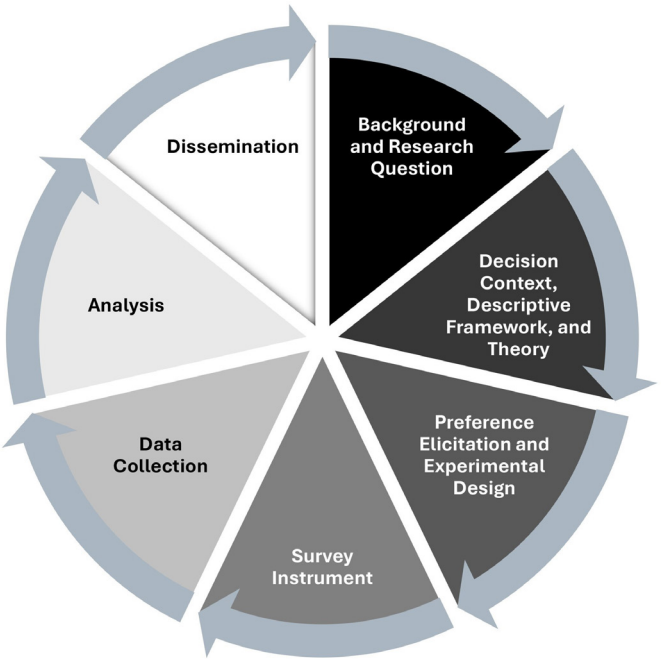

Figure 3 Circular phases of a health preference study.

BMJ Open: first published as 10.1136/bmjopen-2025-100897 on 8 July 2025. Downloaded from http://bmjopen.bmj.com/ on July 14, 2025 at Bodleian Libraries of the University of Oxford . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

METHODS AND ANALYSIS

Figure 3 presents the different phases of this study. In this protocol, we describe the background and research questions; decision context, descriptive framework and theory; preference elicitation and experimental design; survey instrument; data collection; and analysis and dissemination.

Overall, we will conduct two waves of DCE surveys that use a common decision context (eg, a 1-month episode for a 1-year-old child) but different choice tasks. For the 12-month study, we plan to start the formative qualitative research in January 2025, initiate participant recruitment directly after ethics approval of the protocol (19 February), and share our findings by the end of 2025 to align with ongoing studies on the psychometric performance of EQ-TIPS-5L.

Further details on the wave-1 experimental design, survey instrument and analysis plan can be found in online supplemental appendices 1–3, respectively. In online supplemental appendix 1 we provide screenshots from the wave-1 online survey instrument with references and a detailed description of each page. In online supplemental appendix 2, we provide the accompanying experimental design including set selection, blocking and subject design, as well as the settings and terminology implemented in the survey software (LimeSurvey). Finally, in online supplemental appendix 3, we provide an analysis plan with a detailed list of tests that we perform for descriptive analysis, primary and secondary analyses. These materials are not provided for wave 2, because we will adapt its materials based on the wave-1 evidence.

Decision context, descriptive framework and theory

Under this protocol, the decision context defines the respondent's role as the decision maker, the setting as general community and the hypothetical scenario as a 1-month episode for a 1-year-old child. Each acute episode starts today, and the child recovers after 1 month. The episodes will be described using definitions, labels and adjectival statements taken from the EQ-TIPS-5L descriptive system (table 1). The hypothetical scenario for both studies is 1 month for a 1-year-old child.

Under this decision context and descriptive framework, each respondent will complete a series of choice tasks (kaizen task or paired comparisons). For each task, the respondent selects the alternative that maximises their decisional utility (ie, *random utility model*). Using this preference evidence, we will estimate the value of each incremental difference in attribute levels (ie, main effects) in the EQ-TIPS-5L descriptive system.

This protocol incorporates good practices from prior studies to describe the decision context and descriptive framework fully to respondents in a manner that they understand so that they can make informed decisions.[17] The survey instruments (online supplemental appendix 1) will include a child representation task, confirmatory checklists, warm-up tasks and user interface elements (ie, information buttons) during the choice tasks to ensure that all respondents have some understanding of key concepts. To further mitigate potential biases, the order of EQ-TIPS-5L attributes will be randomised at the respondent level and task sequence will be randomised within each component.

Preference elicitation and experimental design

We will conduct DCEs with kaizen tasks (wave 1) and paired comparisons (wave 2) within online surveys of the Australian general population.

The wave-1 DCE survey will include 14 kaizen tasks, each with five possible changes. At the start of each kaizen task (figure 1), a respondent is shown an initial profile a_{it0} *(origin profile)* and begins by choosing their preferred change out of five possible changes. The first choice set A_{it1} is composed of all five objects formed by the initial profile with one change. After this initial choice, the profile is updated ($(a_{it0} \rightarrow a_{it1})$), and the respondent chooses a second change from the remaining four possible changes. This process continues until four of the five potential changes are incorporated into the origin profile, revealing a preference path.

In each kaizen task, four of the five possible changes will be improvements (ie, good changes; for example, level 5 to level 4) and the last change will be a bad change (eg, level 4 to level 5). *Kaiaku* is a Japanese term describing the degradation of an object (ie, kai~change; aku~bad). Including a bad change (ie, aku) alongside the good changes allows the identification of two types of discordances: (*type-1 discordance*) when the bad change is selected first, second, third or fourth; and (*type-2 discordance*) when an improvement is not selected (ie, ranked fifth). Each behaviour implies a discordance between severity and preference along a dimension, and the frequency of these behaviours will inform the assessment of the descriptive system.

Apart from these discordances, respondents in the kaizen task rank the four good changes (ie, improvements), inherently making three choices: a choice between four profiles (ie, one improvement out of four), a choice between six profiles (ie, two improvements out of four) and a choice between four profiles (ie, three improvements out of four). Each kaizen task will produce 14 choice probabilities, which is seven times the preference evidence of a paired comparison of two profiles (ie, profile A vs profile B).

For the wave-1 experimental design, we constructed two d-efficient blocks of 14 origin-destination (OD) pairs. The process to select the wave-1 blocks has three steps, each programmed in R Statistical Software (V.4.1.3): define the candidate set, block selection by d-efficiency, block evaluation and hold-out/aku assignment (see online supplemental appendix 2 for the experimental design).

In step 1, we define the candidate OD pairs. To deter lexicographic behaviour, each OD pair will initially include at least 3 holdouts (ie, an attribute that the act of choosing will not change its level). In wave 1, the descriptive system has seven attributes, which implies four

BMJ Open: first published as 10.1136/bmjopen-2025-100897 on 8 July 2025. Downloaded from http://bmjopen.bmj.com/ on July 14, 2025 at Bodleian Libraries of the University of Oxford . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

differential attributes and three hold-outs per OD pair. The full factorial set has 350 000 OD pairs (ie, 35 possible holdout combinations x (10 possible improvements) four differential attributes = 35×10^4). We further reduced the full factorial set to a candidate set of OD pairs by eliminating those with atypical combinations of severe attributes or that contain alternatives with low choice probabilities (<0.0833; that is, one in 12).

In step 2, we began the iterative process of randomly selecting 14 OD pairs along three properties: (1) Each of the 28 incremental improvements appears at least once; each improvement between level 2 and 1 and between levels 5 and 4 appears twice; and each improvement between levels 3 and 1 and between levels 5 and 3 appears once. In summary, each single-block design has the same frequency of improvements. (2) Each origin profile appears only once in a design and each combination of two improvements appears only once in a design (ie, no duplicates). This reduces repetition between tasks. (3) Each attribute is a hold-out six times (ie, hold-out balance). Next, we compute the determinant of the block's information matrix (ie, d-error). If a block's d-error is lower than its predecessors, it is considered more 'efficient'. We repeated this iterative process until d-efficiency failed to improve. Overall, we selected two d-efficient single-block designs (ie, two blocks of 14 OD pairs) for wave 1 (see online supplemental appendix 2).

In step 3, we assign the two hold-out attributes (level 2 if level 5 is present; otherwise, level 1) and one of the 28 bad changes (ie, aku) to each OD pair. This assignment has no effect on the main-effect estimates. With 14 kaizen tasks per respondent and 28 main effects, each respondent will evaluate half of the bad changes (ie, aku). The R-code for this experimental design is provided on request.

In the wave-2 paired comparisons, respondents will be presented with two health profiles and asked to choose the one they prefer (figure 2). Each choice reflects a preference for a child experiencing the selected profile of health problems over its counterpart. In other words, both profiles indicate that a 1-year-old child will face some problems for 1 month; therefore, this task requires respondents to select the health problems a child experiences (ie, 'bads'), rather than alleviating their problems by selecting improvements ('goods'). To construct the wave-2 experimental design, we will follow a similar d-efficient process based on the wave-1 evidence (ie, five blocks of 28 paired comparisons).

Survey instrument

The wave-1 survey instrument consists of four components (for screenshots, see online supplemental appendix 1). The first component (six questions) includes a consent form followed by five demographic questions that are necessary for implementing the inclusion criteria and ensuring that the 32 recruitment quotas are met.

The second component (28 questions) asks about the respondents' HRQoL, introduces the descriptive system and confirms respondent knowledge of child HRQoL. First, respondents will complete the EQ-5D-5L (six questions) as well as 13 additional questions about their HRQoL (ie, EQ-5D bolt-ons). Next, the respondents will be asked to imagine a 1-year-old child whom they know well (over 1 month) and report their relationship to the child (eg, parent). Next, the respondent is asked to 'write a unique description of this child playing with an object or toy' (up to 400 characters) with the request that it contain no identifying information. This task involves forming or articulating a mental model or representation of a child (real referent). Next, the respondent will complete the EQ-TIPS-5L for this referent child. In addition to securing a basic understanding of key concepts, this child representation task (nine questions) facilitates the exclusion of fraudulent and inattentive respondents prior to the DCE component.

The third component is the DCEs (16 or 30 questions), a confirmation checklist, a warm-up task and 14 kaizen tasks (wave 1) or 28 paired comparisons (wave 2). The checklist verifies understanding of the decision context, and the warm-up task confirms understanding of the task instructions and functionality prior to preference elicitation.

The fourth component (10 questions) consists of one debriefing question followed by questions about socioeconomic status, work and family; and an optional feedback question. Debriefing questions ask the respondent about the experience with the choice tasks. After the debriefing questions, we ask about parental status, residency, marital status, educational attainment, employment status and experience with caregiving and looking after a child.[18] The final survey question is an open text response that allows respondents to provide feedback on the overall survey experience.

Like prior studies with other instruments, members of the general population were involved in the beta-testing of the wave-1 survey instruments, similar to user-testing consultants. Once the survey instrument has passed their evaluations, members of the general population were recruited as respondents for the pilot to confirm its technical functionality. All respondents should be able to complete the entire survey instrument in less than 20 min based on experience from previous studies and pilot evidence (see online supplemental appendix 1).

Data collection

We will recruit 1400 Australian adults from a panel vendor to complete a survey instrument using a strategy similar to prior national health preference studies.[19 20] In wave 1, the DCE survey has two blocks with 200 respondents each (total of 400 respondents). In wave 2, the DCE survey has five blocks with 200 respondents each (total of 1000 respondents). Potential participants will receive an invitation with information about the survey, estimated time commitment, compensation and a survey link. Data will be monitored by regularly downloading spreadsheets from the platform to ensure recruitment goals are met. No further oversight will be required.

BMJ Open: first published as 10.1136/bmjopen-2025-100897 on 8 July 2025. Downloaded from http://bmjopen.bmj.com/ on July 14, 2025 at Bodleian Libraries of the University of Oxford . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Participation in this survey will require access to a reliable internet connection, a computer or other suitable device, the ability to read English text and the capacity to respond using their chosen device. To align with the Australia Census demographics, the recruitment strategy employs 32 quotas stratified by gender (female and other), age (18–29, 30–44, 45–64, 55+) and ancestry (Aboriginal or Torres Strait Islander, African or Middle Eastern, Asian, other). To ensure inclusivity, our recruitment will prioritise participants from unfilled quotas until the target sample size is reached. We acknowledge that some participants may be excluded after enrolment if they withdraw before completing the survey, if their description of the referent child is invalid, complete the survey instrument within 10 min (ie, speeders) or attempt to retake the survey after being excluded. To be specific, the study team will evaluate whether the open-text description includes (1) a child; (2) a toy or object; (3) any form of personalisation; and (4) a minimum of two complete sentences.

Analysis

The analysis plan (online supplemental appendix 3) describes the descriptive, primary and secondary analyses typical of any stand-alone health preference study. In the descriptive analysis, we will use various tests to analyse external and internal validity of the DCE survey evidence.

In wave 1, the primary analysis begins by calculating the frequency of discordances by attribute, main effect and type (aim 1). Using the preference evidence from waves 1 and 2, we will estimate the main-effects conditional logit model by maximum likelihood (aim 2).[21] In wave 2, we will also compare the wave-1 and wave-2 results in terms of main effects (aim 3). More specifically, we will illustrate the differences in each main effect, test their significance (p value<0.05), and assess their overall agreement using Lin's concordance coefficient.[22] Furthermore, we will compare the wave-1 predictions with the actual choices in wave 2, which is complementary to the comparison of main effects (aim 3).

As a secondary analysis, we will assess whether excluding the discordant respondents (ie, respondents with a high frequency of type-1 and type-2 discordance) affects the wave-1 main effect estimates (aim 2) or the agreement between main effect estimates by wave (aim 3). As sensitivity analyses, we will estimate the effects of block, task sequence and attribute order on the frequency of discordance in wave-1 and main effect estimates in both waves.

Table 2 summarises the hypotheses of the primary, secondary and sensitivity analyses. Due to the panel format of the preference evidence, all p values, 95% CIs and SEs will be estimated using cluster bootstrap techniques with replacement, block-specific strata and 1000 iterations.

Furthermore, future analyses could explore econometric methods, experimental and task design and preference heterogeneity, though these areas fall beyond the scope of this protocol. Recent studies indicate that estimates from random-parameter models (eg, mixed logit) can deviate significantly from true parameters, even when correctly specified.[23 24] However, we are particularly interested in estimating scale-adjusted latent-class models to distinguish between differences in behavioural factors and preference structures across classes.

Table 2 Hypotheses in the primary, secondary and sensitivity analyses

Wave-1 primary analyses	H1.1: bad changes are rarely selected (type-1 discordance; aim 1)
	H1.2: improvements are rarely skipped (type-2 discordance; aim 1)
	H2.1: the main effects are non-zero (kaizen evidence; aim 2)
Wave-2 primary analyses	H2.2: the main effects are non-zero (paired comparison evidence; aim 2)
	H3.1: the main effects agree between waves (aim 3)
	H3.2: wave-1 predictions and wave-2 choices agree (aim 3)
Secondary analyses	H4.1: excluding discordant respondents does not alter the main effects (aim 2)
	H4.2: excluding discordant respondents does not alter the agreement between waves (aim 3)
Sensitivity analyses	H5.1: the frequency of discordance and main effects agree by block
	H5.2: the frequency of discordance and main effects agree by task sequence
	H5.3: the frequency of discordance and main effects agree by attribute order

H, hypothesis.

ETHICS AND DISSEMINATION PLAN

This research project involves two waves of anonymous surveys of adults in the Australian general population. The independent review board at Includovate reviewed the application for ethical clearance and approved the study on 19 February 2025. The study poses no physical risks. However, some questions ask respondents to consider hypothetical health scenarios in very young children, which could potentially cause psychological distress in a small number of participants.

All survey data will be anonymised on receipt, removing identifiers that could link information to individuals or establishments. Furthermore, we will ensure the secure storage of these data on encrypted servers with restricted access, requiring multifactor authentication for authorised personnel. Regular security audits and compliance checks will be conducted to ensure data integrity and protection against breaches. Backup systems will also comply with security standards to safeguard against data loss.

BMJ Open: first published as 10.1136/bmjopen-2025-100897 on 8 July 2025. Downloaded from http://bmjopen.bmj.com/ on July 14, 2025 at Bodleian Libraries of the University of Oxford . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

To disseminate our findings, we will prepare at least two manuscripts for publication in peer-reviewed journals (wave-1 and wave-2 results) and present highlights at scientific meetings, such as the EuroQol Plenary Meeting and the International Academy of Health Preference Research. We will also share the statistical code, survey instrument screenshots and deidentified data on reasonable request to enhance the transparency and interpretability of our results.

Author affiliations
[1]Economics, University of South Florida, Tampa, Florida, USA
[2]Nuffield Department of Population Health, University of Oxford, Oxford, UK
[3]Health Economics, The University of Melbourne School of Population and Global Health, Melbourne, Victoria, Australia
[4]Paediatrics and Child Health, University of Cape Town, Rondebosch, South Africa

Acknowledgements We would like to thank Juan M Ramos-Goñi at Maths in Health for his consultation, assistance and support with the survey instrument, including the survey and question themes for LimeSurvey. Kaizen task is a trademark of Benjamin M Craig.

Contributors BMC is the guarantor. Each coauthor developed the original ideas for the study and secured funding, was responsible for the study design, carried out the experimental design, implemented the online survey instrument, developed the data analysis strategy, drafted the manuscript and appendices.

Funding This work was supported by the EuroQol Research Foundation grant number 1850-RA (PI: Craig). The views expressed by the authors in the publication do not necessarily reflect the views of the EuroQol Group.

Competing interests The authors are members of the EuroQol Group.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

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 Kwon J, Freijser L, Huynh E, *et al*. Systematic Review of Conceptual, Age, Measurement and Valuation Considerations for Generic Multidimensional Childhood Patient-Reported Outcome Measures. *Pharmacoeconomics* 2022;40:379–431.
2 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.
3 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.
4 Stolk EA, Craig BM, Mulhern B, *et al*. Health Valuation: Demonstrating the Value of Health and Lifespan. *Patient* 2017;10:515–7.
5 Craig BM. Health valuation protocol for dual discrete choice experiment (dual-DCE) surveys to estimate the effects of different scenarios and attributes on main effects. *BMJ Open* 2025;15:e091097.
6 Rivero-Arias O, Buckell J, Knight M, *et al*. Defining treatment success in children with surgical conditions. *Arch Dis Child* 2024;109:377–86.
7 Jumamyradov M, Craig BM. Measuring Effectiveness Based on Patient Experience (Instead of QALYs) in US Value Assessments. *Pharmacoeconomics* 2025;43:171–6.
8 Verstraete J, Kreimeier S, Dalziel K, *et al*. PCR245 Stakeholder Engagement and Expert Consultation on the EuroQol Toddler and Infant Populations (EQ-TIPS) Measure of Health-Related Quality of Life (HRQoL). *Value Health* 2023;26:S496–7.
9 Verstraete J, Amien R. The reliability and validity of the isixhosa version of the euroqol toddler and infant populations (EQ-TIPS) health related quality of life instrument. *Afr Health Sci* 2023;23:598–610.
10 Verstraete J, Amien R, Herdien L. EuroQol Toddler and Infant Populations (EQ-TIPS): Age-Related Differences in Performance. *Children (Basel)* 2024;11:1034.
11 Verstraete J, Ramma L, Jelsma J. Item generation for a proxy health related quality of life measure in very young children. *Health Qual Life Outcomes* 2020;18:11.
12 Verstraete J, Ramma L, Jelsma J. Validity and reliability testing of the Toddler and Infant (TANDI) Health Related Quality of Life instrument for very young children. *J Patient Rep Outcomes* 2020;4:94.
13 Craig BM, Pickard AS, Rand-Hendriksen K. Do health preferences contradict ordering of EQ-5D labels? *Qual Life Res* 2015;24:1759–65.
14 Craig BM, Monteiro AL, Herdman M, *et al*. Further evidence on EQ-5D-5L preference inversion: a Brazil/U.S. collaboration. *Qual Life Res* 2017;26:2489–96.
15 Craig BM, Jumamyradov M, Rivero-Arias O. The Performance of Kaizen Tasks Across Three Online Discrete Choice Experiment Surveys: An Evidence Synthesis. *Patient* 2024;17:635–44.
16 Craig BM. United States COVID-19 Vaccination Preferences (CVP): 2020 Hindsight. *Patient* 2021;14:309–18.
17 Hollin IL, Craig BM, Coast J, *et al*. Reporting Formative Qualitative Research to Support the Development of Quantitative Preference Study Protocols and Corresponding Survey Instruments: Guidelines for Authors and Reviewers. *Patient* 2020;13:121–36.
18 Libreri C. *Census personal form*. Australian Bureau of Statistics, 2021.
19 Craig BM, Rand K. Choice Defines QALYs: A US Valuation of the EQ-5D-5L. *Med Care* 2018;56:529–36.
20 Craig BM, Rand K, Bailey H, *et al*. Quality-Adjusted Life-Years without Constant Proportionality. *Value Health* 2018;21:1124–31.
21 McFadden D. Conditional logit analysis of qualitative choice behavior. In: *Frontiers in econometrics*. New York: Academic Press, 1974: 105–42.
22 Lin L-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255–68.
23 Jumamyradov M, Craig BM, Jakubczyk M. Scale and rate heterogeneity in the EQ-5D-5L valuation. *Health Qual Life Outcomes* 2024;22:55.
24 Jumamyradov M, Craig BM, Greene WH, *et al*. Comparing the Mixed Logit Estimates and True Parameters under Informative and Uninformative Heterogeneity: A Simulated Discrete Choice Experiment. *Comput Econ* 2025;65:3295–324.

BMJ Open: first published as 10.1136/bmjopen-2025-100897 on 8 July 2025. Downloaded from http://bmjopen.bmj.com/ on July 14, 2025 at Bodleian Libraries of the University of Oxford . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.