Social Science & Medicine 246 (2020) 112707

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

# A new tool for creating personal and social EQ-5D-5L value sets, including valuing 'dead'

Trudy Sullivan[a,*], Paul Hansen[b,c], Franz Ombler[c], Sarah Derrett[a], Nancy Devlin[d]

[a] Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand
[b] Department of Economics, University of Otago, Dunedin, New Zealand
[c] 1000minds Ltd, Wellington, New Zealand
[d] Centre for Health Policy, University of Melbourne, Melbourne, Australia

ARTICLE INFO

*Keywords:*
New Zealand
EQ-5D-5L
Value sets
Health-related quality of life (HRQoL)
Health state preferences
Discrete choice experiment (DCE)
PAPRIKA method
Binary search

ABSTRACT

The EuroQol Group's health descriptive systems, the EQ-5D-3L and its successor introduced in 2009, the EQ-5D-5L, are widely used worldwide for valuing health-related quality of life for cost-utility analysis and patient-reported health outcome measures. A new online tool for creating personal and social EQ-5D-5L value sets was recently developed and trialled in New Zealand (NZ). The tool, which includes extensive checks of the quality of participants' data, implements the PAPRIKA method – a novel type of adaptive discrete choice experiment in the present context – and a binary search algorithm to identify any health states worse than dead. After development and testing, the tool was distributed in an online survey in February and March 2018 to a representative sample of NZ adults (N = 5112), whose personal value sets were created. The tool's extensive data quality checks resulted in a 'high-quality' sub-sample of 2468 participants whose personal value sets were, in effect, averaged to create a social value set for NZ. These results overall as well as feedback from participants indicates that the new valuation tool is feasible and acceptable to participants, enabling valuation data to be relatively easily and cheaply collected. The tool could also be used in other countries, tested against other methods for creating EQ-5D-5L value sets, applied in personalised medicine and adapted to create value sets for other health descriptive systems.

## 1. Introduction

The EuroQol Group's health descriptive systems, the EQ-5D-3L (Brooks, 1996) and its successor introduced in 2009, the EQ-5D-5L (Herdman et al., 2011), are widely used worldwide for valuing health-related quality of life (HRQoL) for cost-utility analysis (CUA) and patient-reported health outcome measures (PROMs). Both versions of the system have five HRQoL dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. But instead of just three levels of severity for the EQ-5D-3L, the EQ-5D-5L has five levels: e.g. no, slight, moderate, severe and extreme problems. Thus, the EQ-5D-5L represents 3125 ($5^5$) health states, compared to just 243 ($3^5$) for the EQ-5D-3L. This increased granularity comes at the expense of more complexity and cost to create EQ-5D-5L value sets relative to its simpler predecessor.

An increasingly common approach for creating patient-reported HRQoL instruments, including EQ-5D-5L value sets, is to use discrete choice experiments (DCEs) (McFadden, 1973). In the present context, DCEs involve participants repeatedly choosing between hypothetical health states to reveal the relative importance of the EQ-5D-5L's dimensions. Compared to other choice-based valuation techniques such as the time trade-off (TTO) and standard gamble, DCEs are cognitively less challenging, and so they can be implemented relatively easily and cheaply using online surveys (e.g. Krabbe et al., 2014).

The EuroQol Group's protocol for creating EQ-5D-5L value sets, known as the "EuroQoL Valuation Technology" (EQ-VT), involves computer-supported personal interviews to collect preferences data (Oppe et al., 2014). The EQ-VT comprises 10 TTO questions, including a lead-time TTO for states worse than dead (Robinson and Spencer, 2006; Devlin et al., 2013), and seven DCE questions involving choosing between two hypothetical health states defined on all five dimensions at a time. 'Hybrid' models incorporating both TTO and DCE data have been used to create EQ-5D-5L value sets for: England (Devlin et al., 2018; Feng et al., 2018), Germany (Ludwig et al., 2018), Hong Kong

[*] Corresponding author. Department of Preventive and Social Medicine, University of Otago, Dunedin 9054, New Zealand.
*E-mail addresses:* trudy.sullivan@otago.ac.nz (T. Sullivan), paul.hansen@otago.ac.nz (P. Hansen), franz@1000minds.com (F. Ombler), sarah.derrett@otago.ac.nz (S. Derrett), nancy.devlin@unimelb.edu.au (N. Devlin).

https://doi.org/10.1016/j.socscimed.2019.112707
Received 28 March 2019; Received in revised form 20 November 2019; Accepted 27 November 2019
Available online 30 November 2019

(Wong et al., 2018), Indonesia (Purba et al., 2017), Ireland (Hobbins et al., 2018), Malaysia (Shafie et al., 2019), Poland (Golicki et al., 2019), Portugal (Ferreira et al., 2019), Spain (Ramos-Goñi et al., 2018), Taiwan (Lin et al., 2018) and Thailand (Pattanaphesaj et al., 2018). DCEs have also been used in Australia to value EQ-5D-5L health states (Norman et al., 2013; Viney et al., 2014) and to test DCE design strategies (e.g. Mulhern et al., 2017).

Notwithstanding its widespread use, the EQ-VT, and some applications, has been criticised. Stolk et al. (2019) and Ramos-Goñi et al. (2017) discuss problems with the original version of the protocol used in the first wave of valuation studies (in Canada, England, China and the Netherlands) and mitigations in later versions. A formal review of the EQ-5D-5L value set for England commissioned by the Department of Health for England criticised the EQ-VT (including the latest version) and the quality of the English data and raised serious concerns about the TTO method (Hernández-Alava et al., 2019); for a response, see Van Hout et al. (2019). Subsequently, NICE issued a position statement that they “do not recommend the use of the EQ-5D-5L value set for England” but are committed to working with key stakeholders “to ensure that a 5L value set of an acceptable quality” is developed (NICE, 2019).

This article presents a new DCE-based online tool for creating EQ-5D-5L value sets. The tool, which includes extensive checks of the quality of participants' data, implements a novel type of adaptive DCE in the HRQoL context, known as the PAPRIKA method (Hansen and Ombler, 2008). Relative to other DCE methods, the PAPRIKA method has the major advantage of producing, as well as a *social* value set, a *personal* value set for each participant (i.e. 3125 health state values for each participant). This capability represents a unique opportunity to personalise value sets, consistent with personalised (precision) medicine (Mirnezami et al., 2012). Individual-level data also enable the heterogeneity of individual health state preferences and differences between sub-groups (e.g. healthy vs chronically ill, etc) to be examined. A third benefit of personal value sets is that, as implemented in the tool, a binary search algorithm can be used for participants to identify any health states worse than dead (e.g. Devlin et al., 2019).

The objective of this article is to present the new valuation tool, especially the DCE and binary search algorithm, and to report on the tool's feasibility and acceptability to participants. The setting for developing and trialling the tool is New Zealand (NZ). As well as the authors being from there, the EQ-5D-5L's adoption internationally suggests the need for a NZ value set given the EQ-5D-3L value set created in 1999 (Devlin et al., 2003) is used extensively by researchers and policy-makers – e.g. for CUA by NZ's Pharmaceutical Management Agency (PHARMAC, 2017) and PROMs by the Ministry of Health (Shuker et al., 2017). Hence, the results from creating a NZ social value set are also presented.

## 2. Methods

### 2.1. Tool development and testing

For testing and development purposes, the preliminary version of the tool comprised a DCE (explained in the next sub-section) and two methods for identifying health states worse than dead: a binary search algorithm (explained in sub-section 2.3) and an interactive visual analogue scale (VAS). An example of a VAS question appears in Appendix 1 along with an earlier presentation of the binary search question (superseded by the final version below). The feasibility and acceptability of the tool was evaluated in 12 ‘think-aloud’ (Gilhooly and Green, 1996) sessions, each lasting about an hour in which a participant verbalised their thoughts while using the tool in the presence of two interviewers who observed and asked questions. Purposive sampling (Etikan et al., 2016) was used to recruit 12 participants who varied in terms of age, gender and ethnicity. The interviewers comprised two of the authors (TS or SD) and a third person.

In short, despite most participants finding it more challenging to answer questions involving a binary assessment between a health state and dead (Fig. 2) than to use a VAS slider (Fig. A1 in Appendix), their answers more accurately reflected their thoughts; therefore, the binary search algorithm was adopted for the tool. Participants' feedback was used to improve the tool's instructions and user interface. The tool was then pilot-tested on a snowball sample of 270 participants and refined further.

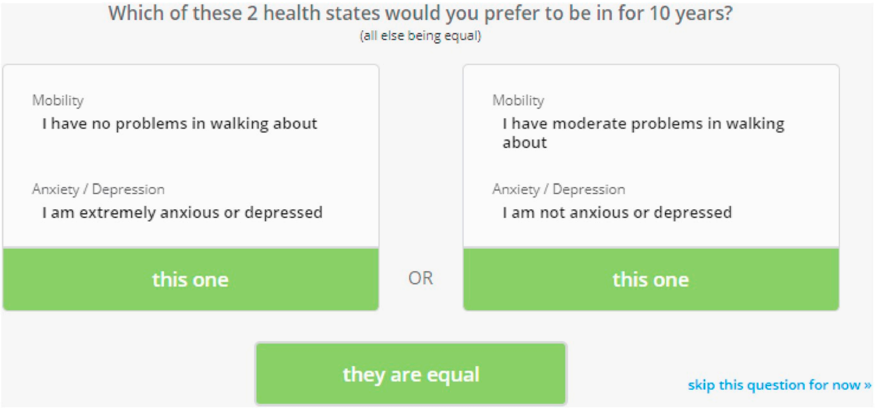
The tool's two main components, in their final forms, are now explained in turn.

### 2.2. Discrete choice experiment

The DCE is based on the PAPRIKA method (Hansen and Ombler, 2008) – an acronym for ***P***otentially ***A***ll ***P***airwise ***R***an***K***ings of all possible ***A***lternatives – as implemented by 1000minds software (www.1000minds.com). Since 2004, this method and software have been used in a wide range of health applications (but not HRQoL until now): health technology prioritisation (Martelli et al., 2016; Sullivan and Hansen, 2017), patient prioritisation (Fitzgerald et al., 2011; Hansen et al., 2012), disease classification and diagnosis (Shiboski et al., 2017) and prioritising diseases for R&D (Tacconelli et al., 2018).

In the present context, the PAPRIKA method involves the participant being repeatedly asked to choose between two hypothetical health states defined on just two dimensions at a time with respect to which state they would prefer to be in for 10 years. Each choice involves a trade-off between the levels for the two dimensions, where implicitly the levels on the other three dimensions are the same for both states (i.e. “all else being equal”). An example of a question appears in Fig. 1.

Such questions (always involving a trade-off between the dimensions,



Which of these 2 health states would you prefer to be in for 10 years?
(all else being equal)

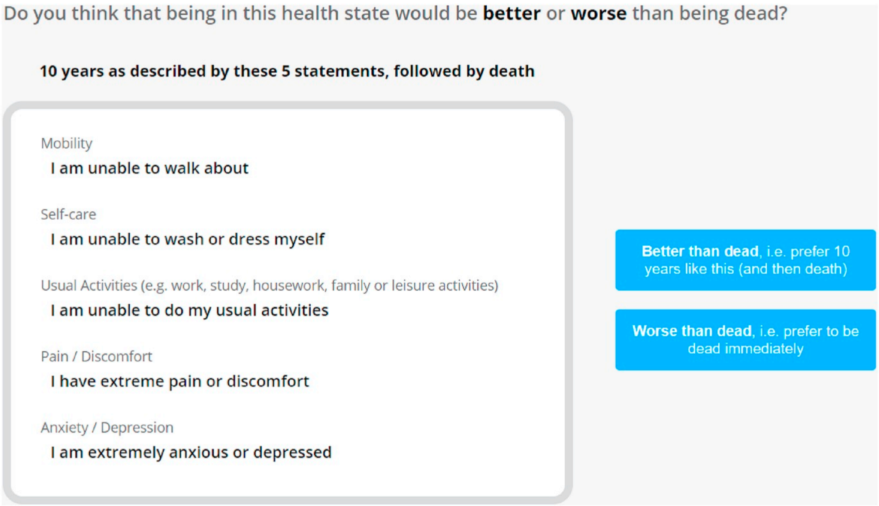
Mobility
I have no problems in walking about

Anxiety / Depression
I am extremely anxious or depressed

Mobility
I have moderate problems in walking about

Anxiety / Depression
I am not anxious or depressed

**Fig. 1.** Example of a DCE question from the 1000minds software.



Do you think that being in this health state would be **better** or **worse** than being dead?

**10 years as described by these 5 statements, followed by death**

Mobility
I am unable to walk about

Self-care
I am unable to wash or dress myself

Usual Activities (e.g. work, study, housework, family or leisure activities)
I am unable to do my usual activities

Pain / Discomfort
I have extreme pain or discomfort

Anxiety / Depression
I am extremely anxious or depressed

**Better than dead**, i.e. prefer 10 years like this (and then death)
**Worse than dead**, i.e. prefer to be dead immediately

**Fig. 2.** Example of a binary search question to identify states worse than dead.

two at a time) are repeated with different pairs of hypothetical health states, which are randomly drawn and positioned with respect to being on the left- or right-hand side of the computer screen. Each time a participant ranks a pair of states, all other states that can be pairwise ranked via transitivity are identified and eliminated, thereby minimising the number of questions asked; for example, if a person prefers state *A* to *B* and *B* to *C*, then, by transitivity, *A* is also preferred to *C* (and is not asked about). Also, each time a person answers a question, based on all preceding answers PAPRIKA adapts with respect to choosing the next question (always one whose answer is not implied by earlier answers). Thus, PAPRIKA is a type of *adaptive* DCE, which when combined with the above-mentioned elimination procedure serves to minimise the number of questions a participant is asked while ensuring they end up having pairwise ranked all possible states defined on two dimensions at a time, either explicitly or implicitly (by transitivity).

Finally, from the participant's explicit pairwise rankings (i.e. answers to the questions) the software uses linear programming techniques to derive weights for the levels on each dimension; for technical details, see Hansen and Ombler (2008). As well as each participant's weights and corresponding personal value set (immediately reportable to the participant), these individual outputs are averaged across all participants to produce social weights and a social value set.

To restrict the number of questions asked, only levels 1, 3 and 5 of each of the EQ-5D-5L dimensions were included in the DCE. The weights for levels 2 and 4 are interpolated using Bézier interpolation (Farin et al., 2002) – in essence, fitting a monotonic smoothed curve through the weights for levels 1, 3 and 5. Also, five combinations of levels (health states) likely to be unrealistic to most people were suppressed: e.g. “no problems doing my usual activities” and either “extreme pain or discomfort” or “extremely anxious or depressed” or “unable to wash or dress myself” or “unable to walk about”.

### 2.3. Binary search to identify health states worse than dead

Enabled by the existence of a personal value set for each participant, an interactive binary search (or bisection) algorithm was implemented for participants to identify any health states they consider to be worse than dead.

The algorithm begins with the participant being asked if they think that being in the lowest-ranked heath state, 55555, for 10 years – the conventional time period used in the EQ-VT (Oppe et al., 2014) – would be better than dead (BTD) or worse than dead (WTD); this question is shown in Fig. 2. If the person answers 55555 is BTD, the algorithm stops. If instead they answer 55555 is WTD, the algorithm proceeds to search for, in effect, the ‘dividing line’ that splits their ranking of the 3125 states into ones BTD and WTD respectively.

Thus, if the person answers 55555 is WTD, they are asked if another, higher-ranked health state – set by the tool to 33333 – is BTD or WTD. Depending on their answer, another higher- or lower-ranked state is evaluated: if 33333 is WTD, 22222 is posed next; instead if 33333 is BTD, 44444 is posed next. Having identified the range of health states in which dead lies, the algorithm proceeds to repeatedly bisect (halve) the participant's personal ranking of states.

For example, with reference to the questions above, suppose the person answers 33333 is BTD and then 44444 is WTD; they are then asked if the state in the middle of *their* ranking of 33333–44444 – e.g. perhaps 34432 (it depends on *their* ranking) – is BTD or WTD. Suppose 34432 is BTD; they are then asked if the state in the middle of their ranking of 34432–44444 – e.g. perhaps 44433 – is BTD or WTD. This process continues: repeatedly halving the range of values until the dividing line is found that splits their ranking of the 3125 states into ones BTD and WTD respectively.

In summary, three results with respect to the location of dead within the 3125 health states are possible: either dead is worse than 55555, and so dead and 55555 are both valued at 0 (customary for EQ-5D valuations); or 11111 is WTD (uncommon), and so dead = 1; or (most often) dead is spanned by two adjacent states in the person's ranking (one BTD, the other WTD), and so dead's value (before rescaling) is the average of these two states' values.

The binary search algorithm explained above differs from traditional implementations in the following four ways that were designed to reduce participants' cognitive effort.

One, the interface does not permit the health state under consideration to be valued as equal to dead (indifference). When confronted with a choice between two undesirable alternatives (a poor health state and dead), some people equate them, meaning “they are both unacceptable” – logically, not the same as “they are *equally* unacceptable”. To eliminate this potential misunderstanding and resulting error in valuing dead, an “equal to dead” button was not provided (e.g. Fig. 2).

Two, the algorithm begins by presenting 55555 rather than the state in the middle of the participant's personal value set. This comparison results in people who regard any HRQoL (including 55555) as BTD having to answer only one question.

Three, for the first three questions asked by the algorithm, each participant is presented with ‘balanced’ states comprising identical levels (e.g. 33333) rather than the states exactly bisecting their personal ranking of states. Balanced states are easier to think about, and so people are less likely to make mistakes.

Four, instead of performing the binary search across all 3125 states, for simplicity a subset is used for each participant that still ensures reasonable accuracy. This subset (for each participant) is formed by grouping the 3125 states by their values rounded to two decimal places and selecting one from each group; thus, the subset has a maximum of 101 states with rounded values in the range 0–1. Consistent with 'balanced' states (as above), the state selected from each group has the lowest standard deviation across its levels (because this state is easier to think about); e.g. if 33233 and 12345 are in the same group, 33233 is selected.

Despite these refinements reducing participants' elicitation burden with respect to cognitive difficulty, the first three refinements may increase the number of questions asked relative to a 'pure' binary search algorithm – i.e. ranging from one (if 55555 is BTD) to approximately nine questions.

### 2.4. New Zealand survey

International research company Dynata was engaged to recruit a sample representative of the NZ adult population with respect to age, gender, ethnicity and geographic location. To ensure that the final sample was as representative of the NZ adult population as possible, if a particular demographic group was under-represented – e.g. due to low participation in the survey, or because their data were excluded for data quality reasons (discussed below) – an on-going recruitment process ensured that this group was over-sampled, thereby delivering a final sample as representative of the NZ adult population as possible.

In February and March 2018, participants were asked to answer anonymously three parts of the online survey: (1) rate their current health status on the EQ-5D-5L questionnaire and EQ VAS, (2) complete the DCE and identify any health states worse than dead, and (3) provide their socio-demographic and background information and feedback about the tool and survey overall. The research was approved by the University of Otago Human Ethics Committee (D17/297).

### 2.5. Participant exclusions

The quality of participants' data is assessed in several ways, with the objective of enabling participants with 'low-quality' data to be identified and excluded – in the process, retaining a sub-sample of 'high-quality' data for creating the social value set.

First, participants who indicated that their responses were invalid – e.g. "my initial responses could have been invalid, because I was floundering with the technology" – were excluded. Also excluded were participants who value dead at unity – because rescaling their value set relative to 11111 = 1 and dead = 0 (see next sub-section) is mathematically impossible.

To test the consistency (reliability) of each participant's answers, the tool repeats two questions at the end of the DCE. The time taken by each participant to answer their DCE questions is also recorded. Thus, participants who fail to answer their two repeated questions identically (consistently) and/or who answer their questions implausibly fast – i.e. below a threshold to be determined after scrutiny of the survey dataset – are excluded. The tool also records for each participant how many DCE questions are answered by clicking "this one" on the left-hand side versus "this one" on the right-hand side versus "they are equal" (see Fig. 1 again). Finally, after being shown their ranking of the five dimensions produced by the DCE, participants are asked whether the ranking is as they expected.

### 2.6. Calculating personal and social value sets

Each participant's personal value set of 3125 health state values can be calculated by summing their DCE preference weights for each state's combination of levels on the five dimensions (where, by definition, states 11111 and 55555 are equal to unity and zero respectively). The value $v$ of each health state can then be rescaled using the usual rescaling formula $\frac{v - z}{1 - z}$, where $z$ is the participant's value for dead (from the binary search) (Oppe et al., 2007).

Each participant's personal value set can be represented by a linear equation analogous to ones used in the EQ-5D literature to create social value sets (Oppe et al., 2007). These equations comprise negative 'disutility' coefficients corresponding to level decrements from 'full health' state 11111 = 1, where health state values are calculated by adding unity to the sum of the negative coefficients corresponding to each state's combination of levels on the five dimensions. Each participant's disutility coefficients can be calculated from their DCE weights using the formula $\frac{w - w_1}{1 - z}$, where $w$ is the weight for the level and dimension whose coefficient is being calculated, $w_1$ is the weight of level 1 (the highest weight) of the same dimension and $z$ is the participant's value for dead.

A social value set can be constructed by calculating the mean for each of the 3125 rescaled values across all individuals. The obvious method is to generate all participants' personal value sets and then calculate the mean value across all individuals for each health state. However, a simpler and equivalent method – i.e. less computationally intensive and resulting in the same social value set – is to average the above-mentioned personal disutility coefficients across all participants to determine the *social* disutility coefficients, from which the social value set is calculated by adding unity to the sum of the social coefficients for each state's combination of levels.

### 2.7. Correlations with other country EQ-5D-5L value sets

A cursory comparison of the NZ social value set with other countries' value sets is possible by calculating Spearman ρ and Pearson $r$ correlation coefficients using other countries' value sets available from Ramos-Goñi et al. (2019). Significance levels ($\alpha$) presented in Ombler et al. (2018) – available as an online resource from www.1000minds.com/sectors/health/hrqol – were used to test the coefficients' statistical significance.

## 3. Results

The survey was completed by 5112 people (0.10% of NZ's population of 5 million). Fig. 3 represents the 2644 participant exclusions to obtain a high-quality sub-sample of 2468. Three participants whose feedback indicates their survey responses are invalid were excluded immediately. As represented by the Venn diagram in the figure, also excluded were participants who: value dead at unity, or are inconsistent (failed to answer the two repeated DCE questions identically), or are implausibly fast – i.e. a median time of less than 6 s per answer. This 6 s threshold (corresponding to participants answering half their DCE questions in under a minute) is supported by Fig. 4, which reveals, overall, that the more time participants take to answer, the more consistent they are (also evident in Fig. 3, where of the 2256 inconsistent participants, 683 are implausibly fast). In summary, of the 5112 who completed the survey, 2644 were excluded, leaving a high-quality sub-sample of 2468 for the analysis that follows.

In addition, a reassuring difference between the 2644 exclusions and the n = 2468 sub-sample emerges with respect to people who answered all their DCE questions by clicking the same button. Of the 2644 exclusions, 225 (9%) always clicked the same button: mostly "they are equal". Given most of these 225 also answered implausibly fast, it seems very likely that they carelessly 'clicked their way through' the DCE questions without due consideration. In contrast, just 132 (5%) in the n = 2468 sub-sample always clicked the same button: "they are equal". A close examination of each of these people's data revealed nothing else 'suspicious' about their quality, including their answer to the question about the accuracy of their ranking of the five dimensions and any other general feedback, that might indicate they should also be excluded. And so, given always answering "they are equal" (i.e. indifference) is theoretically plausible, these participants were retained

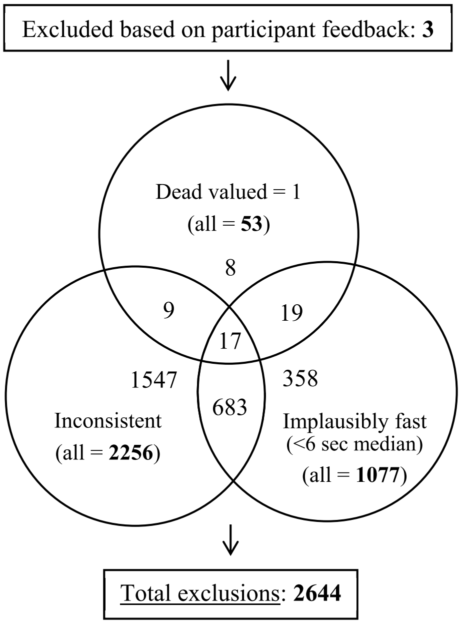


**Fig. 3.** Exclusions to obtain a high-quality sub-sample for creating a social value set (n = 2468).

in the high-quality sub-sample.

The sub-sample's socio-demographic characteristics are reported in Table 1, together with NZ population statistics for comparison purposes. The sub-sample is generally representative of the NZ adult population. The sub-sample's self-assessed health using the EQ-5D-5L is summarised in Appendix 2.

Participants answered 20 DCE questions on average, with a median time of 12.6 s per answer. As reported in Table 2, most participants (63.7%) found the survey instructions and design easy to understand, and almost half found the DCE questions difficult to answer. For most participants (87%), their ranking of the five dimensions produced by the DCE was as they expected. However, of the remainder (13%), usually either just one dimension was mis-ranked or two dimensions were reversed (often despite equal weights, so that if ordinals such as “1st = ” had been displayed, the ranking may have been correct).

The sub-sample's mean weights from the DCE are reported in the second column of Table 3, where they are normalised so that level 1 on all dimensions (i.e. ‘full health’) sums to unity: i.e. 11111 = 1. The disutility coefficients are in the third column – where the social value set can be calculated by adding unity to the sum of the coefficients

**Table 1**
Socio-demographic characteristics of the sub-sample (n = 2468), and for the NZ population.

| Characteristic | Participants | | NZ population %[+] |
|---|---|---|---|
| | n | % | |
| Age (years) | | | |
| 18-24 | 252 | 10.2 | 12.8 |
| 25-34 | 439 | 17.8 | 16.1 |
| 35-44 | 440 | 17.8 | 17.9 |
| 45-54 | 381 | 15.4 | 18.8 |
| 55-64 | 402 | 16.3 | 15.4 |
| 65 + | 554 | 22.5 | 19.0 |
| Gender | | | |
| Male | 1157 | 46.9 | 47.9 |
| Female | 1306 | 52.9 | 52.1 |
| Gender diverse | 5 | 0.2 | unrecorded |
| Ethnicity[a] | | | |
| New Zealand European | 1563 | 63.3 | 64.3 |
| Māori | 390 | 15.8 | 14.1 |
| Pacific | 108 | 4.4 | 6.9 |
| Asian | 340 | 13.8 | 11.1 |
| Other | 329 | 13.3 | 13.6 |
| Education | | 18 + yrs | 15 + yrs |
| No qualifications/Secondary school | 834 | 33.8 | 54.2 |
| University degree or equivalent | 1079 | 43.7 | 20.1 |
| Other post-secondary school qualification | 555 | 22.5 | 25.7 |
| Individual Income | | 18 + yrs | 15 + yrs |
| $20,000 or less | 552 | 22.4 | 38.2 |
| $20,001 - $30,000 | 425 | 17.2 | 13.7 |
| $30,001 - $50,000 | 537 | 21.7 | 21.4 |
| $50,001 - $70,000 | 414 | 16.8 | 12.9 |
| $70,001 - $100,000 | 336 | 13.6 | 7.8 |
| $100,001 or more | 204 | 8.3 | 6.0 |
| Economic Activity | | | |
| Full-time work for pay (30 h or more per week) | 965 | 39.1[#] | |
| Part-time work for pay (< 30 h per week) | 380 | 15.4 | |
| Not in paid work | 237 | 9.6 | |
| Student/Homemaker | 330 | 13.4 | |
| Retired | 491 | 19.9 | |
| Other (including self-employed) | 65 | 2.6 | |
| Long-term Disability (lasting 6 months or more) | | | |
| Yes | 634 | 25.7 | 24.0 |
| No | 1834 | 74.3 | 76.0 |

Notes.
[a] Sums to > 100% as people identify with multiple ethnicities; [+] Statistics from the NZ 2013 Census; # 2017 September quarter employment and unemployment rates were 67.8% and 4.5%.

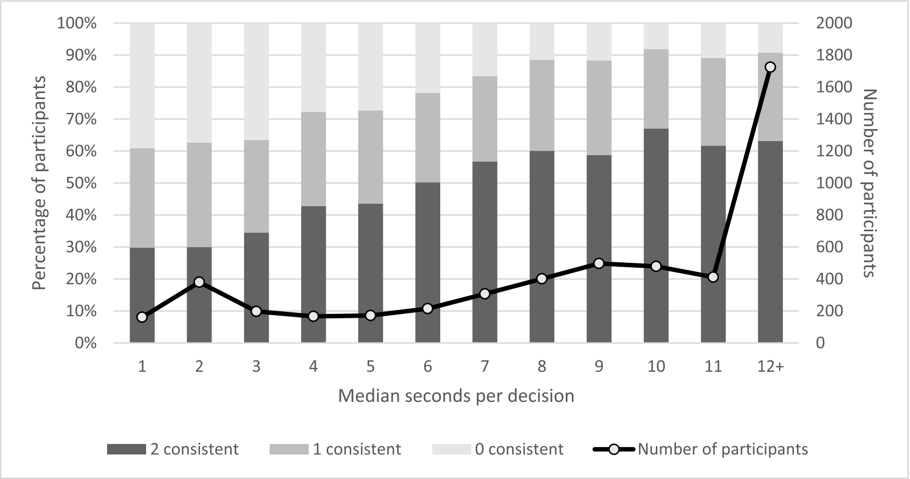


**Fig. 4.** Median time per decision, compared with consistency of answers to two repeated DCE questions.

**Table 2**
Participant feedback about the DCE.

| Feedback | Sub-sample (n = 2468) | |
|---|---|---|
| **Understanding instructions/survey design** | No. | % |
| Very easy/Easy | 1573 | 63.7 |
| Neutral | 668 | 27.1 |
| Very difficult/Difficult | 227 | 9.2 |
| **Choosing between two health states** | | |
| Very easy/Easy | 651 | 26.4 |
| Neutral | 605 | 24.5 |
| Very difficult/Difficult | 1212 | 49.1 |
| **Ranking of the EQ-5D-5L dimensions** | | |
| As expected | 2148 | 87.0 |
| Not as expected | 320 | 13.0 |

corresponding to each state's combination of levels; e.g. 55555's value: 1–0.350–0.370–0.340–0.381–0.389 = −0.830. Analogous calculations for all 3125 health states results in a social value set for NZ. This value set – available on request from the authors – is summarised graphically in Fig. 5, where 780 (25%) of the 3125 states are worse than dead.

As a robustness check, additional participant exclusions based on increasing the 'implausibly fast' threshold to median times of 8 and 10 s per answer respectively were applied to obtain two smaller, arguably higher-quality, sub-samples of n = 2189 and n = 1657. Relative to the original n = 2468 sub-sample, the social value sets from these two smaller sub-samples are very similar: their Pearson *r* correlation coefficients all equal 1.000 and their mean values for dead are 0.338, 0.337 and 0.333. These results suggest that more stringent 'data quality' exclusions would be unlikely to significantly alter the social value set.

With reference to the second column in Table 3, the most important dimension for participants on average is Anxiety/Depression with a mean weight of 0.215, followed by Pain/Discomfort (0.206), Self-Care (0.202), Mobility (0.191) and – least important – Usual Activities (0.186). This result for Anxiety/Depression, for example, means that participants value *not* being anxious/depressed the most; in other words, being extremely anxious/depressed confers the most disutility.

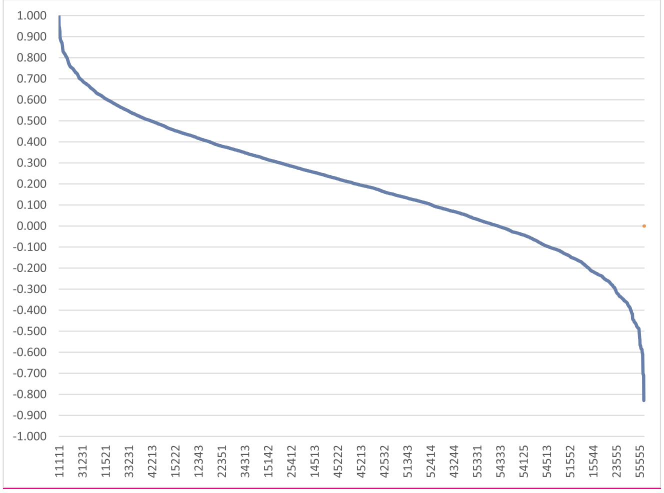


**Fig. 5.** The 3125 health state values (means), from highest (11111 = 1) to lowest (55555 = −0.830).

With respect to their functional 'shape' in terms of marginal effects of level changes, all dimensions exhibit increasing disutility as the levels progressively worsen from "no problems" (level 1) to "moderate problems" (level 3) to "extreme problems" (level 5). Recall that the weights for levels 2 and 4 are interpolated instead of directly created by the DCE – resulting, as can be seen in Table 3, in increasing disutility across all five levels for all dimensions.

Table 4 reports correlation coefficients for the NZ value set vis-à-vis other countries' value sets, their values for 55555 and the number of states worse than dead.

**Table 3**
Mean DCE weights and social disutility coefficients (n = 2468).

| Dimension | Mean DCE weight | Social disutility coefficient |
|---|---|---|
| **Mobility** | | |
| I have no problems in walking about | 0.191 | 0.000 |
| I have slight problems in walking about | 0.159 | −0.056 |
| I have moderate problems in walking about | 0.119 | −0.129 |
| I have severe problems in walking about | 0.065 | −0.229 |
| I am unable to walk about | 0.000 | −0.350 |
| **Self-Care** | | |
| I have no problems washing or dressing myself | 0.202 | 0.000 |
| I have slight problems washing or dressing myself | 0.165 | −0.066 |
| I have moderate problems washing or dressing myself | 0.121 | −0.145 |
| I have severe problems washing or dressing myself | 0.065 | −0.249 |
| I am unable to wash or dress myself | 0.000 | −0.370 |
| **Usual Activities (e.g. work, study, housework, family or leisure activities)** | | |
| I have no problems doing my usual activities | 0.186 | 0.000 |
| I have slight problems doing my usual activities | 0.158 | −0.050 |
| I have moderate problems doing my usual activities | 0.120 | −0.117 |
| I have severe problems doing my usual activities | 0.066 | −0.217 |
| I am unable to do my usual activities | 0.000 | −0.340 |
| **Pain/Discomfort** | | |
| I have no pain or discomfort | 0.206 | 0.000 |
| I have slight pain or discomfort | 0.175 | −0.055 |
| I have moderate pain or discomfort | 0.133 | −0.130 |
| I have severe pain or discomfort | 0.073 | −0.242 |
| I have extreme pain or discomfort | 0.000 | −0.381 |
| **Anxiety/Depression** | | |
| I am not anxious or depressed | 0.215 | 0.000 |
| I am slightly anxious or depressed | 0.174 | −0.072 |
| I am moderately anxious or depressed | 0.126 | −0.157 |
| I am severely anxious or depressed | 0.067 | −0.265 |
| I am extremely anxious or depressed | 0.000 | −0.389 |

**Table 4**
Correlation of the NZ value set with other countries' value sets (EQ-5D-5L).

| | EQ-VT model[a] | Spearman ρ (significance) | Pearson *r* (significance) | 55555 value | No. states worse than dead (%) |
|---|---|---|---|---|---|
| Canada | TTO | 0.957 (0.00) | 0.962 (0.00) | −0.284 | 131 (4) |
| China | TTO | 0.970 (0.00) | 0.973 (0.00) | −0.391 | 315 (10) |
| England | hybrid | 0.948 (0.00) | 0.953 (0.00) | −0.285 | 159 (5) |
| Germany | hybrid | 0.895 (0.13) | 0.905 (0.12) | −0.661 | 471 (15) |
| Hong Kong | hybrid | 0.960 (0.00) | 0.964 (0.00) | −0.865 | 1114 (36) |
| Indonesia | hybrid | 0.913 (0.06) | 0.923 (0.05) | −0.865 | 1108 (35) |
| Ireland | hybrid | 0.904 (0.09) | 0.915 (0.07) | −0.974 | 1124 (36) |
| Japan | TTO | 0.970 (0.00) | 0.972 (0.00) | −0.025 | 1 (0) |
| Netherlands | TTO | 0.903 (0.10) | 0.912 (0.09) | −0.446 | 483 (15) |
| South Korea | TTO | 0.941 (0.01) | 0.945 (0.01) | −0.066 | 4 (0) |
| Spain | hybrid | 0.946 (0.00) | 0.951 (0.00) | −0.416 | 258 (8) |
| Thailand | hybrid | 0.970 (0.00) | 0.972 (0.00) | −0.421 | 187 (6) |
| Uruguay | TTO | 0.949 (0.00) | 0.954 (0.00) | −0.264 | 27 (1) |
| New Zealand | | - - | - - | −0.830 | 780 (25) |

Notes.
[a] Some implementations of the EQ-VT protocol are based entirely on TTO data (in contrast to implementations of 'hybrid' models incorporating both TTO and DCE data).

## 4. Discussion

Fundamental to the new DCE-based tool for creating EQ-5D-5L value sets presented in this article are two important features. One, the tool can be administered using a self-completed online survey, typically taking just 5–10 min to complete – e.g. in contrast to the EuroQol Group's EQ-VT protocol which involves computer-supported personal interviews to collect DCE and TTO data. Two, the tool includes extensive checks of the quality of participants' data – which enabled the identification of a high-quality sub-sample (n = 2468) from which a social value set for NZ was created (as well as 2468 personal value sets).

A cursory comparison of this NZ value set vis-à-vis other countries' value sets reveals it is not obviously atypical; however, further analysis in this respect would be worthwhile. Although three countries (Hong Kong, Indonesia and Ireland) have more states worse than dead than for NZ, the large number for NZ relative to other countries could be further investigated. Are New Zealanders' preferences fundamentally different from other people's? Or does this result arise from the way dead is valued? For example, might such personal questions be answered more honestly in private with the new tool than in the presence of an EQ-VT interviewer?

Although n = 2468 is a large sample (e.g. 0.05% of NZ's population) from which to create a social value set, on the face of it, excluding half – i.e. 2644 (51.6%) – of the survey's original 5112 participants could be considered a high rate of data wastage. However, a low level of participant engagement or understanding is a risk with any survey of the general population, and this is especially so when recruitment is via an online commercial research panel (as for this study). Identifying and excluding disengaged participants is important for "assurance of reliability and validity of the responses" (Schoenherr et al., 2015, p. 294); and reporting such exclusions is important for transparency.

On the other hand, the exclusion of participants who met the time threshold for answering the DCE questions but failed to answer the two repeated questions consistently could perhaps be regarded as too stringent – because, decision-makers are not always perfectly consistent (Thaler, 1980). There are several possible explanations for inconsistencies in the present context, including: as mentioned earlier, participants were disengaged or did not understand the questions; or they were fully engaged and comprehending but were inconsistent in their preferences (e.g. perhaps they changed); or they simply made mistakes. Because unpicking the reasons for inconsistencies was impossible, and because it is important to have confidence in the quality of the data for creating social value sets, participants who were inconsistent – for whatever reason – were excluded. After these exclusions, the 'high quality' sub-sample is generally representative of the NZ adult population due to the sampling strategy.

Both the PAPRIKA DCE method implemented in the tool and the DCE in the EQ-VT are based on participants pairwise ranking hypothetical health states. However, PAPRIKA involves states defined on just two dimensions at a time (i.e. '*partial* profile' DCE) whereas the EQ-VT involves states defined on all five dimensions together ('*full* profile' DCE). In general, a major advantage of pairwise ranking alternatives (here, health states) defined on two dimensions is that this is the cognitively least complex of *all* possible choice tasks; and so people's answers are likely to be more accurate (valid and reliable) than answers to more complex questions.

On the other hand, the validity of people's answers to PAPRIKA's simple questions depends on their preferences not violating 'joint factor' independence (Krantz, 1972). This property, in the present context, arises from the linearity of the equation used for creating social value sets (as is common in the EQ-5D literature mentioned in the Introduction). Joint factor independence requires that when a person is choosing between two health states defined on just two dimensions their choice does not depend on the particular levels on the other three dimensions respectively that are implicitly assumed to be the same for both states (i.e. "all else being equal"; see Fig. 1). From the think-aloud sessions, pilot-testing and population survey, no evidence was found to suggest that these other levels are important when people make their choices.

Notwithstanding the relative simplicity of PAPRIKA's questions, almost half of the high-quality sub-sample reported finding them difficult to answer. This finding is unsurprising given the questions involve confronting trade-offs between EQ-5D-5L dimensions, which is unlikely to be a familiar cognitive task for most people. Future applications of the tool will include a preliminary warm-up exercise involving several practice DCE questions, with the objective of increasing the overall quality of participants' answers by reducing the number of inconsistent responses.

Another important advantage of the PAPRIKA DCE method relative to the DCE in the EQ-VT is that a personal value set is produced for each participant, enabling a binary search algorithm to be used for participants to identify any health states worse than dead. The algorithm developed for the tool includes several refinements to reduce participants' cognitive effort relative to more traditional implementations.

Feedback overall – from the think-aloud sessions, pilot-testing and population survey – indicates the new tool is acceptable in general to participants. The tool's user-friendliness and online delivery could significantly lower the cost of creating EQ-5D-5L value sets. The tool could also support CUA and PROMs at the individual patient level, incorporating the patient's preferences into treatment decisions in 'real time'. For example, the tool could be available on computer tablets in doctor waiting rooms or as a mobile app for patients to quickly create their own personal value sets. The availability of personal value sets also enables any relationships between individuals' preference weights and their observable characteristics, such as age, ethnicity, health status, etc, to be investigated (e.g. using cluster analysis; Späth, 1980).

A potentially important limitation of the study – and another area for future research – concerns the determination of weights for levels 2 and 4 for each dimension. In order to restrict the number of questions asked, these weights were interpolated, with only levels 1, 3 and 5 included in the DCE: resulting in just 20 questions on average, typically taking just 5–10 min in total. (Another potential advantage of level 4 not being included in the DCE is the avoidance of the confusion likely to be experienced by some people when asked to differentiate between levels 4 and 5 on the Pain/Discomfort and Anxiety/Depression dimensions with respect to which of the near-synonyms "severe" versus "extreme" problems is worse.) In future applications, all five levels could easily be included: resulting in approximately 55 DCE questions on average (e.g. 12–25 min in total). Such an elicitation burden is likely to be acceptable for people who are sufficiently motivated. Alternatively, an additional DCE could determine mean weights for levels 2 and 4 for each dimension and then be combined with the current results to refine the social value set. Both approaches would enable each dimension's functional 'shape' in terms of the marginal effects of changes across levels to be investigated.

Possible areas for future research internationally include trialling the new tool in other countries – including leveraging the tool's cost advantages for low and middle-income countries – and testing the tool against other methods for creating EQ-5D-5L value sets. An obvious potential comparator is the EQ-VT protocol. The tool can readily be adapted to create value sets for other health descriptive systems, such as the SF-6D (Brazier et al., 2002), PROMIS (Cella et al., 2010), etc.

## 5. Conclusion

The new valuation tool for creating *personal* and *social* EQ-5D-5L value sets is feasible and acceptable to participants, enabling valuation data to be relatively easily and cheaply collected. As reported on in this article, the tool was first applied in NZ. It could also be used in other countries, tested against other methods for creating EQ-5D-5L value sets, applied in personalised medicine and adapted to create value sets for other health descriptive systems.

## Author Contributions

**Trudy Sullivan**: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualization, Supervision, Project Administration, Funding Acquisition. **Paul Hansen**: Conceptualization, Methodology, Validation, Writing – Original Draft, Writing – Review and Editing, Visualization. **Franz Ombler**: Conceptualization, Methodology, Software, Validation, Formal Analysis, Resources, Data Curation, Writing - Original Draft, Writing - Review and Editing, Visualization. **Sarah Derrett**: Conceptualization, Methodology, Investigation, Formal Analysis, Writing - Review and Editing. **Nancy Devlin**: Conceptualization, Methodology, Writing – Review and Editing.

## Declaration of competing interest

The second and third authors have ownership interests in 1000minds Ltd, whose software is used for the valuation tool; the fourth and fifth authors are members of the EuroQol Group.

## Acknowledgments

Thank you to Josh Ward for helping to develop the survey and facilitating the think-aloud sessions and snowball sample, to our colleagues, friends and families who took part in the pilot-testing, participants in the think-aloud sessions and everyone who completed the snowball and general population surveys. Thanks also for feedback received at the 35th EuroQol Group Scientific Plenary Meeting 2018, and to the editor and anonymous referees for their comments on the manuscript. The Department of Preventive and Social Medicine, University of Otago provided funding for data collection and software administration; 1000minds Ltd supplied the software.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2019.112707.

## References

Brazier, J., Roberts, J., Deverill, M., 2002. The estimation of a preference-based measure of health from the SF-36. J. Health Econ. 21 (2), 271–292.
Brooks, R., 1996. EuroQol: the current state of play. Health Policy 37 (1), 53–72.
Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Cook, K., 2010. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J. Clin. Epidemiol. 63 (11), 1179–1194.
Devlin, N.J., Hansen, P., Kind, P., Williams, A., 2003. Logical inconsistencies in survey respondents' health state valuations – a methodological challenge for estimating social tariffs. Health Econ. 12 (7), 529–544.
Devlin, N., Buckingham, K., Shah, K., Tsuchiya, A., Tilling, C., Wilkinson, G., Van Hout, B., 2013. A comparison of alternative variants of the lead and lag time TTO. Health Econ. 22 (5), 517–532.
Devlin, N.J., Shah, K.K., Feng, Y., Mulhern, B., van Hout, B., 2018. Valuing health-related quality of life: an EQ-5D-5L value set for England. Health Econ. 27 (1), 7–22.
Devlin, N.J., Shah, K.K., Mulhern, B.J., Pantiri, K., van Hout, B., 2019. A new method for valuing health: directly eliciting personal utility functions. Eur. J. Health Econ. 1–14.
Etikan, I., Musa, S.A., Alkassim, R.S., 2016. Comparison of convenience sampling and purposive sampling. Am. J. Theor. Appl. Stat. 5 (1), 1–4.
Farin, G., Hoschek, J., Kim, M.-S., 2002. Handbook of Computer Aided Geometric Design. Elsevier Science Ltd, North Holland.
Feng, Y., Devlin, N.J., Shah, K.K., Mulhern, B., van Hout, B., 2018. New methods for modelling EQ-5D-5L value sets: an application to English data. Health Econ. 27 (1), 23–38.
Ferreira, P.L., Antunes, P., Ferreira, L.N., Pereira, L.N., Ramos-Goñi, J.M., 2019. A Hybrid Modelling Approach for Eliciting Health State Preferences: the Portuguese EQ-5D-5L Value Set. Quality of Life Research, pp. 1–13.
Fitzgerald, A., de Coster, C., McMillan, S., Naden, R., Armstrong, F., Barber, A., ... Lacaille, D., 2011. Relative urgency for referral from primary care to rheumatologists: the Priority Referral Score. Arthritis Care Res. 63 (2), 231–239.
Gilhooly, K., Green, C., 1996. Protocol Analysis: Theoretical Background. Handbook of Qualitative Research Methods for Psychology and the Social Sciences, vol. 14. The British Psychological Society, Leicester, pp. 43–54.
Golicki, D., Jakubczyk, M., Graczyk, K., Niewada, M., 2019. Valuation of EQ-5D-5L health states in Poland: the first EQ-VT-based study in Central and Eastern Europe. PharmacoEconomics 1–12.
Hansen, P., Ombler, F., 2008. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. J. Multi-Criteria Decis. Anal. 15 (3-4), 87–107.
Hansen, P., Hendry, A., Naden, R., Ombler, F., Stewart, R., 2012. A new process for creating points systems for prioritising patients for elective health services. Clin. Gov. Int. J. 17 (3), 200–209.
Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., ... Badia, X., 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual. Life Res. 20 (10), 1727–1736.
Hernández-Alava, M., Pudney, S., Wailoo, A., 2019. The EQ-5D-5L: Value Set for England: Findings of a Quality Assurance Programme. *Value in Health*. (forthcoming).
Hobbins, A., Barry, L., Kelleher, D., Shah, K., Devlin, N., Goni, J.M.R., O'Neill, C., 2018. Utility values for health states in Ireland: a value set for the EQ-5D-5L. PharmacoEconomics 36 (11), 1345–1353.
Krabbe, P.F., Devlin, N.J., Stolk, E.A., Shah, K.K., Oppe, M., Van Hout, B., ... Xie, F., 2014. Multinational evidence of the applicability and robustness of discrete choice modeling for deriving EQ-5D-5L health-state values. Med. Care 52 (11), 935.
Krantz, D.H., 1972. Measurement structures and psychological laws. Science 175 (4029), 1427–1435.
Lin, H.W., Li, C.I., Lin, F.J., Chang, J.Y., Gau, C.S., Luo, N., et al., 2018. Valuation of the EQ-5D-5L in Taiwan. PLoS One 13 (12), e0209344.
Ludwig, K., von der Schulenburg, J.-M.G., Greiner, W., 2018. German value set for the EQ-5D-5L. PharmacoEconomics 36 (6), 663–674.
Martelli, N., Hansen, P., van den Brink, H., Boudard, A., Cordonnier, A.-L., Devaux, C., ...

Borget, I., 2016. Combining multi-criteria decision analysis and mini-health technology assessment: a funding decision-support tool for medical devices in a university hospital setting. J. Biomed. Inform. 59, 201–208.

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, Cambridge, MA.

Mirnezami, R., Nicholson, J., Darzi, A., 2012. Preparing for precision medicine. N. Engl. J. Med. 366 (6), 489–491.

Mulhern, B., Bansback, N., Hole, A.R., Tsuchiya, A., 2017. Using discrete choice experiments with duration to model EQ-5D-5L health state preferences: testing experimental design strategies. Med. Decis. Mak. 37 (3), 285–297.

NICE, 2019. Position Statement on Use of the EQ-5D-5L Value Set for England. (updated October 2019). Retrieved 31 October 2019. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l.

Norman, R., Cronin, P., Viney, R., 2013. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. Appl. Health Econ. Health Policy 11 (3), 287–298.

Ombler, F., Albert, M., Hansen, P., 2018. How significant are "high" correlations between EQ-5D value sets? Med. Decis. Mak. 38 (6), 635–645.

Oppe, M., Szende, A., de Charro, F., 2007. Comparative review of visual analogue scale value sets. In: Szende, A., Oppe, M., Devlin, N.J. (Eds.), EQ-5D Value Sets: Inventory, Comparative Review and User Guide. Springer, Berlin, Germany.

Oppe, M., Devlin, N.J., van Hout, B., Krabbe, P.F., de Charro, F., 2014. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health 17 (4), 445–453.

Pattanaphesaj, J., Thavorncharoensap, M., Ramos-Goñi, J.M., Tongsiri, S., Ingsrisawang, L., Teerawattananon, Y., 2018. The EQ-5D-5L valuation study in Thailand. Expert Rev. Pharmacoecon. Outcomes Res. 18 (5), 551–558.

PHARMAC, 2017. *Prescription for Pharmacoeconomic Analysis* (PFPA). Retrieved 10 December 2018. www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa.

Purba, F.D., Hunfeld, J.A., Iskandarsyah, A., Fitriana, T.S., Sadarjoen, S.S., Ramos-Goñi, J.M., et al., 2017. The Indonesian EQ-5D-5L value set. PharmacoEconomics 35 (11), 1153–1165.

Ramos-Goñi, J.M., Oppe, M., Slaap, B., Busschbach, J.J., Stolk, E., 2017. Quality control process for EQ-5D-5L valuation studies. Value Health 20 (3), 466–473.

Ramos-Goñi, J.M., Oramas-Zarate, J., Rivero-Arias, O., 2019. eq5d5l: A Command to Estimate Preference-Based Values. EuroQol Working Paper Series No. 19001.

Ramos-Goñi, J.M., Craig, B.M., Oppe, M., Ramallo-Fariña, Y., Pinto-Prades, J.L., Luo, N., Rivero-Arias, O., 2018. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. Value Health 21 (5), 596–604.

Robinson, A., Spencer, A., 2006. Exploring challenges to TTO utilities: valuing states worse than dead. Health Econ. 15 (4), 393–402.

Schoenherr, T., Ellram, L.M., Tate, W.L., 2015. A note on the use of survey research firms to enable empirical data collection. J. Bus. Logist. 36 (3), 288–300.

Shafie, A.A., Thakumar, A.V., Lim, C.J., Luo, N., Rand-Hendriksen, K., Yusof, F.A.M., 2019. EQ-5D-5L valuation for the Malaysian population. PharmacoEconomics 37 (5), 715–725.

Shiboski, C.H., Shiboski, S.C., Seror, R., Criswell, L.A., Labetoulle, M., Lietman, T.M., ... Bowman, S.J., 2017. 2016 American College of Rheumatology/European League against Rheumatism classification criteria for primary Sjögren's syndrome: a consensus and data-driven methodology involving three international patient cohorts. Arthritis Rheumatol. 69 (1), 35–45.

Shuker, C., Bohm, G., Hamblin, R., Simpson, A., St George, D., Stolarek, I., Wilson, J., Merry, A.F., 2017. Progress in public reporting in New Zealand since the Ombudsman's ruling, and an invitation. N. Z. Med. J. 130 (1457), 11–22.

Späth, H., 1980. Cluster Analysis Algorithms for Data Reduction and Classification of Objects. Ellis Horwood, Chichester, UK.

Stolk, E., Ludwig, K., Rand, K., van Hout, B., Ramos-Goñi, J.M., 2019. Overview, update, and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L Valuation Protocol. Value Health 22 (1), 23–30.

Sullivan, T., Hansen, P., 2017. Determining criteria and weights for prioritizing health technologies based on the preferences of the general population: a New Zealand pilot study. Value Health 20 (4), 679–686.

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D.L., ... Carmeli, Y., 2018. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect. Dis. 18 (3), 318–327.

Thaler, R., 1980. Toward a positive theory of consumer choice. J. Econ. Behav. Organ. 1 (1), 39–60.

Viney, R., Norman, R., Brazier, J., Cronin, P., King, M.T., Ratcliffe, J., Street, D., 2014. An Australian discrete choice experiment to value EQ-5D health states. Health Econ. 23 (6), 729–742.

Van Hout, B., Feng, Y., Mulhern, B., Shah, K., Devlin, N., 2019. The EQ-5D-5L Value Set for England: Response to the 'quality Assurance'. *Value in Health.* (forthcoming).

Wong, E.L., Ramos-Goñi, J.M., Cheung, A.W., Wong, A.Y., Rivero-Arias, O., 2018. Assessing the use of a feedback module to model EQ-5D-5L health states values in Hong Kong. Patient Patient Center. Outcomes Res. 11 (2), 235–247.