ScienceDirect
Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Preference-Based Assessments

Developing an Australian Value Set for the Recovering Quality of Life-Utility Index Instrument Using Discrete Choice Experiment With Duration

Thao Thai, PhD, Lidia Engel, PhD, Jemimah Ride, PhD, Brendan Mulhern, PhD, Richard Norman, PhD, Cathrine Mihalopoulos, PhD

ABSTRACT

Objectives: The Recovering Quality of Life-Utility Index (ReQoL-UI) instrument was designed to measure the quality-of-life outcomes for people older than 16 years with mental health problems. We aimed to elicit societal preferences for the ReQoL-UI health states to facilitate better decision making in Australia.

Methods: A discrete choice experiment with duration was embedded in a self-completed online survey and administered to a representative sample (n = 1019) of the Australian adult population aged 18 years and older stratified by age, sex, and geographic location. A partial subset design discrete choice experiment was used with 3 fixed attributes and 5 varying attributes, containing 240 choice tasks that were divided into 20 blocks so that each respondent was assigned a block of 12 choice tasks. The value set was modeled using the conditional logit model with utility decrements directly anchored on the 0 to 1 dead-full health scale. Preference heterogeneity was tested using a mixed logit model.

Results: The final value set reflects the monotonic nature of the ReQoL-UI descriptive systems where the best health state defined by the descriptive system has a value of 1 and the worst state has a value of −0.585. The most important dimension was physical health problems, whereas the least important attribute was self-perception. Sensitivity and preference heterogeneity analyses revealed the stability of the value set.

Conclusions: The value set, which reflects the preferences of the Australian population, facilitates the calculation of an index for quality-adjusted life-years in mental health intervention cost-utility analyses.

Keywords: Australia, mental health, QALYs, recovering quality of life, value set.

VALUE HEALTH. 2025; 28(3):460–469

Highlights

- The Recovering Quality of Life-Utility Index measure was designed to address the diverse range of experiences and impacts, focusing on recovery and offering a comprehensive assessment of individuals' quality of life among people older than 16 years with mental health problems.
- An Australian value set for the Recovering Quality of Life-Utility Index was developed using a discrete choice experiment with duration approach, ranging from the best health state value of 1 to the worst state of −0.585, with physical health as the most significant dimension and self-perception as the least.
- The value set, which reflects the preferences of the Australian population, facilitates the calculation of an index value on a quality-adjusted life-year scale, making it applicable to the cost-utility analysis of mental health interventions in Australia.

Introduction

Mental health problems pose significant challenges in Australia, affecting both individuals and the healthcare system. The prevalence of mental disorders, including depression, anxiety, and substance abuse, remains substantial, with approximately 1 in 5 Australians experiencing a mental illness in any given year.[1] The economic burden associated with mental disorders is substantial, with healthcare expenditures increasing from A$10.9 billion in 2017 to 2018 to A$12.2 billion in 2021 to 2022.[2,3] Moreover, untreated mental health illness can lead to a range of adverse consequences, including lost productivity, unemployment, and homelessness, further compounding the societal impact.[4,5]

Considering the limited available healthcare resources, economic evaluation plays a vital role in informing healthcare decisions and resource allocation. Countries such as Australia, the United Kingdom (UK), Canada, and The Netherlands mandate cost-effectiveness assessments, particularly for pharmaceuticals and medical care.[6,7] Among different economic evaluation methods, cost-utility analysis (CUA) is widely favored by health technology assessment agencies.[6,7] CUA quantifies benefits primarily through quality-adjusted life-years (QALYs), which combines both mortality impacts and health-related quality of life (HRQoL) into a single metric, where the time spent in a particular health state is weighted by a corresponding health state utility value that denotes the value of the HRQoL of that health state.[8] These values represent the strength of preference for a particular health state and are anchored on a cardinal scale between 0 (dead) and 1 (full health).[9] Utility weights are most often obtained through the use of preference-based HRQoL measures that comprise a descriptive system that measures various dimensions of HRQoL through a set of items (or questions) and an algorithm used to convert responses from the descriptive system to a single, numeric utility value.[9] This algorithm weights the various dimensions of HRQoL measured by the instrument against each other and allows combinations of answers to be expressed as a weight on the 0 to 1 scale.

Various methods exist for evaluating health states, each differing in complexity, practicality, and theoretical underpinnings. The simplest method is the visual analog scale, in which individuals rate their health on a (usually) linear scale. In contrast, techniques such as time trade-off (TTO) and standard gamble involve more complex decisions, with TTO requiring participants to choose between different health durations and standard gamble involving choices about probabilistic risks to life. Discrete choice experiments (DCEs) have gained popularity in eliciting health preferences to derive utility algorithms and value sets for various HRQoL measures[10-13] because they are arguably less cognitively taxing than other techniques and have strong theoretical foundations.[14]

In mental health, the appropriateness of preference-based instruments in measuring HRQoL remains a question. Although the EQ-5D and SF-6D (derived from SF-36 or SF-12) are sufficiently sensitive to reflect the impact of milder mental health problems, they may lack sensitivity for more severe conditions, such as schizophrenia[15-17] or bipolar disorder.[17] The Recovering Quality of Life (ReQoL) measure was designed to address the diverse range of experiences and impacts, with a focus on recovery and offering a comprehensive assessment of individuals' quality of life among people older than 16 years with mental health problems.[18] It comes in 2 versions: one with 10 items (the ReQoL-10) and one with 20 items (ReQoL-20).[18] The questionnaire was developed to be used as a routinely collected patient-reported outcome measure in the UK. A value set called the ReQoL-Utility Index (ReQoL-UI) using TTO has been developed for the UK context and allows QALY calculation.[19]

To accurately derive QALYs for CUAs, it is crucial to use a preference-based scoring algorithm that aligns with societal preferences for ReQoL-UI health states. However, the currently available algorithm relies on data from the UK general population,[19] inherently reflecting the societal preferences specific to that country. Given the significant variability in health preferences across countries due to cultural, social, economic, and historical factors, relying solely on UK-based data is unlikely to accurately capture the preferences of the Australian population.[20] Indeed, the Australian Pharmaceutical Benefits Advisory Committee specifically states their preference to use Australian-based preference weights in the calculation of utility weights.[6] Our study aimed to elicit societal preferences for the ReQoL-UI health states using DCE within the Australian context to facilitate more informed decision making, resulting in more efficient use of healthcare resources in Australia.

Methods

DCE With Duration

DCEs with duration (DCE_{TTO}) are increasingly recognized for their effectiveness in eliciting preferences and constructing value sets for preference-based measures in health economics.[12,13] This method extracts the relative utility of dimensions and levels within preference-based measures based on random utility theory[21] and Lancaster's consumer demand theory.[22] DCEs use experimental designs to present respondents with hypothetical choice tasks, each offering multiple alternatives described by attribute levels derived from the preference-based measure's descriptive system. Respondents select their most preferred alternative in each task. All responses are then pooled to estimate mean utility weights for each attribute level. The DCE_{TTO} method was designed to anchor relative preferences directly to the utility scale necessary for calculating QALYs by integrating a duration attribute for each health state under evaluation,[10] so directly observing the trade-off both among different attributes of health and between those and time spent in the health state.

The ReQoL-UI Measure

The ReQoL-UI originates from the ReQoL-10 items and ReQoL-20 items.[19] The ReQoL-UI classification system (see Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.12.008) comprises 1 physical health item and 6 mental health items, of which 3 are positively worded and 3 are negatively worded.[19] The same 5 response levels are attached to 6 mental health items ranging from "never" to "most or all of the time" whereas response levels attached to the physical health items range from "no problems" to "very severe problems."

Experimental Design

Choice tasks were presented in pairs of health scenarios described by the 7 items of the ReQoL-UI and a duration attribute (see Appendix Fig. 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.12.008). Following previous practice, the duration attribute was allocated to 4 levels (1, 4, 7, and 10 years).[23,24]

The full factorial generates 312 500 ($5^7 \times 4$) possible ReQoL health states, which is infeasible to present to respondents. As such, we chose the total number of choice tasks of 240 to ensure the estimation of the health state dimensions and duration attributes. The experiment was designed to allow the estimation of 28 main effects (7×4), 1 continuous duration main effect, 28 2-factor interactions between each move from full health in each dimension and the linear term on duration to allow the estimation of utility value decrements comprising the value set, and 21 3-interaction factors between duration and any pair of dimensions where both appear at level 5, for a total of 78. The inclusion of 3-way interactions potentially captures additional decrements when any 2 dimensions are present at the most severe levels. This is based on the assumption that the impact of the most severe level in 1 dimension is not independent of the most severe levels in other dimensions.[25,26]

The design was developed using a modified Fedorov algorithm and incorporating small nonzero priors for 2-factor interaction terms, which signify monotonically increasing severity across levels within each dimension.[27] A C-efficient design that optimizes the coefficient ratios was generated in Ngene (https://www.choice-metrics.com/) to generate 240 choice tasks. These choice tasks were divided into 20 blocks such that each participant saw 12 choice tasks. To reduce cognitive burden, we adopted a partial subset design where we imposed 3 overlapping attributes while varying 5 attributes in each choice task.[28] A candidate set (n = 10 000) of random pairs that imposed the overlapping constraints across 3 attributes was used as input for Ngene 1.2.1 to generate the partial subset design.[29] The order of choice tasks within each block and the overall block order were randomized for each participant. In addition, the left-right order of alternatives within choice tasks was randomized between individuals. Furthermore, attribute order randomization was used, with blocks of positively and negatively worded attributes randomized and attributes within each block also randomized.

Implausible combinations were identified based on the qualitative feedback from respondents through pilot testing. Thus, constraints were applied in the candidate set as input for the DCE design to avoid the occurrence of implausible combinations including hope at level 5 ("Most or all of the time I think my life is not worth living") combined with either well-being at level 1 ("I feel happy most or all of the time") or self-perception at level 1 ("I feel confident in myself most or all of the time").

Survey Structure

The survey comprised 4 main sections. First, respondents answered sociodemographic questions to determine eligibility and ensure the representativeness of the sample compared with the Australian population. Subsequently, respondents completed the ReQoL-10 questionnaire to assess their QoL and familiarize themselves with the ReQoL-UI and response ranges. The third section comprised DCE choice tasks, which began with a warm-up example choice task in which one scenario was less severe than the other in all dimensions (dominant task). A walk-through instructional slideshow was then presented to guide respondents in completing the tasks, including reminders about attribute-level reversals between positively and negatively worded attributes and overlapping attributes. A second dominant choice task was followed to ensure their understanding and engagement with choice tasks. The walk-through instructional slideshow was presented again if respondents failed the second dominant test, with participants then asked to reconsider their choices. Subsequently, a block of 12 choice tasks was randomly administered, followed by some debriefing questions assessing the perceived difficulty and clarity of the valuation tasks, and the choice strategy used in the valuation tasks. The last session comprised the EQ-5D-5L[30] and EQ-Health and Wellbeing[31] questionnaires, followed by demographic questions. Respondents' transitions between choice tasks were recorded to assess time spent on all 12 choice tasks.

To ensure respondents' engagement and understanding, we incorporated quality checks into the questionnaire. For instance, respondents were asked to confirm basic information twice, such as age and location, survey topic, and open-ended questions, to avoid bot responses.[32]

Pretest Phase

The DCE tasks were tested with a convenience sample of 7 academics and students using the think-aloud approach[33] to ensure clarity and effectiveness in eliciting preferences. The survey was tested using a convenience sample of 5 individuals to confirm clarity for the target population. A pilot study followed

Table 1. Sample characteristics.

Variables	Value	n (%)	Comparable Australian data (%)	Difference
Age (years)	18-24	119 (11.68)	11.07%	0.61
	25-34	204 (20.02)	18.07%	1.95
	35-44	195 (19.14)	17.51%	1.63
	45-54	176 (17.27)	16.30%	0.97
	55-64	128 (12.56)	15.11%	−2.55
	65-74	103 (10.11)	12.33%	−2.22
	75-99	94 (9.22)	9.61%	−0.39
Gender	Male	498 (48.87)	48.78%	0.09
	Female	521 (51.13)	51.22%	−0.09
Location	NSW	345 (33.86)	31.78%	2.08
	VIC	269 (26.4)	25.70%	0.70
	QLD	172 (16.88)	20.12%	−3.24
	SA	70 (6.87)	7.13%	−0.26
	WA	110 (10.79)	10.36%	0.43
	ACT	20 (1.96)	2.24%	−0.28
	TAS	21 (2.06)	1.79%	0.27
	NT	12 (1.18)	0.88%	0.30
Country of birth	Australia	793 (77.82)		
Highest education	Year 10 or less	89 (8.73)		
	Year 11/12	197 (19.33)		
	III/IV or diploma	296 (29.05)		
	University degree	425 (41.71)		
Gross household income per annum	<$25 999	73 (7.16)		
	$26 000-$64 999	275 (27)		
	$65 000-$103 999	229 (22.5)		
	$104 000-$207 948	285 (28)		
	>$208 000	75 (7.4)		
	Not answer	82 (8.05)		
Marital status	Single	286 (28.07)		
	Married/de facto	615 (60.36)		
	Separated/divorced	85 (8.34)		
	Widow	30 (2.94)		
	Not answer	3 (0.29)		

ACT indicates Australian Capital Territory; NSW, New South Wales; NT, Northern Territory; QLD, Queensland; SA, South Australia; TAS, Tasmania; VIC, Victoria; WA, Western Australia.

Table 2. Perceived difficulty of the valuation tasks.

Response options	How difficult was it to make a choice in these choice questions?	I found it difficult to imagine the health states	I found it difficult to see the difference between health states	I found it difficult to choose between health states	I found it difficult to consider all aspects when choosing between health states
1	6.48	16.19	27.87	14.33	16.29
2	34.05	27.87	32.78	26.50	29.15
3	25.12	39.94	28.46	33.46	28.36
4	28.26	12.27	8.83	20.22	19.63
5	6.08	3.73	2.06	5.50	6.58

Note. All answers were provided on 5-point Likert response scales ranging from 1 (none of the time) to 5 (most of the time) except for "How difficult was it to make a choice in these choice questions?" ranging from 1 (very easy) to 5 (very difficult).

with the first 100 respondents, assessing completion times, randomization, and open-ended feedback. Given that no changes were needed, pilot data were included in the final data set, and the survey proceeded to the main data collection phase.

Study Sample

Data were collected from the general Australian population aged 18 years and older using an online survey administered by Pureprofile (www.pureprofile.com). Quota sampling based on age, sex, and jurisdiction was used to improve representativeness. Respondents received financial rewards after survey completion, according to Pureprofile's policy. Those who failed quality check tests were excluded from the final data set. The study was approved by the Ethical Review Committee of [Anonymous] (reference no: 39392).

Data Analysis

The utility function specification was based on the work by Bansback et al,[10] where the utility *(U)* of option *j* in choice sets for survey respondent *i* was assumed to be

$$U_{isj} = \alpha TIME_{isj} + \beta X'_{isj} TIME_{isj} + \varepsilon_{isj} \quad (1)$$

where X'_{isj} represents a set of dummies associated with the levels of the ReQoL-UI health state presented in option j, whereas β is the corresponding vector of coefficients. The error term is a random error term with an extreme value type 1 distribution. To develop the value set, we estimated a number of models using the conditional logit with the best level of each attribute as the reference level. To account for respondents answering multiple questions, standard errors are adjusted using a clustered sandwich estimator. First, we generated an unadjusted model (model 1) that can include illogical orderings of attribute levels where an increase in health severity may increase utility (whereas a decrease in utility is expected). If disordering existed, we then combined the disordering levels with the adjacent levels to generate a consistent model (model 2). Following the literature,[24,34,35] we then tested the inclusion of a dummy variable (denoted N5), which takes on a value of 1 if any of the attributes are at the worst levels and 0 otherwise (model 3). Model 4 included 3-factor interaction terms between duration and each of the 2-factor interaction terms between attributes at level 5 (AC5 × HO5 and so on). For model selection, we examined the sign, statistical significance, and logical consistency of coefficients; Akaike information criterion[36]; and Bayesian information criterion.[37] A model was preferred if it produced more statistically significant coefficients that had an expected sign (ie, fewer illogical coefficients) and if it had lower Akaike and Bayesian information criteria.

Based on the selected model, the attribute-level interactions were then anchored onto a 0 to 1 utility scale by dividing the attribute-level interactions (β) by the estimated coefficient for duration (α).[10] The coefficient values (β/α) reflect the relative decrease in value compared with the best quality-of-life state across all dimensions, which were the reference levels. Therefore, negative coefficients are expected, following a logical order of decreasing magnitude. Statistical significance levels were assessed at both 1% and 5%. The relative importance of attributes (dimensions) was computed as follows: (1) compute the attribute utility range by subtracting the lowest utility value from the highest utility value of each attribute, (2) compute the total attribute utility range by summing all utility ranges of all attributes, and (3) determine the relative importance of each attribute by dividing the utility range of a specific attribute by the sum of the utility ranges of all attributes.[38-40] The relative importance of attribute levels was assessed based on the magnitude of their corresponding coefficients across all attributes.

Preference heterogeneity and sensitivity analysis

The robustness of the results was assessed using preference heterogeneity and sensitivity analyses. To capture unobserved preference heterogeneity, we used a mixed logit model with random coefficients. Observable preference heterogeneity was examined by adding interactions with age, gender, and mental health status. The sensitivity analysis involved excluding subsets of respondents with potential biases, such as those who completed tasks in less than the 5th percentile of the duration distribution, individuals who failed either of the 2 dominant tests, or those who consistently chose either the left or right option across all choice tasks.

All analyses were conducted in Stata (StataCorp LLC, College Station, TX), except for preference heterogeneity, which was performed in Nlogit (Econometric Software, Inc, Plainview, NY) owing to Stata's limit of 20 random parameters.

Results

Sample Characteristics

The sample characteristics are presented in Table 1. Of the 1019 respondents who completed the online survey, the sample was broadly representative of the Australian population in terms of sex, age, and geographic location. The mean completion time of the choice tasks was 11 minutes (SD 50; range 0.57-788) and the

Table 3. Model selection.

Variable	Model 1		Model 2		Mixed logit model	
Estimated coefficients	Coefficient	SE	Coefficient	SE	Coefficient	SE
Duration	0.4409*	0.02	0.4689*	0.02	−0.5108*	0.04
Activity (I enjoyed what I did)						
Most of the time	0.0000	NA	0.0000	NA	0.0000	NA
Often	0.0127	0.01	0.0000	NA	0.0000	NA
Sometimes	−0.0332†	0.01	−0.0395*	0.01	−0.0430*	0.01
Only occasionally	−0.0520*	0.01	−0.0610*	0.01	−0.0646*	0.01
Never	−0.0871*	0.01	−0.0891*	0.01	−0.1107*	0.01
Self-perception (I felt confident in myself)						
Most of the time	0.0000	NA	0.0000	NA	0.0000	NA
Often	−0.0056	0.01	−0.0083	0.01	−0.0145	0.01
Sometimes	0.0133	0.01	−0.0123	0.01	−0.0121	0.01
Only occasionally	0.0029	0.01	−0.0123	0.01	−0.0121	0.01
Never	−0.0459*	0.01	−0.0123	0.01	−0.0121	0.01
Well-being (I felt happy)						
Most of the time	0.0000	NA	0.0000	NA	0.0000	NA
Often	0.0202	0.01	0.0000	NA	0.0000	NA
Sometimes	−0.0129	0.01	−0.0359*	0.01	−0.0475*	0.01
Only occasionally	−0.0498*	0.01	−0.0703*	0.01	−0.0977*	0.01
Never	−0.1446*	0.01	−0.1507*	0.01	−0.2157*	0.01
Belonging and relationship (I felt lonely)						
Never	0.0000	NA	0.0000	NA	0.0000	NA
Only occasionally	−0.0161	0.01	−0.0204‡	0.01	−0.0270†	0.01
Sometimes	−0.0372*	0.01	−0.0371*	0.01	−0.0524*	0.01
Often	−0.0587*	0.01	−0.0580*	0.01	−0.0775*	0.01
Most of the time	−0.0642*	0.01	−0.0629*	0.01	−0.0863*	0.01
Choice, control, and autonomy (I felt unable to cope)						
Never	0.0000	NA	0.0000	NA	0.0000	NA
Only occasionally	−0.0019	0.01	−0.0049	0.01	−0.0058	0.01
Sometimes	−0.0110	0.01	−0.0049	0.01	−0.0058	0.01
Often	−0.0466*	0.01	−0.0445*	0.01	−0.0600*	0.01
Most of the time	−0.0620*	0.01	−0.0687*	0.01	−0.0830*	0.01
Hope (I thought my life was not worth living)						
Never	0.0000	NA	0.0000	NA	0.0000	NA
Only occasionally	−0.0010	0.01	−0.0073	0.01	−0.0080	0.01
Sometimes	−0.0108	0.01	−0.0284†	0.01	−0.0382†	0.01
Often	−0.0819*	0.01	−0.0838*	0.01	−0.1062*	0.01
Most of the time	−0.1004*	0.01	−0.1079*	0.01	−0.1675*	0.01
Physical health item (Please describe your physical health: problems with pain, mobility, difficulties caring for yourself, or feeling physically unwell)						
Never	0.0000	NA	0.0000	NA	0.0000	NA
Only occasionally	−0.0270‡	0.01	−0.0278‡	0.01	−0.0232‡	0.01
Sometimes	−0.0755*	0.01	−0.0750*	0.01	−0.0925*	0.01
Often	−0.1904*	0.01	−0.1901*	0.01	−0.2626*	0.01
Most of the time	−0.2524*	0.01	−0.2516*	0.01	−0.3369*	0.01
SD						
Duration					0.5092*	0.03
Activity (I enjoyed what I did)						
Most of the time					NA	NA
Often					NA	NA
Sometimes					0.0678*	0.02
Only occasionally					0.0338	0.03
Never					0.0856†	0.03
Self-perception (I felt confident in myself)						
Most of the time					NA	NA
Often					0.0465	0.03
Sometimes					0.0315‡	0.02
Only occasionally					0.0315‡	0.02
Never					0.0315‡	0.02
Well-being (I felt happy)						
Most of the time					NA	NA

continued on next page

Table 3. Continued

Variable	Model 1		Model 2		Mixed logit model	
Estimated coefficients	Coefficient	SE	Coefficient	SE	Coefficient	SE
Often					NA	NA
Sometimes					0.0931*	0.02
Only occasionally					0.0384	0.02
Never					0.1475*	0.02
Belonging and relationship (I felt lonely)						
Never					NA	NA
Only occasionally					0.0006	0.03
Sometimes					0.0059	0.03
Often					0.0077	0.02
Most of the time					0.0553†	0.02
Choice, control, and autonomy (I felt unable to cope)						
Never					NA	NA
Only occasionally					0.0065	0.02
Sometimes					0.0065	0.02
Often					0.0207	0.03
Most of the time					0.0597*	0.02
Hope (I thought my life was not worth living)						
Never					NA	NA
Only occasionally					0.0419‡	0.02
Sometimes					0.0279	0.03
Often					0.0070	0.02
Most of the time					0.0770*	0.02
Physical health item (Please describe your physical health: problems with pain, mobility, difficulties caring for yourself, or feeling physically unwell)						
Never					NA	NA
Only occasionally					0.0787*	0.02
Sometimes					0.0024	0.02
Often					0.0298	0.02
Most of the time					0.0660*	0.01
Estimation statistics						
AIC	13 700		13 800		13 039	
BIC	14 000		14 000		13 394	
N	24 456		24 456		24 456	
n	1019		1019		1019	

Note. Model 1, unadjusted model; model 2, adjusted model where disordering levels were combined with the adjacent levels to ensure monotonicity of utility decrements.
AIC indicates Akaike information criterion; BIC, Bayesian information criterion; NA, not available; SE, standard error.
*$P < .001$.
†$P < .01$.
‡$P < .05$.

median was 4.69 minutes. Approximately 1.4% of the participants (n = 14) were considered speeders (completing the 12 choice tasks under the 5th percentile), 8.2% (n = 84) failed at least 1 dominance test, and 5.5% (n = 55) consistently chose either the left or right option across all choice tasks.

The perceived difficulty of the valuation tasks is presented in Table 2. On average, respondents found it difficult to make a choice in choice questions some of the time (mean score 3.07). However, they found it was easy to imagine the health states (mean score 2.55), see the difference between health states (mean score 2.12), choose between health states (mean score 2.71), and consider all aspects when choosing between health states (mean score 2.63).

ReQoL-UI Estimates

The estimates of the latent utility and QALY scales (with and without imposed constraints) are presented in Table 3. In the unadjusted model (model 1), the coefficients for the interactions between durations and levels of all dimensions were negative as expected, except for level 2 of activity, levels 3 and 4 of self-perception, and level 2 of well-being. All other coefficients are generally aligned logically, indicating that as severity increases, the decrement in utility also increases. Moreover, the duration coefficient exhibited the anticipated positive coefficient, suggesting that individuals generally prefer longer lives. To smooth out the inconsistent coefficients, we combined levels 1 and 2 of activity; levels 3, 4, and 5 of self-perception; and levels 1 and 2 of well-being. The new model produced new illogical estimates between levels 2 and 3 of the control items, necessitating a combination of these 2 levels. The consistent model (model 2) produced logically consistent coefficients across all dimensions, with 21 of 25 being statistically significant coefficients ($P = .05$). In model 3 (see Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.12.008), adding a variable when any

Table 4. The Australian value set.

Variable	Value set decrement	
Estimated coefficients	Coefficient	SE
Duration		
Activity (I enjoyed what I did)		
Most of the time	0.0000	NA
Often	0.0000	NA
Sometimes	−0.0842*	0.02
Only occasionally	−0.1301*	0.02
Never	−0.1900*	0.02
Self-perception (I felt confident in myself)		
Most of the time	0.0000	NA
Often	−0.0178	0.02
Sometimes	−0.0263	0.01
Only occasionally	−0.0263	0.01
Never	−0.0263	0.01
Well-being (I felt happy)		
Most of the time	0.0000	NA
Often	0.0000	NA
Sometimes	−0.0766*	0.02
Only occasionally	−0.1498*	0.02
Never	−0.3215*	0.02
Belonging and relationship (I felt lonely)		
Never	0.0000	NA
Only occasionally	−0.0435†	0.02
Sometimes	−0.0791*	0.02
Often	−0.1236*	0.02
Most of the time	−0.1341*	0.02
Choice, control, and autonomy (I felt unable to cope)		
Never	0.0000	NA
Only occasionally	−0.0105	0.01
Sometimes	−0.0105	0.01
Often	−0.0950*	0.02
Most of the time	−0.1465*	0.02
Hope (I thought my life was not worth living)		
Never	0.0000	NA
Only occasionally	−0.0156	0.02
Sometimes	−0.0606‡	0.02
Often	−0.1787*	0.02
Most of the time	−0.2301*	0.02
Physical health item (Please describe your physical health: problems with pain, mobility, difficulties caring for yourself, or feeling physically unwell)		
Never	0.0000	NA
Only occasionally	−0.0594†	0.02
Sometimes	−0.1598*	0.02
Often	−0.4054*	0.02
Most of the time	−0.5366*	0.02

SE indicates standard error.
*$P < .001$.
†$P < .05$.
‡$P < .01$.

attribute is at its worst levels results in a reversal of estimated coefficients (reducing the utility). Similarly, in model 4, the incorporation of 21 3-factor interaction terms similarly leads to inversions in estimated coefficients (see Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.12.008).

Based on our a priori criteria, the value set is reported for the adjusted model (model 2), where the utility decrements are anchored onto the 1 to 0 full health-dead scale (Table 4). The anchored decrements are shown in Figures 1 and 2. Among 7 attributes, physical health was the most important, contributing 33% to the total utility. The second most important attribute was well-being (22%), followed by hope (15%) and activity (13%). The least important attribute was self-perception (1%) whereas belonging (6%) and control (9%) were the second and third least important attributes.

The relative importance of attribute levels further reveals the population's preference. Across all attribute levels, physical health level 5 had the largest utility decrement, followed by physical health level 4. The third important level was well-being at level 5, followed by hope at level 5. In contrast, control levels 2 and 3 had the smallest utility decrement (0.01), meaning that these differences in levels of the control item had the least impact on utility.

Utility values for each health state are computed by summing the utility decrements to 1. In this classification system, the best health state is assigned a value of 1, whereas the worst health state (5555555) has a value of −0.5851 (worse than being dead) (1+ [−0.1900−0.0263−0.3215−0.1341−0.1465−0.2301−0.5366]). The distribution of utility values across all health states described by the descriptive system of the ReQoL-UI is shown in Appendix Figure 2 (see Appendix Figure 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.12.008).

Preference Heterogeneity and Sensitivity Analysis

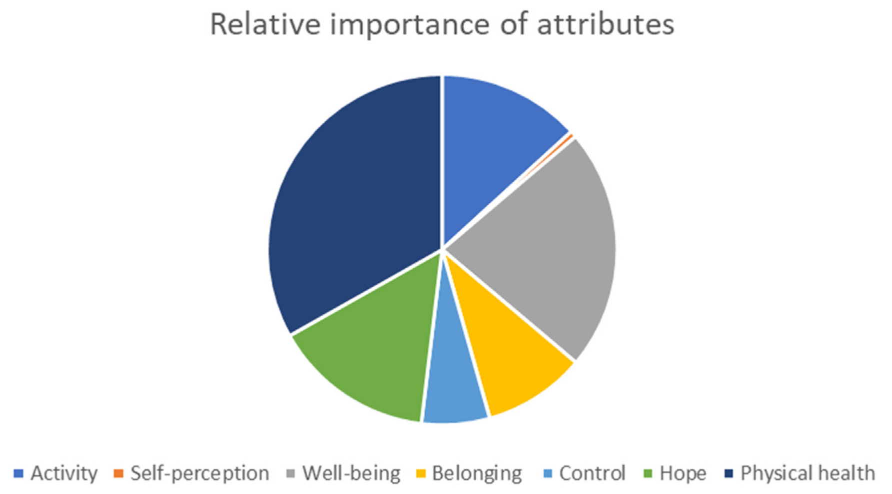
The mixed logit model results (model 3, Table 3) revealed notable preference heterogeneity, particularly within the duration attribute and the worst levels across all items, as evidenced by their statistically significant SDs. Despite this heterogeneity, the mixed logit model results closely aligned with those of the conditional logit model on mean attribute effects, suggesting a stable value set. The observable preference heterogeneity regarding age, gender, and mental health status and sensitivity analysis reinforced the robustness of coefficient estimates (Appendices Tables 2 and 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.12.008).

Discussion

We developed an Australian value set for the ReQoL-UI using a DCE_{TTO} approach. This value set, which reflects the preferences of the Australian population, facilitates the calculation of an index value on a QALY scale, making it applicable to CUAs of mental health interventions in the Australian context. Given that the ReQoL-UI is derived from the ReQoL-10 and ReQoL-20, responses from both questionnaires can be used to calculate utility values using our value set. Given that Australia mainly uses the CUA technique that relies on the calculation of QALYs in guiding public finance decisions for medical services and pharmaceuticals,[41] this study could have a considerable influence on resource allocation procedures in mental health area within the country.

Compared with the UK value set, the Australian value set had a lower value for the worst health state (5555555). In particular, the Australian value set produced a utility of −0.5851 whereas the UK value set produced a utility of −0.195. This could be caused by several reasons. First, different methods may generate different values given that we used a DCE_{TTO} approach whereas the UK value set was derived using the TTO method. Previous studies have shown significant differences in value sets derived from TTO and DCEs.[26,42,43] Furthermore, variations observed between the value sets may partly indicate disparities in preference between the 2 populations. Additional psychometric analyses comparing ReQoL-UI value sets derived from different populations or between known groups would provide further insights into how

Figure 1. Relative importance of attributes.

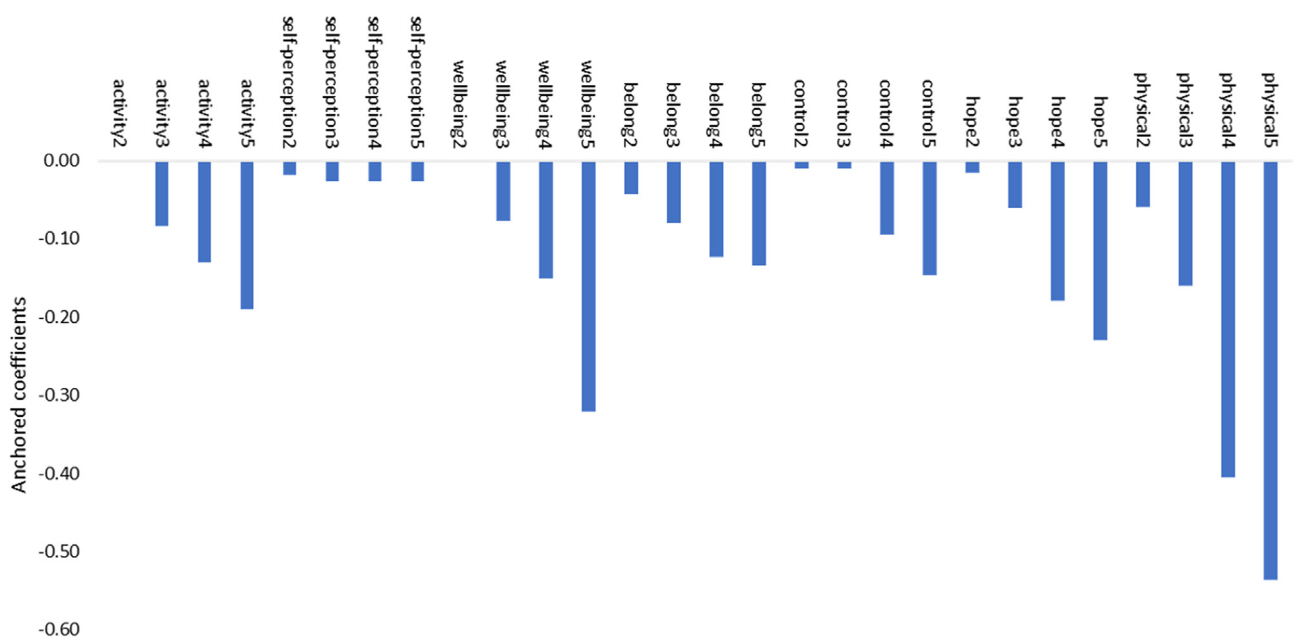


various value sets influence the psychometric properties of the ReQoL-UI and the estimation of QALYs.

We found that the most important dimension for the general Australian population was physical health, which contributed 33% of the utility score, whereas the 6 mental-related items contributed 67%. This finding aligns well with the UK value set of the ReQoL-UI or the value set developed for the Mental Health Quality of Life[44] or the value sets of other generic measures using the general population.[24] The large contribution of the physical health dimension may be due to the descriptor for physical health in the measure, which includes many examples of physical health problems (pain, mobility, difficulties caring for myself, or feeling physically unwell). These may signal a high-severity dimension that may attract respondents' attention. There could be a framing bias in how response options were presented in the ReQoL-UI questionnaire, where physical health items include response options with the word "problems" such as "no problems/ slightly problems/ moderate problems/ severe problems/very severe problems" whereas mental health items do not adopt the same response options but "never/only occasionally/sometimes/often/most or all of the times."

In developing the value set, we tested different specifications including the inclusion of interaction terms as described in the literature.[24,34,35] In our case, including the interaction terms did not improve the value set development in terms of the number of illogical coefficients; thus, we used the main effects approach commonly used in the literature.[12,13] Furthermore, we tested the

Figure 2. Relative importance of item levels of the ReQoL-UI.



ReQoL-UI indicates Recovering Quality of Life-Utility Index.

stability of our estimates using mixed logit modeling and sensitivity analysis by excluding potentially biased respondents. The value set was developed based on a carefully designed DCE with reporting aligned with the DIRECT checklist.[45] Clear instructions on completing choice tasks were presented to participants through a slideshow at the outset.

This study has some limitations. The online sampling method used may not adequately represent all population values, potentially excluding certain disadvantaged groups, such as those unfamiliar with technology or lacking internet access in rural areas. Furthermore, older individuals, particularly those aged 55 years and older, were underrepresented in our study sample. This may be attributed to the tendency for older adults to be less represented in general population online panel cohorts and the potential barriers of limited digital literacy and lack of internet access that may hinder their participation in online recruitment.[46] Given the high prevalence of mental health issues among older people, often driven by loneliness and social isolation,[47,48] future research should aim for a more balanced sample to better represent this population. Resource constraints limited us to a single administration mode; therefore, we cannot entirely rule out its potential influence on our primary findings. However, the sample size of the study was large, covering a broad representation of the population of interest (gender, sex, education, and geographic location). Second, evaluating engagement in online samples poses challenges.[32] To address this, we incorporated quality checks, asking respondents to confirm basic information (eg, age, location) twice and reviewing their engagement and text answers to screen out bot responses. Finally, it is important to acknowledge the ongoing debate surrounding the selection of the most appropriate DCE_{TTO} approach in health valuation studies, including the conventional “pairs” approach commonly adopted,[12,13] the triplets with death approach,[35] and the triplets with full health assuming nonlinear preferences on time approach.[49] Concerns have been raised regarding the use of the death approach, because it may potentially contradict the principles of random utility theory, which underpins the DCE method.[50] Alternatively, the triplets with full health assuming nonlinear preferences on time approach, proposed since 2018, holds promise but has not yet gained widespread acceptance in the literature. Given the absence of consensus on the best approach in health valuation, we opted for the conventional “pair” approach that has been widely adopted in the literature.[12,13]

Conclusion

This study developed an Australian value set for the ReQoL-UI using a DCE_{TTO} approach. This value set, which reflects the preferences of the Australian population, facilitates the calculation of an index value on a QALY scale, making it applicable to CUAs of mental health interventions in the Australian context.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Drs Engel, Mulhern, and Norman are editors of *Value in Health* and have no role in the peer review process of this article.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2024.12.008.

Article and Author Information

Accepted for Publication: December 13, 2024

Published Online: February 8, 2025

doi: https://doi.org/10.1016/j.jval.2024.12.008

Author Affiliations: Monash University Health Economics Group, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia (Thai, Engel, Ride, Mihalopoulos); Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, Australia (Mulhern); Curtin School of Population Health, Curtin University, Perth, Australia (Norman).

Correspondence: Thao Thai, PhD, Monash University Health Economics Group, School of Public Health and Preventive Medicine, Monash University, Level 4, 553 St Kilda Rd, Melbourne, Victoria 3004, Australia. Email: thao.thai@monash.edu

Authorship Confirmation: All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support: This work was supported by the Monash University Health Economics Group grant.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: The authors thank all the participants in this research. The authors also acknowledge the Recovering Quality of Life team at the University of Sheffield for assisting them with early discussions regarding this work.

REFERENCES

1. Australian Bureau of Statistics. Prevalence and impact of mental illness. Mental Health. https://www.aihw.gov.au/mental-health/overview/prevalence-and-impact-of-mental-illness; Published 2024. Accessed February 3, 2025.
2. Australian Bureau of Statistics. Expenditure on mental health-related services. Mental Health. https://www.aihw.gov.au/mental-health/topic-areas/expenditure; Published 2024. Accessed February 3, 2025.
3. Schofield D, Cunich M, Shrestha R, et al. Indirect costs of depression and other mental and behavioural disorders for Australia from 2015 to 2030. *BJPsych Open*. 2019;5(3):e40.
4. Paul KI, Moser K. Unemployment impairs mental health: meta-analyses. *J Vocat Behav*. 2009;74(3):264–282.
5. Fischer PJ, Breakey WR. Homelessness and mental health: an overview. *Int J Ment Health*. 1985;14(4):6–41.
6. Department of Health, Australian Government. Guidelines for preparing a submission to the pharmaceutical benefits advisory committee. https://pbac.pbs.gov.au/content/information/files/pbac-guidelines-version-5.pdf; Published 2016. Accessed February 3, 2025.
7. National Institute for Health Care excellence. Guide to the Methods of Technology Appraisal 2013. National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781; Published 2013. Accessed February 3, 2025.
8. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. 4th ed. Oxford, England: Oxford University Press; 2015.
9. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford, England: Oxford University Press; 2016.
10. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *J Health Econ*. 2012;31(1):306–318.
11. Bansback N, Hole AR, Mulhern B, Tsuchiya A. Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues. *Soc Sci Med*. 2014;114(100):38–48.
12. Wang H, Rowen DL, Brazier JE, Jiang L. Discrete choice experiments in health state valuation: a systematic review of progress and new trends. *Appl Health Econ Health Policy*. 2023;21(3):405–418.
13. Mulhern B, Norman R, Street DJ, Viney R. One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. *Pharmacoeconomics*. 2019;37(1):29–43.
14. Brazier J, Rowen D, Yang Y, Tsuchiya A. 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(5):575–587.
15. Papaioannou D, Brazier J, Parry G. How valid and responsive are generic health status measures, such as EQ-5D and SF-36, in schizophrenia? A systematic review. *Value Health J Int Soc Pharmacoecon Outcomes Res*. 2011;14(6):907–920.
16. Mulhern B, Mukuria C, Barkham M, et al. Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D. *Br J Psychiatry*. 2014;205(3):236–243.
17. Brazier J, Connell J, Papaioannou D, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technol Assess*. 2014;18(34):1–188. vii-viii, xiii-xxv.
18. Keetharuth AD, Brazier J, Connell J, et al. Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. *Br J Psychiatry*. 2018;212(1):42–49.
19. Keetharuth AD, Rowen D, Bjorner JB, Brazier J. Estimating a preference-based index for mental health from the recovering quality of life measure: valuation of recovering quality of life utility index. *Value Health*. 2021;24(2):281–290.
20. Ferreira LN, Ferreira PL, Rowen D, Brazier JE. Do Portuguese and UK health state values differ across valuation methods? *Qual Life Res*. 2011;20(4):609–619.
21. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. *Frontiers in Econometrics*. Cambridge, MA: Academic Press; 1974:105–142.
22. Lancaster KJ. A new approach to consumer theory. *J Pol Econ*. 1966;74(2):132–157.
23. Rowen D, Mulhern B, Stevens K, Vermaire JH. Estimating a Dutch value set for the pediatric preference-based CHU9D using a discrete choice experiment with duration. *Value Health*. 2018;21(10):1234–1242.
24. Mulhern BJ, Bansback N, Norman R, Brazier J, SF-6Dv2 International Project Group. Valuing the SF-6Dv2 classification system in the United Kingdom using a discrete-choice experiment with duration. *Med Care*. 2020;58(6):566–573.
25. Norman R, Viney R, Brazier J, et al. Valuing SF-6D health states using a discrete choice experiment. *Med Decis Mak*. 2013;34(6):773–786.
26. Viney R, Norman R, Brazier J, et al. An Australian discrete choice experiment to value eq-5d health states. *Health Econ*. 2014;23(6):729–742.
27. Rowen D, Powell P, Mukuria C, Carlton J, Norman R, Brazier J. Deriving a preference-based measure for people with Duchenne muscular dystrophy from the DMD-QoL. *Value Health*. 2021;24(10):1499–1510.
28. Jonker MF, Donkers B, de Bekker-Grob E, Stolk EA. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. *Health Econ*. 2019;28(3):350–363.
29. ChoiceMetrics [computer program]. Ngene 1.2 user manual and reference guide. https://choice-metrics.com/NgeneManual120.pdf; Published 2018. Accessed February 3, 2025.
30. Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. *Qual Life Res*. 2021;30(3):647–673.
31. Brazier J, Peasgood T, Mukuria C, et al. The EQ-HWB: overview of the development of a measure of health and wellbeing and key results. *Value Health*. 2022;25(4):482–491.
32. Gonzalez JM, Grover K, Leblanc TW, Reeve BB. Did a bot eat your homework? An assessment of the potential impact of bad actors in online administration of preference surveys. *PLoS One*. 2023;18(10):e0287766.
33. Ryan M, Watson V, Entwistle V. Rationalising the “irrational”: a think aloud study of discrete choice experiment responses. *Health Econ*. 2009;18(3):321–336.
34. Mulhern B, Norman R, Brazier J. Valuing SF-6Dv2 in Australia using an international protocol. *Pharmacoeconomics*. 2021;39(10):1151–1162.
35. Norman R, Mulhern B, Lancsar E, et al. The use of a discrete choice experiment including both duration and dead for the development of an EQ-5D-5L value set for Australia. *Pharmacoeconomics*. 2023;41(4):427–438.
36. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974;19(6):716–723.
37. Schwarz G. Estimating the dimension of a model. *Ann Stat*. 1978;6(2):461–464.
38. Malhotra NK, Nunan D, Birks DF. *Marketing Research : an Applied Approach*. 5th ed. Harlow, UK: Pearson; 2017.
39. Orme B. *Getting Started With Conjoint Analysis: Strategies for Product Design and Pricing Research*. Manhattan, CA: Research Publishers, LLC; 2005.
40. Thai T, Bliemer M, Chen G, Spinks J, de New S, Lancsar E. Comparison of a full and partial choice set design in a labeled discrete choice experiment. *Health Econ*. 2023;32(6):1284–1304.
41. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the pharmaceutical benefits advisory committee [electronic resource]. https://pbac.pbs.gov.au/content/information/archived-versions/pbac-guidelines-v4-5.pdf. Accessed February 3, 2025.
42. Viney R, Norman R, King MT, et al. Time trade-off derived EQ-5D weights for Australia. *Value Health*. 2011;14(6):928–936.
43. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Econ*. 2018;27(1):7–22.
44. van Krugten FCW, Jonker MF, Himmler SFW, Hakkaart-van Roijen L, Brouwer WBF. Estimating a preference-based value set for the mental health Quality of Life Questionnaire (MHQoL). *Med Decis Mak*. 2024;44(1):64–75.
45. Ride J, Goranitis I, Meng Y, LaBond C, Lancsar E. A reporting checklist for discrete choice experiments in health: the DIRECT checklist. *Pharmacoeconomics*. 2024;42(10):1161–1175.
46. Thomas J, McCosker A, Parkinson S, et al. *Measuring Australia's Digital Divide: Australian Digital Inclusion Index: 2023*. Melbourne, Victoria: RMIT university, Swinburne University of Technology, and Telstra; 2023.
47. Lee SL, Pearce E, Ajnakina O, et al. The association between loneliness and depressive symptoms among adults aged 50 years and older: a 12-year population-based cohort study. *Lancet Psychiatry*. 2021;8(1):48–57.
48. Engel L, Mihalopoulos C. The loneliness epidemic: a holistic view of its health and economic implications in older age. *Med J Aust*. 2024;221(6):290–292.
49. Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. Advocating a paradigm shift in health-state valuations: the estimation of time-preference corrected QALY tariffs. *Value Health*. 2018;21(8):993–1001.
50. Flynn TN, Louviere JJ, Marley AAJ, Coast J, Peters TJ. Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. *Popul Health Metr*. 2008;6(1):6.