Social Science & Medicine 354 (2024) 117073

Contents lists available at ScienceDirect

Social Science & Medicine

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

Creating an SF-6Dv2 social value set for New Zealand

Trudy Sullivan[a,*], Georgia McCarty[b], Franz Ombler[c], Robin Turner[d], Brendan Mulhern[e], Paul Hansen[c,f]

[a] Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand
[b] Ngāi Tahu Māori Health Research Unit, University of Otago, Dunedin, New Zealand
[c] 1000minds Ltd, Dunedin, New Zealand
[d] Biostatistics Centre, University of Otago, Dunedin, New Zealand
[e] Centre for Health Economics Research and Evaluation, University of Technology Sydney, Australia
[f] Department of Economics, University of Otago, Dunedin, New Zealand

ARTICLE INFO

Handling editor: Richard Smith

Keywords:
SF-6D
SF-6Dv2
New Zealand
Health-related quality of life (HRQoL)
Value sets
Discrete choice experiment (DCE)
PAPRIKA method
Binary search

ABSTRACT

The SF-6D health descriptive system and its second version published in 2020, the SF-6Dv2, is used worldwide for valuing health-related quality of life (HRQoL) for economic evaluation and measuring patient-reported health outcomes. In this study, a valuation tool was developed and applied to create a social value set, comprising 18,750 health state values, for the SF-6Dv2 for New Zealand (NZ). This tool was adapted and extended from the one used to create a social value set for the EQ-5D-5L, a simpler health descriptive system with fewer dimensions and health states. The tool implements the PAPRIKA method, a type of adaptive discrete choice experiment, and a binary search algorithm to identify health states worse than dead and has extensive data quality controls to ensure the validity and reliability of the social value set derived from participants' personal value sets. The tool, accompanied by a short introductory video designed specifically for the SF-6Dv2, was distributed via an online survey to a large representative sample of adult New Zealanders in June–July 2022. The tool's data quality controls enabled participants who failed to understand or sincerely engage with the valuation tasks to be identified and excluded, resulting in the participants being pared down to a sub-sample of 2985 'high-quality' participants whose personal value sets were averaged for the social value set. These results, including participants' positive feedback, demonstrate the feasibility and acceptability of using the tool to value larger health descriptive systems such as the SF-6Dv2. Having successfully created an SF-6Dv2 social value set for NZ, the valuation tool can be readily applied to other countries, used to generate personal value sets for personalised medicine and adapted to create value sets for other health descriptive systems.

1. Introduction

Developed by researchers at the University of Sheffield in 2002, the SF-6D (Brazier et al., 1998, 2002) is a health descriptive system that measures health-related quality of life (HRQoL) on six dimensions: physical functioning, role limitations, social functioning, pain, mental health and vitality. The SF-6D is derived from the Short Form 36 (SF-36) and Short Form 12 (SF-12) (Ware and Sherbourne, 1992) and is used worldwide to inform economic evaluations of health care interventions and as a patient-reported outcome measure (PROM) (CADTH, 2006; Brazier et al., 2007; Delwel, 2008; PBAC, 2015; HIQA, 2020).

In response to criticisms of the SF-6D descriptive system and its value sets (Brazier et al., 2002; Longworth and Bryan, 2003; Brazier et al., 2004; Ferreira et al., 2008; Brazier et al., 2020), the system was revised in 2020, resulting in version 2, the SF-6Dv2 (Brazier et al., 2020). As well as changes to the wording of the system's levels, the SF-6Dv2 has one less level for physical functioning and one more level for role limitations than the SF-6D; for details see Brazier et al. (2020). Thus, the SF-6Dv2 has six levels on the pain dimension and five levels on the five other dimensions and is therefore capable of representing 18,750 possible health states (Brazier et al., 2020). An SF-6Dv2 social value set consists of values for each of these 18,750 health states representing the HRQoL preferences of the population group concerned. These values are anchored in the range of 1 = perfect health (no problems on any of the dimensions) to 0 = dead, and with negative values for states considered worse than dead.

* Corresponding author. Department of Preventive and Social Medicine, University of Otago, Dunedin, 9054, New Zealand.
E-mail address: trudy.sullivan@otago.ac.nz (T. Sullivan).

https://doi.org/10.1016/j.socscimed.2024.117073
Received 28 November 2023; Received in revised form 15 May 2024; Accepted 22 June 2024
Available online 26 June 2024

SF-6Dv2 value sets have been published for Australia (Mulhern et al., 2021), China (Wu et al., 2021) and UK (Mulhern et al., 2020). As New Zealand (NZ) does not have SF-6D or SF-6Dv2 value sets, this study's primary objective was to create an SF-6Dv2 value set to support decision-making about the allocation of health care resources in NZ.

The methods used to create SF-6D value sets, including SF-6Dv1 value sets for 10 countries, have evolved over time; for a review of elicitation methods and modelling strategies used to create SF-6D value sets, see Wang and Poder (2023). To create the Australian and UK SF-6Dv2 value sets, the same protocol based on a discrete choice experiment (DCE) with duration (DCE_{TTO}) was used, as described below (Mulhern et al., 2020, 2021). For China, time trade-off (TTO) and DCE_{TTO} were used, with the final value set generated using the TTO data (Wu et al., 2021). In a separate study, best-worst scaling (BWS) was explored as a method for valuing SF-6Dv2 health states (Osman et al., 2021).

DCEs are increasingly used for creating value sets (Mulhern et al., 2019; Wang et al., 2023). They involve participants repeatedly choosing between two or more hypothetical health states defined by a descriptive system such as the SF-6Dv2 to reveal the relative importance, or weights, of the dimensions and their levels, which are used to produce a value set.

An international protocol for administering the DCE_{TTO} method using online surveys was developed to create the Australian (Mulhern et al., 2021) and UK (Mulhern et al., 2020) SF-6Dv2 value sets. The protocol consists of two choice tasks for participants: (1) choosing between pairs of health states with durations of 1, 4, 7 and 10 years respectively; and (2) choosing between two health states with different durations and immediate death, with the best and worst alternatives chosen (Mulhern et al., 2020, 2021).

In the present study, a valuation tool for creating an SF-6Dv2 social value set for New Zealand was developed and applied via an online survey. The tool has two main components: (1) an adaptive DCE implementing the **P**otentially **A**ll **P**airwise **R**an**K**ings of all **P**ossible **Al**ternatives (PAPRIKA) method (Hansen and Ombler, 2008), and (2) a binary search algorithm to identify health states worse than dead. The tool generates personal SF-6Dv2 value sets corresponding to 18,750 health state values for each participant. These personal value sets are averaged to produce a social value set reflecting population preferences. To ensure the value set's validity and reliability, the tool has extensive data-quality controls to identify and exclude participants who fail to understand or sincerely engage with the valuation tasks.

A similar valuation tool was successfully used to create a social value set for NZ for another widely used health descriptive system, the EQ-5D-5L (Sullivan et al., 2020). Compared to the EQ-5D-5L, which represents 3125 possible health states, the SF-6Dv2 has six times as many states (18,750), requiring participants in the present study to make more choices in the DCE component of the tool. Also, an important lesson from the tool's previous application was the need to better inform and motivate participants, and so a short introductory video was created for participants to watch before they started the present survey.

Thus, as well as the above-mentioned primary objective of creating an SF-6Dv2 social value set for NZ – intended for use by policy-makers and researchers – this study also investigates the feasibility and acceptability of the valuation tool for the 'larger' SF-6Dv2 and the effect of the introductory video on participants' understanding and engagement and hence the quality of their data. The resulting NZ SF-6v2 social value set is also compared with the social value sets for Australia, UK and China.

2. Methods

Implemented using 1000minds software (Ombler et al., 2022), the valuation tool consists of a DCE based on the PAPRIKA method (Hansen and Ombler, 2008) to generate a personal SF-6Dv2 value set (18,750 values) for each participant and a binary search algorithm to identify any health states they consider are worse than dead. The DCE and binary search algorithm are explained in the next two sub-sections, followed by a discussion of the survey and introductory video, the tool's data-quality controls, the analysis for generating value sets, and, lastly, the international comparisons.

2.1. The discrete choice experiment

The PAPRIKA method (Hansen and Ombler, 2008) has been used extensively in a wide variety of health applications based on DCEs, including prioritising patients for access to health care, health technology assessments and prioritising pathogens for research purposes (Srikumar et al., 2020; Powers et al., 2023; Khanal et al., 2023; Sullivan and Hansen, 2017; Jakubczyk et al., 2022; Tacconelli et al., 2018).

In the present context, the PAPRIKA method involves participants being repeatedly presented with a pair of hypothetical health states described by two of the SF-6Dv2 dimensions at a time (i.e. partial profiles) and asked which state they would prefer to be in for 10 years (Fig. 1). The health states in each pair are randomly positioned with respect to being on the left- or right-hand side of the pair (screen). The 10-year duration was chosen because it was used in our earlier study to create an EQ-5D-5L value set for NZ (Sullivan et al., 2020) – to align with the "EuroQol Valuation Technology" (EQ-VT) protocol where 10 years is used in the "conventional" TTO (Oppe et al., 2014) – and 10 years is the maximum duration in the Australian, Chinese and UK SF-6Dv2 studies (Mulhern et al., 2020, 2021; Wu et al., 2021). In the introductory video discussed later below, participants are advised to think about only the two dimensions in each question and assume the other four dimensions (not included in the question) are the same.

Each time a person chooses their preferred health state in a pair, all other pairs whose rankings are logically implied are identified and eliminated and the next pair (question) is presented; thus, PAPRIKA is recognised as a type of *adaptive* DCE. For example, if health state A is preferred to state B, and B is preferred to state C, then the participant will not be asked to rank A versus C, because – by transitivity – this question has already been answered implicitly (A is preferred to C). This adaptive elimination procedure minimises 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.

Based on each participant's explicit pairwise rankings, the software uses quantitative methods based on linear programming to determine the weights, sometimes referred to as 'utilities', for the five or six levels on the six dimensions, representing their relative importance to the participant; for technical details, see Hansen and Ombler (2008). As well as creating a personal value set for each participant, their individual weights are averaged to produce a social value set.

To limit the number of questions participants are asked, only levels 1, 4 and 6 on the pain dimension and levels 1, 3 and 5 on the five other dimensions are included in the health states included in the DCE questions. The weights (utilities) for the remaining levels 2, 3 and 5 on the pain dimension and levels 2 and 4 on the other dimensions are determined using linear interpolation. The values for pain's levels 2 and 3 are the two equally-spaced points in the interval between levels 1 and 4, and the values for pain's level 5 and the other dimensions' levels 2 and 4 are the midpoints between the two values either side of the interpolated value.

Five health states considered to be unimaginable to most people were excluded from the DCE: 'not at all limited in vigorous activities' and 'very severe pain' or 'worn out all of the time'; 'role limitations none of the time' and 'very severe pain' or 'worn out all of the time'; and 'very severe pain' and 'worn out none of the time'. Computer simulations revealed that these exclusions, together with the interpolation process, reduce the mean number of questions asked by 48% from 65 to 34.

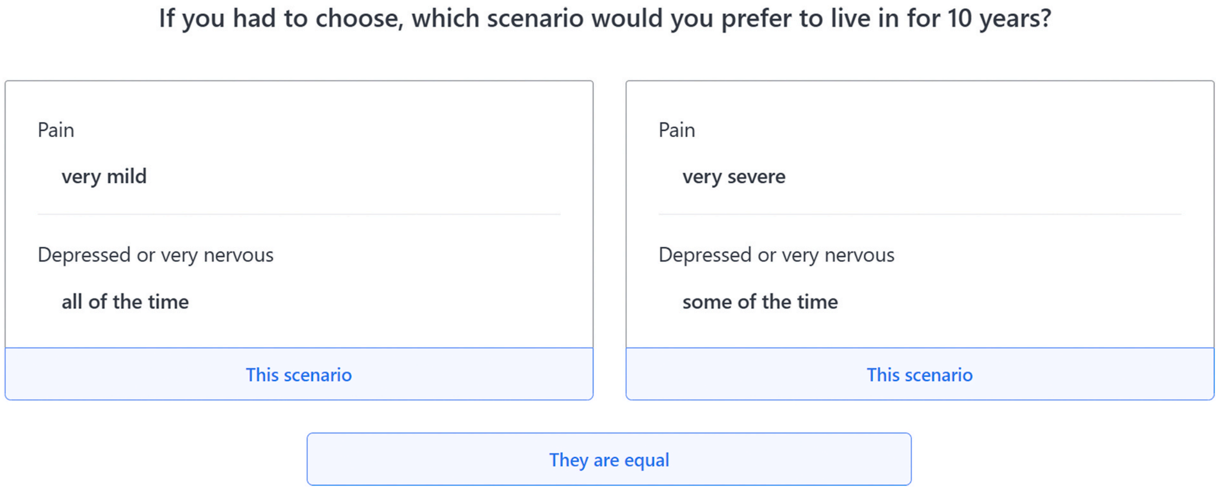

Fig. 1. Example of a DCE question.

2.2. The binary search algorithm

For each participant, having answered the DCE questions, their personal value set is immediately created by the 1000minds software, thereby generating their personal ranking of the 18,750 health states. The valuation tool then asks the participant a second series of questions implementing the binary search algorithm explained below to identify at which point 'dead' slots into their ranking of the 18,750 states (including, potentially, at the bottom of the ranking), with states below that point worse than dead.

In the following explanation, the health states referred to are represented by a six-digit code, where each digit corresponds to a level for the SF-6Dv2's six dimensions in the order they were listed earlier (physical functioning, role limitations, social functioning, pain, mental health and vitality); e.g. health state 111111 corresponds to no problems on any of the six dimensions.

The algorithm begins by presenting the participant with the worst possible health state – state 555655 (Fig. 2A) – and asking whether they think that living in this state for 10 years would be better or worse than being dead. If the person answers 'better than dead' (BTD), no more questions are asked (because dead is this person's worst imaginable health state). If instead they answer 'worse than dead' (WTD), they are presented with another, higher-ranked state from their ranking, perhaps 333233 (Fig. 2B), and again asked whether living in this state for 10 years is BTD or WTD. Having identified the range of states in which dead lies, the algorithm proceeds to repeatedly bisect their ranking of health states until it positions dead to split the ranking into states BTD and WTD respectively. For simplicity, and to ensure reasonable accuracy, instead

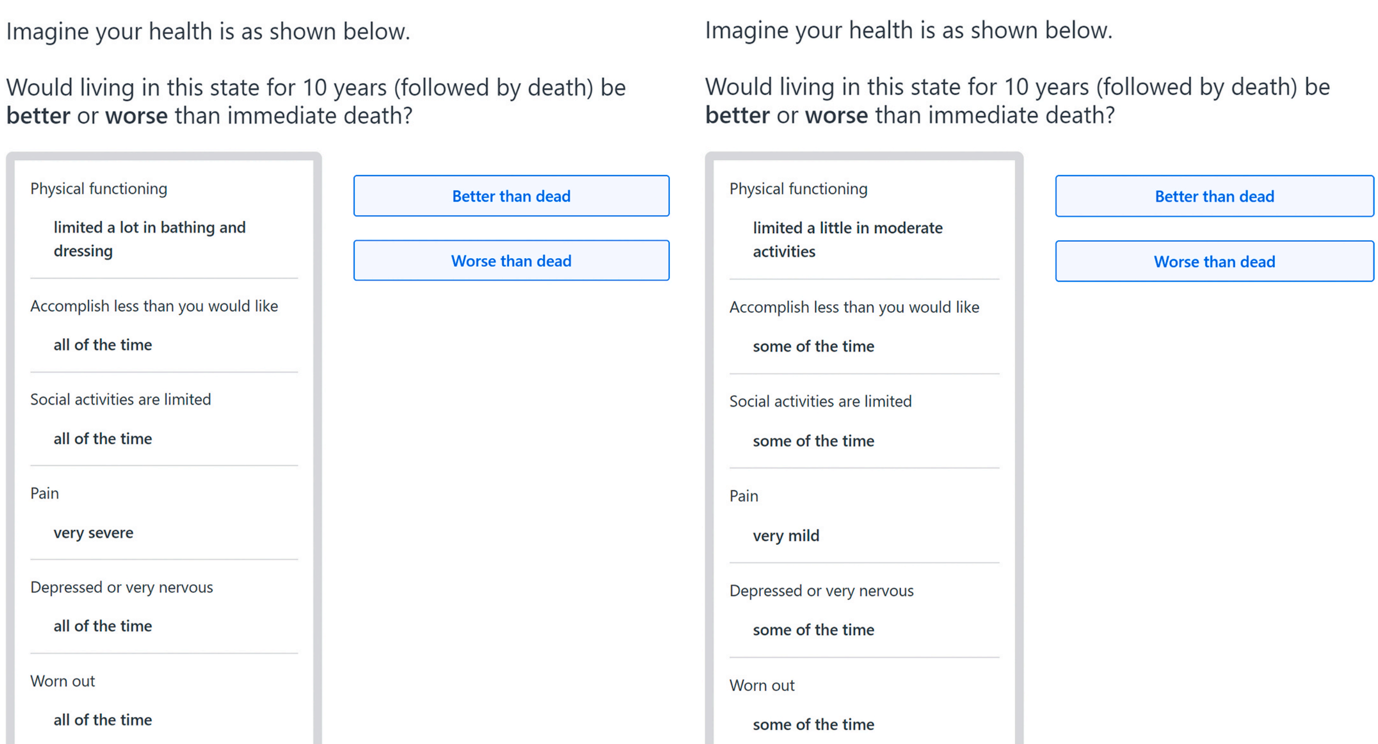

Fig. 2. Two examples of binary-search questions to identify states worse than dead.

of performing the binary search across all 18,750 health states, a maximum of 101 health states is potentially presented to each participant. These health states are determined by grouping each participant's 18,750 values by their value after rounding to two decimal places and selecting one from each group. Therefore, a maximum of eight binary-search questions are required, depending on how the person answers the questions.

Thus, at the end of the binary search, there are three possibilities for dead's location in the participant's ranking of the 18,750 health states and their values: (1) either dead is worse than 555655 (extreme problems on all dimensions), and so dead and 555655 are both valued at 0; or (2) 111111 is WTD (uncommon), and so dead = 1; or (3) dead is spanned by two adjacent states in the person's ranking (most often) – one BTD and the other WTD – and so dead's value (before rescaling, as discussed later) is the average of these two states' values.

2.3. The survey

The survey comprises six steps for participants to complete after indicating their consent by clicking the survey link.

1) Watch a short video (2.5 min) explaining the SF-6Dv2's background and how to complete the survey;
2) Rate their current health status on the SF-6Dv2 questionnaire;
3) Answer the DCE questions;
4) Answer the binary-search questions to identify any health states worse than dead;
5) Provide their socio-demographic and background information; and
6) Answer several questions about how they found the tool and the survey overall.

The survey was pilot-tested with a small sample (n = 15) of participants who were asked to 'think aloud' (Fonteyn et al., 1993) while completing the survey in the presence of two of the authors (TS and GM). Minor adjustments were made to the final survey.

2.4. The introductory video

A short video (2.5 min) explaining the SF-6Dv2's background and how to complete the survey was created (McCarty et al., 2022). During the survey's soft launch it became clear that many participants were not watching the video before starting the valuation tasks. Various options were tested such as making it mandatory to watch the video (at least to play it), whereby participants were screened out if they skipped it – i.e. they could not proceed to the survey – and clarifying the instructions: e.g. "Please start by watching this video: You need to watch the whole video before you can start the survey." The usefulness of the video was evaluated by randomly including it for some participants and not for others.

2.5. Participant recruitment

Two research companies, Dynata and CINT, were engaged to recruit a representative sample of the NZ adult (≥18 years) population with respect to age, gender, ethnicity and geographic location. In anticipating the possibility that some people might poorly understand or engage with the valuation tasks, sampling was conducted iteratively to ensure the sample remained representative of the NZ population. Given the considerable inequities in health status and outcomes for Māori (NZ's indigenous people) and to strengthen responsiveness to Māori in health research (Reid et al., 2017), purposive sampling (Patton, 2002) ensured a minimum of 20% Māori in the final sample. In line with other SF-6Dv2 valuation studies, the target was for approximately 3000 'high-quality' responses to create the social value set. Ethics approval was obtained from the University of Otago Human Ethics Committee (D17/297).

2.6. Data-quality controls

The tool has a consistency test and several other checks to identify and exclude 'low-quality' responses that are indicative of participants' failure to understand or sincerely engage with the valuation tasks.

The consistency test involves participants, at the end of the DCE and unbeknown to them, being re-presented with two questions answered earlier. Failure to answer each question identically resulted in the participant's responses being excluded from the 'high-quality' sub-sample for creating the social value set.

Participants were also excluded if they: (1) answered the DCE questions unreasonably quickly, with a median time of less than 4 s per answer or less than 2 s for any answer; (2) answered all questions in the same way by choosing the left, right or 'equal' option for every question; or (3) valued the 'best' health state (111111) worse than or equal to dead, which would result in dead being valued as 1 so that rescaling the participant's values relative to 111111 = 1 and dead = 0 for their value set is impossible.

Finally, participants were also excluded if, in response to several open-ended questions at the end of the survey, they indicated that their valuation-task responses were invalid.

2.7. Generating value sets

To produce a personal value set for each participant, i.e. 18,750 values ranging from 111111 = 1 (perfect health) to dead = 0 with values for health states considered worse than dead < 0, each participant's DCE preference weights were rescaled using this formula: $\frac{v-z}{1-z}$, where v is the value of each health state (obtained by summing the preference weights for each state's combination of levels on the six dimensions) and z is the participant's value for dead from the binary search.

Disutility coefficients (i.e. decrements in full health by dimension level) were also calculated for each participant using this formula: $\frac{w-w1}{1-z}$, where w is the preference weight being considered (i.e. by dimension level), w_1 is the highest level of that dimension and z is the participant's value for dead.

To generate the social value set, the personal disutility coefficients were averaged across all participants to produce social disutility coefficients which were then summed across the dimensions for each health state and added to one. This calculation is equivalent to generating a personal value set for each participant and then finding the average value for each health state.

2.8. International comparisons

The NZ SF-6Dv2 social value set created in this study was compared with the social value sets for Australia (Mulhern et al., 2021), China (Wu et al., 2021) and UK (Mulhern et al., 2020). The comparison was with

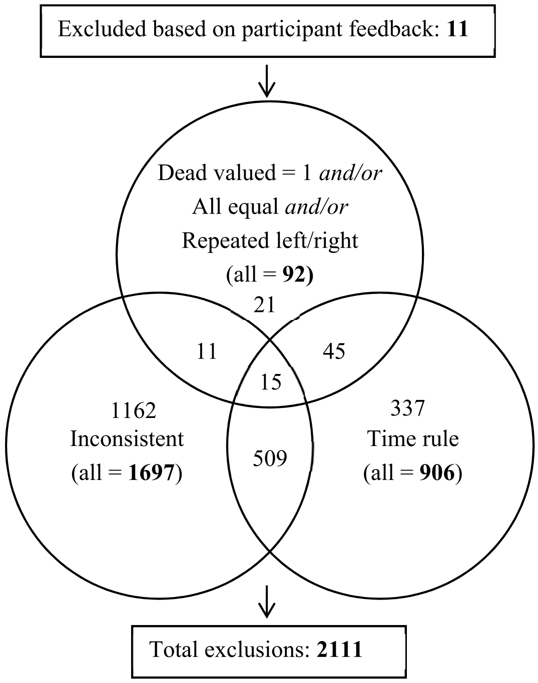

Fig. 3. Exclusions to obtain a 'high quality' sub-sample (n = 2985).

respect to rankings of the SF-6Dv2 dimensions, values of the worst possible health state and number of states worse than dead, and Spearman ρ and Pearson r correlation coefficients for pairs of value sets. The disutility coefficients of the levels for each dimension were also charted.

3. Results

The survey was distributed in June–July 2022 and completed by 5096 people (0.1% of NZ's population of approximately 5.1 million). Of these 5096 responses, 2111 were excluded for the following reasons, also summarised in Fig. 3, leaving a 'high-quality' sub-sample of 2985.

Eleven responses were immediately excluded based on participants' feedback that indicated their valuation responses were invalid: e.g. due to computer problems. As shown in the Venn diagram (Fig. 3), some responses were excluded based on just one exclusion criterion: i.e. 1162 people did not answer the two repeated DCE questions the same way; 337 completed the DCE implausibly fast; and 21 either valued dead at 1 or answered every question the same way. However, many responses were excluded based on multiple exclusion criteria: e.g. of the 1697 people who answered inconsistently, 509 did not meet the time threshold, 11 either valued dead at 1 or answered all questions the same way, and 15 were excluded based on all three criteria.

3.1. The participants

The socio-demographic characteristics of the sub-sample participants are reported in Table 1, together with NZ population census data (where available) for comparison purposes. The sub-sample is generally representative of the NZ population; however, people aged 18–24 years, with income ≤ $20,000 and of Pacific ethnicity are under-represented, whereas people with a University degree or equivalent are over-

Table 1
Socio-demographic characteristics of sub-sample (n = 2985) and NZ census data.

Characteristic	Participants n	Participants %	NZ population[a] %
Gender			
Male	1637	54.8	48.8
Female	1331	44.6	51.2
Gender diverse	17	0.5	unrecorded
Age (years)			
18-24	164	5.5	12.2
25-34	494	16.5	18.4
35-44	502	16.8	16.3
45-54	539	18.1	17.5
55-64	496	16.6	15.7
65-74	548	18.4	11.5
75-84	229	7.7	6.1
≥85	13	0.4	2.3
Ethnicity[b]			
NZ European	2308	77.3	71.1
Māori	763	25.6	13.5
Pacific	91	3.0	6.4
Asian	231	7.7	15.0
MELAA (Middle Eastern/Latin American/African)	23	0.8	1.4
Others	175	5.9	1.2
Education level			
No formal qualifications	216	7.2	18.2
Secondary school qualifications	772	25.9	38.3
Other post-secondary school qualifications	801	26.8	18.7
University degree or equivalent	1196	40.1	24.8
Employment status			
Full-time work (≥ 30 h per week)	1243	41.6	50.1[c]
Part-time work (< 30 h per week)	350	11.7	14.7
Not in paid work (including people on a benefit)	284	9.5	
Student	66	2.2	
Homemaker	198	6.6	
Retired	633	21.2	
Self-employed	194	6.5	
Others (including self-employed)	17	0.6	
Individual income			
≤ $20,000	338	11.3	34.5
$20,001-$30,000	392	13.1	13.7
$30,001-$50,000	564	18.9	20.2
$50,001-$70,000	542	18.2	14.4
$70,001-$100,000	469	15.7	9.6
≥ $100,001	390	13.1	7.6
Prefer not to say	290	9.7	
Region			
Northern	941	31.5	37.3
Midland	586	19.6	19.9
Central	700	23.5	19.4
Southern	755	25.3	23.5
Outside NZ	3	0.1	0.0
Long-term disability (lasting ≥ 6 months)			
No	2150	72.0	76.0[d]
Yes	835	28.0	24.0
Chronic disease			
No	1197	40.1	
Yes	1788	59.9	

Notes.
[a] NZ 2018 Census population statistics include people ≥18 years for the age and gender categories and people ≥15 years for all other categories (StatsNZ, 2018).
[b] Sums to > 100% as some people identify with multiple ethnicities.
[c] 2022 June quarter employment and unemployment rates were 68.5% and 3.3% (StatsNZ, 2022).
[d] 2013 New Zealand Disability Survey (StatsNZ, 2014).

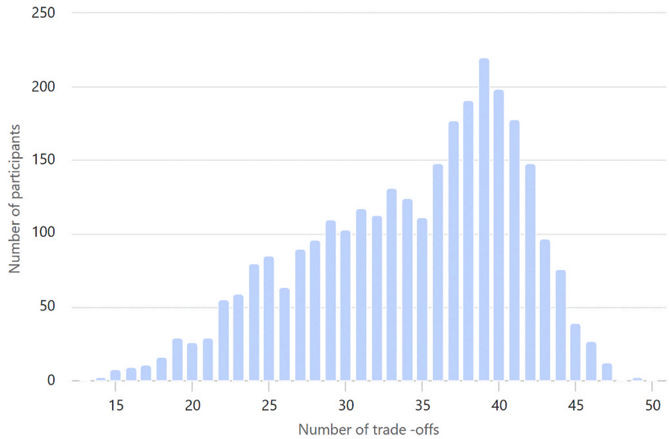

Fig. 4. Number of DCE questions answered.

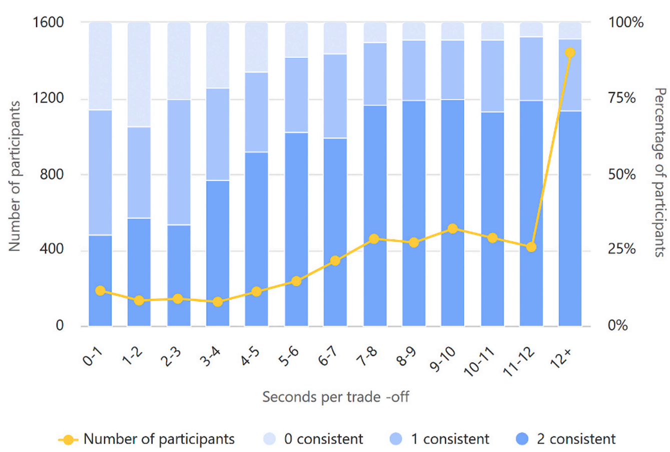

Fig. 5. Median time per DCE question answered.

Table 2
Participant feedback about the survey.

Feedback	Sub-sample (n = 2985) No.	%
Understanding instructions/survey design		
Very easy/Easy	1583	53.0
Neutral	912	30.6
Very difficult/Difficult	490	16.4
Choosing between two health states		
Very easy/Easy	786	26.3
Neutral	921	30.9
Very difficult/Difficult	1278	42.8
Choosing between living in a health state or dead		
Very easy/Easy	867	29.0
Neutral	694	23.2
Very difficult/Difficult	1424	47.7
Ranking of the SF-6Dv2 dimensions		
As expected (yes)	2669	89.4
Not as expected (all other responses)	316	10.6

Table 3
Mean DCE weights and social disutility coefficients (n = 2985).

Dimensions	Mean DCE weights	Social disutility coefficients
Physical functioning		
Not at all limited in vigorous activities	**0.171**	0
Limited a little in vigorous activities	0.147	−0.037
Limited a little in moderate activities	0.122	−0.075
Limited a lot in moderate activities	0.061	−0.173
Limited a lot in bathing and dressing	0	−0.271
Role limitations		
None of the time	**0.115**	0
A little of the time	0.097	−0.028
Some of the time	0.079	−0.055
Most of the time	0.040	−0.118
All of the time	0	−0.182
Social functioning		
None of the time	**0.105**	0
A little of the time	0.091	−0.022
Some of the time	0.076	−0.045
Most of the time	0.038	−0.106
All of the time	0	−0.167
Pain		
No	**0.241**	0
Very mild	0.198	−0.066
Mild	0.164	−0.119
Moderate	0.129	−0.172
Severe	0.065	−0.274
Very severe	0	−0.376
Mental health		
None of the time	**0.218**	0
A little of the time	0.179	−0.060
Some of the time	0.139	−0.122
Most of the time	0.069	−0.231
All of the time	0	−0.340
Vitality		
Worn out none of the time	**0.149**	0
Worn out a little of the time	0.128	−0.034
Worn out some of the time	0.106	−0.067
Worn out most of the time	0.053	−0.151
Worn out all of the time	0	−0.236

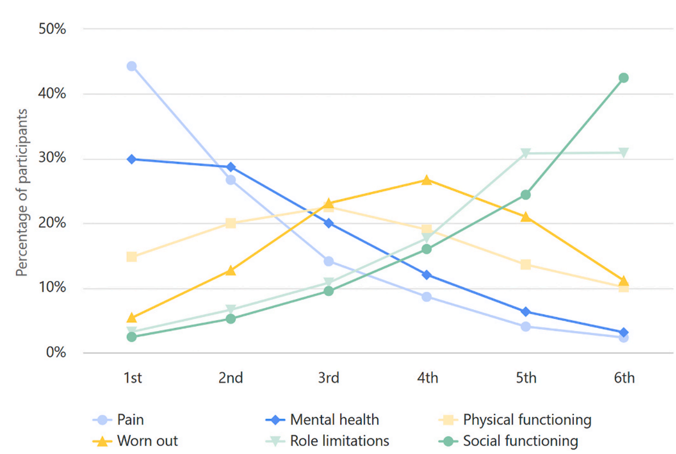

Fig. 6. Percentage of participants ranking dimensions from 1st to 6th place.

represented. Just over a quarter of the sub-sample identified as Māori.

3.2. The survey

Participants answered a mean of 34.3 (SD 6.9) DCE questions each (Fig. 4), with a median time of 10.4 s per question (Fig. 5).

The relationship between the median number of seconds per question and the number of repeated questions answered consistently is shown in Fig. 5. In general, the more time people spent per question, the more consistent they are.

As reported in Table 2, most participants had no difficulty understanding the survey instructions or design (83.6%), choosing between health states in the DCE (57.2%) and choosing between various health states or dead (52.5%). When shown their ranking of the six SF-6Dv2 dimensions, as determined by their answers to the DCE questions, most participants (89.4%) said the ranking was as they expected.

Table 4
Rankings of the SF-6Dv2 dimensions for the value sets of NZ and other countries.

	NZ	Australia	China	UK
1st	Pain	Pain	Pain	Pain
2nd	Mental	Mental	Physical	Mental
3rd	Physical	Physical	Mental	Physical
4th	Vitality	Social	Vitality	Social
5th	Role	Role	Social	Vitality
6th	Social	Vitality	Role	Role

Table 5
Value of worst possible health state and number of states worse than dead for NZ and other countries.

	Valuation method	Value of **555655**	States WTD[a]	%
NZ	DCE	−0.572	1804	9.6
Australia	DCE_{TTO}	−0.685	3679	19.6
China	TTO	−0.277	924	4.9
UK	DCE_{TTO}	−0.574	2826	15.1

Note: [a]WTD = worse than dead.

3.3. The introductory video

Participants who voluntarily watched at least 90% of the video were more than twice as likely to pass the data-quality controls. Therefore, for the full launch, the video was included, and participants who did not watch at least 90% of the video were screened out so that they could not begin the survey.

Participants' comments in the survey's (optional) free-text section included compliments about the video's usefulness: e.g. "The video at the start was really good. So when the questions came up, I knew exactly what I had to do."; "I think more surveys should contain a short video at the start to explain the survey and the type of questions they will get."; and "The video was great, it was warm and welcoming (especially loved the reo Māori throughout)."

3.4. Creating an SF-6Dv2 value set for New Zealand

The mean weights are reported in the second column of Table 3. These weights are normalised so that perfect health (no problems on any dimension) is equal to 1; i.e. the bolded values on the first level of each dimension sum to 1. The dimension with the highest weight – i.e. that people most want to avoid – is pain, with a mean weight of 0.241, followed by mental health (0.218), physical functioning (0.171), vitality (0.149), role limitations (0.115) and – least important – social functioning (0.105).

The social disutility coefficients are in the third column of Table 3. The values for the social value set can be calculated by adding these coefficients to 1: e.g. the value for state 333333 = 1 + (−0.075) + (−0.055) + (−0.045) + (−0.119) + (−0.122) + (−0.067) = 0.517. Analogous calculations for all 18,750 health states result in the SF-6Dv2 social value set for NZ. An Excel workbook for generating these health state values for NZ – also for Australia, China and UK – is available on request from the authors.

Fig. 6 shows the overall ranking of the six dimensions, including the percentage of participants ranking each dimension from 1st to 6th place. For example, 44.3% of participants ranked pain 1st and 2.3% ranked it 6th, whereas, in contrast, just 2.4% ranked social functioning 1st and 42.5% ranked it 6th.

Table 6
Correlation coefficients for the SF-6Dv2 value sets of NZ and other countries.

	NZ	Australia	China	UK
NZ	-	0.901 (0.08)	0.858 (0.36)	0.905 (0.07)
Australia	0.904 (0.10)	-	0.894 (0.11)	0.995 (0.00)
China	08.56 (0.46)	0.890 (0.19)	-	0.888 (0.15)
UK	0.907 (0.09)	0.995 0.00	0.885 (0.22)	-

Note: Spearman's ρ correlation coefficients are in the table's upper triangular and Pearson's r correlation coefficients in the lower triangular (shaded), with p-values for statistical significance in parentheses.

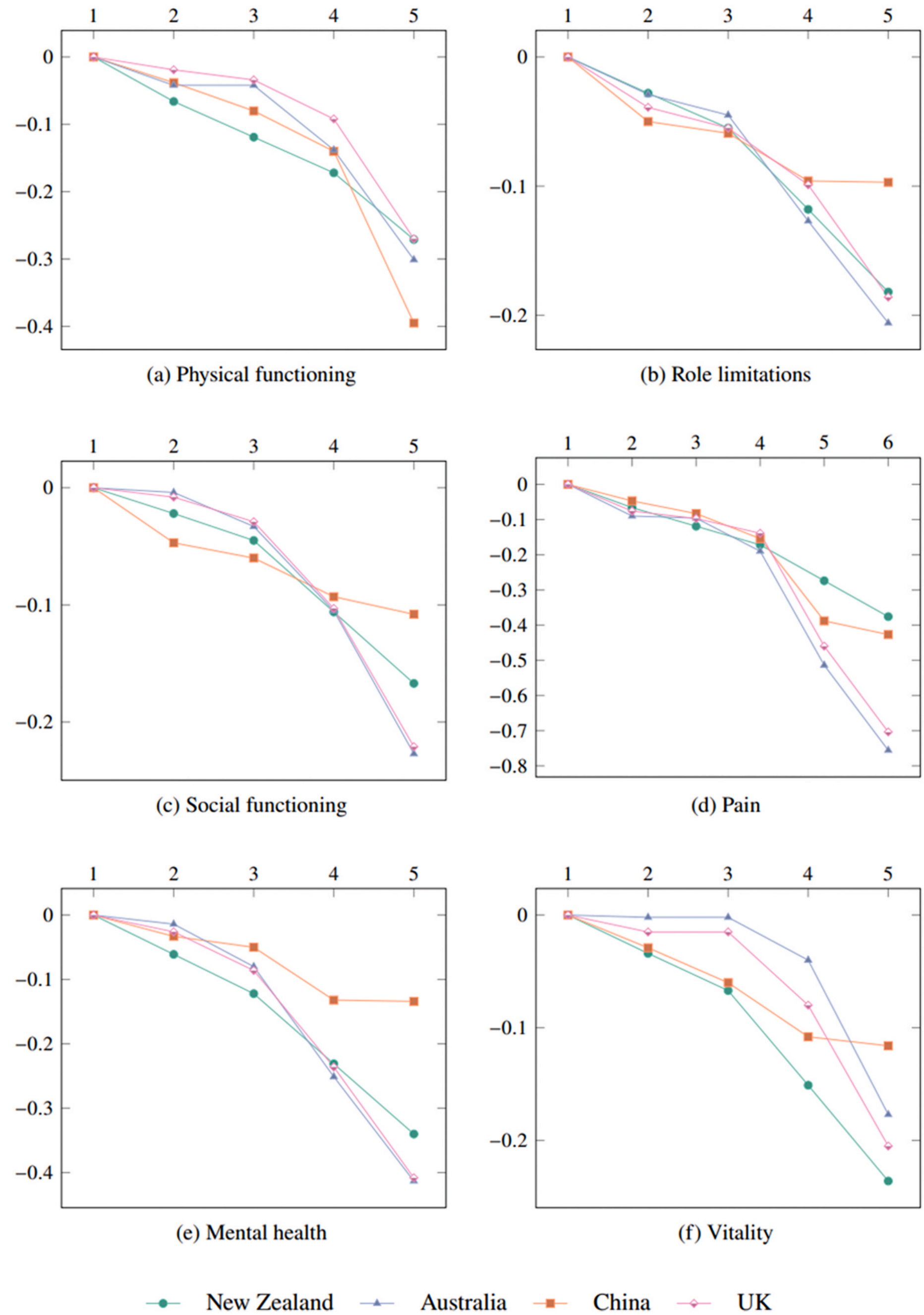

Fig. 7. Disutility coefficients for the levels of the SF-6Dv2's dimensions for NZ and other countries.

3.5. *Comparisons with other SF-6Dv2 value sets*

Table 4 reports the rankings of the SF-6Dv2 dimensions for the NZ, Australian (Mulhern et al., 2021), Chinese (Wu et al., 2021) and UK (Mulhern et al., 2020) value sets. The values of the worst possible health state (555655) and the number of states worse than dead (WTD) are in Table 5. Table 6 presents Spearman's ρ and Pearson's r correlation coefficients of the value sets of pairs of countries, where ρ correlation coefficients are in the table's upper triangular and r in the lower triangular (shaded), with p-values for statistical significances in parentheses (Ombler et al., 2018; 1000minds HRQoL correlation significance calculator (1000minds, 2023)). There are strong positive relationships between the value sets for NZ and UK ($\rho = 0.905$ $p = 0.07$; $r = 0.907$ $p = 0.09$), NZ and Australia ($\rho = 0.901$ $p = 0.08$; $r = 0.904$ $p = 0.10$) and Australia and UK ($\rho = 0.995$ $p < 0.001$; $r = 0.995$ $p < 0.001$).

Fig. 7 charts the disutility coefficients for each dimension's levels for NZ and the three other countries. For all dimensions for all countries, their functional 'shape' in terms of marginal effects of level changes exhibits increasing disutility as the levels progressively worsen from 'no problems' (level 1) to 'extreme problems' (level 5 or 6).

4. Discussion

The SF-6Dv2 social value set created in this study is the first for the SF-6Dv2 for the NZ population and only the fourth SF-6Dv2 social value set worldwide (Mulhern et al., 2020, 2021; Wu et al., 2021). This is the second time a valuation tool implementing the PAPRIKA method (Hansen and Ombler, 2008) has been used to create a social value set, building on the first implementation that resulted in an EQ-5D-5L value set for NZ (Sullivan et al., 2020).

Unlike the valuation tool for the EQ-5D-5L, and in addition to other enhancements, the SF-6Dv2 tool was accompanied by a short introductory video. As well as creating an SF-6Dv2 social value set for NZ, an objective of this latest application of the tool was to investigate the video's effect on participants' understanding and engagement and hence the quality of their data. According to feedback from the pilot-testing and the survey's free-text responses, participants found the video useful in explaining what the survey was about and how to complete it. Also, the inclusion of te reo Māori (Māori language) was appreciated and is likely to have contributed to the higher proportion of Māori completing the survey.

An important finding is that the video served as an additional data-quality control (i.e. in addition to the tool's built-in controls). People who watched at least 90% of the video were more than twice as likely to pass the tool's data-quality controls (e.g. to answer the repeated questions consistently), indicating they understood and sincerely engaged with the survey. This finding has obvious implications for the design of future valuation surveys: including an introductory video that engages participants and clearly explains what the survey entails is likely to increase the number of high-quality responses.

Compared to the EQ-5D-5L, the SF-6Dv2 has one more dimension with six levels and is therefore capable of representing six times as many possible health states: 18,750 compared to just 3125. Participants in the SF-6Dv2 survey were required to answer a mean of 34.3 DCE questions compared to 20 for the EQ-5D-5L survey. Most people (83.6%) reported no difficulty understanding the survey format or instructions and provided positive general feedback. However, 42.8% said they found the DCE questions intrinsically difficult to answer, which is unsurprising given the cognitively difficult and time-consuming nature of trade-off tasks and is consistent with other studies (Sullivan et al., 2020; Xie et al., 2020). Similarly, 52.5% of participants reported having some difficulty answering the questions about health states worse than dead; nonetheless, most people were able to do so.

An important feature of the valuation tool is its extensive data-quality controls identifying and excluding 'low-quality' responses indicative of participants' failure to understand or sincerely engage with the valuation tasks. The 41% exclusion rate (2111 of 5096 responses) in the SF-6Dv2 survey and the 52% rate (2644 of 5112) in the NZ EQ-5D-5L survey (Sullivan et al., 2020) are higher than is typical in other online surveys. However, as for other studies whose general-population samples are recruited using commercial research companies, these apparently high exclusion rates (reflecting relatively low participant understanding or engagement) are not unexpected and reflect the stringent data-quality controls imposed. High data-quality standards are important because they ensure policy-makers and researchers can have confidence in the validity and reliability of the resulting social value set – which is especially important given one likely application is supporting decision-making about the allocation of health care resources.

Overall, these results indicate that the valuation tool is feasible and acceptable to participants in the present study and that the tool used in the earlier EQ-5D-5L study was successfully adapted and extended for the larger SF-6Dv2.

4.1. Comparison with other DCE-based methods used by other valuation studies

Other DCE-based methods for creating SF-6Dv2 and EQ-5D-5L value sets are based on *full*-profile DCEs. Full-profile DCEs involve the hypothetical health states (i.e. 'profiles') that participants are asked to choose between being defined on all the system's dimensions together at once: e.g. all six SF-6Dv2 dimensions for the DCE_{TTO} (Mulhern et al., 2020, 2021; Wu et al., 2021) and all five EQ-5D-5L dimensions for the 'EuroQoL Valuation Technology' (EQ-VT) (Oppe et al., 2014).

In contrast, the PAPRIKA method used in this study is based on *partial* profiles, whereby participants choose between pairs of hypothetical health states defined on just two SF-6Dv2 dimensions at a time (and involving a trade-off). With just two partial health state profiles to choose between – instead of two (or more) full profiles with all dimensions, and sometimes including an immediate death or full health option (e.g. Viney et al., 2014) – PAPRIKA's questions are the simplest possible questions involving trade-offs.

In general, partial profiles are easier for people to think about than full profiles because they involve fewer dimensions, and DCEs based on partial profiles have been shown to reflect people's true preferences more accurately than full-profile DCEs (Chrzan, 2010; Meyerhoff and Oehlmann, 2023). The simplicity of PAPRIKA's DCE questions ensures they are relatively easy and quick for people to think about and to have confidence in their answers. Although full profiles are required for the binary search task, participants only consider one profile at a time.

Another benefit of the PAPRIKA method is that a personal value set is created for each participant – in contrast to most other DCE-based methods for creating social value sets which generate *aggregate* data from the group of participants (which can be modelled assuming both homogenous and heterogenous preferences). PAPRIKA enables the heterogeneity of health state preferences to be explored at the individual level as well as in clusters or by sub-group (e.g. by ethnicity, age or health status). Custom value sets can also be created to inform decision-making; e.g. the health state preferences of male participants could be used to evaluate prostate treatments and of 18–24 year olds to design mental health services for young people, etc.

4.2. Meaning of the study: possible mechanisms and implications

An important practical outcome of this research is that policy-makers and researchers have access to a second contemporary HRQoL value set for NZ. The SF-6Dv2 value set provides an alternative to the EQ-5D value set, which has been the only system used by NZ's Pharmaceutical Management Agency (Pharmac, 2015) for generating estimates of quality-adjusted life years (QALYs) for cost-utility analysis. As a stand-alone questionnaire, the SF-6Dv2 can be used in trials and studies and its values applied to any administration of the SF-36v2 in both existing and future data. The value set can also be used for PROMs by NZ health agencies, including the Health Quality and Safety Commission, Te Whatu Ora Health New Zealand and the Accident Compensation Corporation (ACC).

A particular strength of the NZ SF-6Dv2 value set is that it is based on purposive targeting of a representative sample of adult New Zealanders in terms of the main socio-demographic characteristics (gender, age, ethnicity and region). A higher proportion of Māori – who experience inequities in health status and outcomes (Gurney et al., 2020) – was purposefully recruited (Patton, 2002). Over a quarter of the 'high-quality' sub-sample ($n = 2985$) used to create the social value set identified as Māori (13.5% of NZ people $\geq$ 15 years identified as Māori in the 2018 NZ Census), which is a considerably higher proportion than for the earlier NZ social value sets: 15.8% for the EQ-5D-5L (Sullivan et al., 2020) and 7.6% for the EQ-5D-3L (Devlin et al., 2003). Given the significant and pervasive health inequities faced by Māori, it is very important to understand and include Māori preferences in health care policy and decision-making (Sullivan et al., 2023). The results from this study – specifically, the 763 personal value sets for Māori and the social value set – can be used more effectively to target resources to improve Māori health outcomes.

Though the sub-sample used to create the social value set is generally representative of the NZ population overall, there were proportionately more people with higher education and fewer in the lowest income bracket, young people and of Pacific Island ethnicity. However, the benefit of having personal value sets is that the preferences of population sub-groups can be compared to determine whether there are significant differences between them, as will be explored in future research.

4.3. Comparison with other countries' value sets

Comparisons between the NZ, Australian, Chinese and UK SF-6Dv2 value sets reveal similar rankings of the six dimensions. On the other hand, the greatest contrasts are for the social functioning and vitality dimensions: social functioning is the least important dimension for NZ whereas it is the fourth most important for Australia and UK; and vitality is the least important dimension for Australia but the fourth most important for NZ and China.

Notwithstanding these ranking differences, it is the magnitude of the differences in disutility coefficients that directly affects QALY calculations. For example, although pain is the highest-ranked dimension for all countries, NZ and China place much less weight on pain than Australia and UK. The implication is that QALY gains may be different for the same health state in different countries. Similarly, an intervention impacting social functioning could be evaluated more favourably in Australia and UK than in NZ. However, the impact of the differences – in overall rankings and magnitude – is nuanced due to the shape of the value functions, as shown in Fig. 7; e.g. the difference in preference weights between levels 2 and 3 might be smaller than the difference between levels 5 and 6.

The percentage of health states worse than dead is 9.6% for NZ, which is higher than for China (4.9%) and lower than for Australia (19.6%) and UK (15.1%). The value of the worst possible health state is −0.572, very close to UK (−0.574), higher than Australia (−0.685) and lower than China (−0.277).

There is a positive relationship between the NZ value set and the UK and Australian value sets. However, this relationship is not as significant as the one between the Australian and UK value sets – which may be due, at least in part, to Australia and UK using the same protocol based on $\mathrm{DCE_{TTO}}$ to create their value sets.

The implications of differences in preference weights and values of the worst possible health state between the different countries could be explored further.

4.4. Potential limitations of the research

A 10-year duration for the health states was used in this study as this was the longest duration used in other SF-6Dv2 studies and aligns with conventional TTO for states better than dead (e.g. in the EQ-VT protocol). It is possible, however, that the duration used could lead to difficulties for DCE participants in imagining different amounts of time, and also their choices, depending on interactions between the health states' severity and the time spent in them. These effects have been shown to impact preference weights, with the value of moderate and severe states decreasing as duration increases (Stalmeier et al., 2007). These possibilities could be explored in the context of the PAPRIKA method by conducting multiple DCEs with durations of 1, 4, 7 and 10 years, for example.

Another potential limitation concerns how the weights for the dimensions' intermediate levels – levels 2, 3 and 5 on the pain dimension and 2 and 4 on the five other dimensions – were determined. To restrict the number of DCE questions, so as to not over-burden survey participants, these weights were interpolated, with only three levels from each dimension included in the questions. In contrast, the usual approach in other DCEs is to include all levels in the questions. Such an approach, which for the SF-6Dv2, would require suitably motivated participants to answer approximately 75 questions (compared to 34 in the present study), was tested for the EQ-5D-5L in a small pilot study and is recommended if time and resources permit (O'Hara, 2020). However, this approach has the obvious disadvantage of risking participants being over-burdened, thereby resulting in lower quality answers and lower completion rates, and hence potentially less representative samples.

A related potential limitation is the use of *linear* interpolation. In the absence of prior knowledge about the shape of the dimensions' value functions, linear interpolation is likely to result in the smallest possible error on average as a result of the interpolated value lying in the middle of the two values from which it is derived. Reassuringly, if the shape of the value functions for NZ (using linear interpolation) mirror other countries, then NZ's intermediate levels appear reasonably consistent with those of other countries – albeit, two exceptions might be a slight under-weighting of 'severe' on the pain dimension, which could arguably be a little closer to 'very severe', and a slight over-weighting of 'most of the time' on the vitality dimension, which could perhaps be a little closer to 'some of the time'. Nonetheless, linear interpolation may still be problematic. Other interpolation methods, such as Bézier (Farin et al., 2002), are available but have their own limitations – and, hence why linear interpolation was chosen for the present study. Further work to determine whether there is a better method would be worthwhile.

An alternative approach – something that could be investigated in future studies – is to run multiple 'partial' DCEs based on all levels being included but for subsets of the dimensions for each DCE (e.g. four dimensions per DCE), thereby requiring fewer DCE questions per participant. These partial DCEs could be used to establish the shape of the value functions for each dimension (addressing the limitation above), which could then be used to inform interpolation when running a full DCE. The advantage of this approach is that the binary search for 'dead' can still be based on the participant's personal value set, albeit with socially-informed interpolation.

Alternatively, the partial DCEs can be combined to produce the social value set. This technique is being tested in a valuation feasibility study of the EQ-HWB system (Brazier et al., 2022) in which the dimensions are divided between two DCEs and a 'common' dimension is shared across the two DCEs. This common dimension is used to, in effect, merge the results of the two DCEs into a single social value set comprising all dimensions (Burge et al., 2010; Potoglou et al., 2011). Unfortunately, however, a separate binary-search survey would be required to identify states worse than dead, and because it is based on the social value set rather than personal value sets, the binary search would be less accurate.

Another potential limitation is selection bias. A sample may be representative in terms of population demographics, or could be adjusted accordingly, but as for any online survey, people with poor access to computer technology or who are averse to, unconfident or less engaged when completing online surveys, are likely to be under-represented. Participants' feedback indicates that answering 34 DCE questions was achievable. However, for some participants, the number of questions may have affected their level of engagement. In mitigation, the survey was designed to be as user-friendly as possible and accessible via the usual range of online devices that people use, including smartphones and other portable devices.

4.5. Future opportunities

The opportunity to access individual-level data could be fully leveraged to enable researchers and policy-makers to 'dial up' tailored value sets for specific socio-demographic groups – e.g. a value set for 'Māori women between the ages of 35 and 55 with chronic illnesses'. The overall population data set could be continually updated as participants are recruited to the sample – safeguarded by the tool's data-quality controls – so that it remains as up-to-date, representative and inclusive as possible (e.g. as population characteristics evolve).

5. Conclusion

An SF-6Dv2 social value set is now available for NZ. The ease with which personal and social value sets for the SF-6Dv2 were created in the study reported on here, combined with a similar finding in an earlier study for the EQ-5D-5L (Sullivan et al., 2020), confirms the feasibility and acceptability of the valuation tool. The tool enables value sets to be relatively easily and cheaply created at both the personal level – e.g. suitable for personalised medicine – and the population level to create social value sets. Having been applied twice in NZ, the tool could also be used in other countries and for other health descriptive systems.

Funding/support

This project was funded by the Health Research Council of New Zealand (HRC19/640).

Ethics statement

Ethics approval for this study was obtained from the University of Otago Human Ethics Committee (D17/297). All participants provided informed consent and their privacy maintained.

CRediT authorship contribution statement

Trudy Sullivan: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Georgia McCarty:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Franz Ombler:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Robin Turner:** Writing – review & editing, Visualization, Formal analysis, Data curation. **Brendan Mulhern:** Writing – review & editing, Validation, Conceptualization. **Paul Hansen:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Conceptualization.

Data availability

Data will be made available on request.

Acknowledgments

Thank you to Viktor Sherepa for creating the instructional video and to Petra Kelly-Voicu for help with the audio recording. Thank you also to the participants who took part in the think-aloud (pre-testing) sessions and who completed the survey.

References

Brazier, J., Usherwood, T., Harper, R., Thomas, K., 1998. Deriving a preference-based single index from the UK SF-36 Health Survey. J. Clin. Epidemiol. 51 (11), 1115–1128. https://doi.org/10.1016/S0895-4356(98)00103-6.

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. https://doi.org/10.1016/S0167-6296(01)00130-8.

Brazier, J., Roberts, J., Tsuchiya, A., Busschbach, J., 2004. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ. 13 (9), 873–884. https://doi.org/10.1002/hec.866.

Brazier, J., Ratcliffe, J., Salomon, J., Tsuchiya, A., 2007. Methods for obtaining health state values: generic preference-based measures of health and the alternatives. Measuring and valuing health benefits for economic evaluation. In: Brazier, J., Ratcliffe, J., Salomon, J., Tsuchiya, A. (Eds.), Measuring and Valuing Health Benefits for Economic Evaluation. Oxford University Press, pp. 147–206. https://doi.org/10.1093/med/9780198725923.001.0001.

1000minds, 2023. HRQoL Correlation Significance Calculator. https://www.1000minds.com/health/hrqol/significance.

Brazier, J.E., Mulhern, B.J., Bjorner, J.B., Gandek, B., Rowen, D., Alonso, J., Vilagut, G., Ware, J.E., 2020. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. Med. Care 58 (6), 557–565. https://doi.org/10.1097/MLR.0000000000001325.

Brazier, J., Peasgood, T., Mukuria, C., Marten, O., Kreimeier, S., Luo, N., et al., 2022. The EQ-HWB: overview of the development of a measure of health and wellbeing and key results. Value Health 25 (4), 482–491.

Burge, P., Netten, A., Gallo, F., 2010. Estimating the value of social care. J. Health Econ. 29 (6), 883–894.

Canadian Agency for Drugs Technologies in Health (CADTH), 2006. *Guidelines for the Economic Evaluation of Health Technologies: Canada*. Ottawa: Canadian Agency for Drugs and Technologies in Health. https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-0.

Chrzan, K., 2010. Using partial profile choice experiments to handle large numbers of attributes. Int. J. Mark. Res. 52 (6), 827–840. https://doi.org/10.2501/S1470785310201673.

Delwel, G., 2008. *Guidance for Outcomes Research for the Assessment of the Cost-Effectiveness of In-Patient Medicines* (Report No. 270: 28113706). Health Care Insurance Board: Diemen, The Netherlands.

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. https://doi.org/10.1002/hec.741.

Farin, G., Hoschek, J., Kim, M.S. (Eds.), 2002. Handbook of Computer Aided Geometric Design. Elsevier, Vancouver.

Ferreira, P.L., Ferreira, L.N., Pereira, L.N., 2008. How consistent are health utility values? Qual. Life Res. 17, 1031–1042. https://doi.org/10.1007/s11136-008-9368-8.

Fonteyn, M.E., Kuipers, B., Grobe, S.J., 1993. A description of think aloud method and protocol analysis. Qual. Health Res. 3 (4), 430–441. https://doi.org/10.1177/104973239300300403.

Gurney, J., Stanley, J., Sarfati, D., 2020. The inequity of morbidity: disparities in the prevalence of morbidity between ethnic groups in New Zealand. J. Comorbidity 10. https://doi.org/10.1177/2235042X20971168, 2235042X20971168.

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. https://doi.org/10.1002/mcda.428.

Health Information and Quality Authority (HIQA), 2020. Guidelines for the Economic Evaluation of Health Technologies in Ireland, pp. 38–41. https://www.hiqa.ie/sites/default/files/2020-09/HTA-Economic-Guidelines-2020.pdf.

Jakubczyk, M., Niewada, M., Plisko, R., Władysiuk, M., Jachimowicz, M., Pruszko, C., Katarzyna, W., Lis, J., 2022. What matters in treating non-oncological rare diseases?—eliciting experts' preferences in Poland with PAPRIKA. J. Multi-Criteria Decis. Anal. 29 (1–2), 110–121. https://doi.org/10.1002/mcda.1754.

Khanal, S., Schmidtke, K.A., Talat, U., Turner, A.M., Vlaev, I., 2023. Using multi-criteria decision analysis to describe stakeholder preferences for new quality improvement initiatives that could optimise prescribing in England. Frontiers in Health Services 3, 1155523. https://doi.org/10.21203/rs.3.rs-2308438/v1.

Longworth, L., Bryan, S., 2003. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. Health Econ. 12 (12), 1061–1067. https://doi.org/10.1002/hec.787.

McCarty, G., Sullivan, T., Ombler, F., Hansen, P., Mulhern, B., Kelly-Voicu, P., Sherepa, V., 2022. *Aotearoa New Zealand SF-6Dv2 – Health Preferences Survey – 2022* [Video]. YouTube. https://www.youtube.com/watch?v=rS1bxlE_eRk.

Meyerhoff, J., Oehlmann, M., 2023. The performance of full versus partial profile choice set designs in environmental valuation. Ecol. Econ. 204, 107665 https://doi.org/10.1016/j.ecolecon.2022.107665.

Mulhern, B., Norman, R., Street, D.J., Viney, R., 2019. One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. Pharmacoeconomics 37, 29–43. https://doi.org/10.1007/s40273-018-0714-6.

Mulhern, B.J., Bansback, N., Norman, R., Brazier, J., 2020. Valuing the SF-6Dv2 classification system in the United Kingdom using a discrete-choice experiment with duration. Med. Care 58 (6), 566–573. https://doi.org/10.1097/MLR.0000000000001324.

Mulhern, B., Norman, R., Brazier, J., 2021. Valuing SF-6Dv2 in Australia using an international protocol. Pharmacoeconomics 39 (10), 1151–1162. https://doi.org/10.1007/s40273-021-01043-4.

Ombler, F., Albert, M., Hansen, P., 2018. How significant are "high" correlations between EQ-5D value sets? Med. Decis. Making 38 (6), 635–645. https://doi.org/10.1177/0272989X18778295.

Ombler, F., Moller, J., Paton, K., Sherepa, V., Hansen, P., 2022. 1000minds [software]. https://www.1000minds.com.

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. https://doi.org/10.1016/j.jval.2014.04.002.

Osman, A.M., Wu, J., He, X., Chen, G., 2021. Eliciting SF-6Dv2 health state utilities using an anchored best-worst scaling technique. Soc. Sci. Med. 279 https://doi.org/10.1016/j.socscimed.2021.114018.

O'Hara, R., 2020. An Assessment of the 2020 EQ-5D-5L Value Set for New Zealand. University of Otago [Unpublished Master's dissertation].

Patton, M.Q., 2002. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Sage Publications, Thousand Oaks.

Pharmaceutical Management Agency (Pharmac), 2015. Prescription for pharmacoeconomic analysis, version 2.2 – methods for cost-utility analysis. Available from: https://pharmac.govt.nz/assets/pfpa-2-2.pdf.

Potoglou, D., Burge, P., Flynn, T., Netten, A., Malley, J., Forder, J., Brazier, J.E., 2011. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. Soc. Sci. Med. 72 (10), 1717–1727.

Powers, J., McGree, J.M., Grieve, D., Aseervatham, R., Ryan, S., Corry, P., 2023. Managing surgical waiting lists through dynamic priority scoring. Health Care Manag. Sci. 26 (3), 1–25. https://doi.org/10.1007/s10729-023-09648-1\.

Reid, M.J., Paine, S.J., Curtis, E., Jones, R., Anderson, A., Willing, E., Harwood, M., 2017. Achieving health equity in Aotearoa: strengthening responsiveness to Māori in health research. N. Z. Med. J. 130 (1465), 96–103.

Srikumar, G., Eglinton, T., MacCormick, A.D., 2020. Development of the general surgery prioritisation tool implemented in New Zealand in 2018. Health Pol. 124 (10), 1043–1049. https://doi.org/10.1016/j.healthpol.2020.07.018.

Stalmeier, P.F., Lamers, L.M., Busschbach, J.J., Krabbe, P.F., 2007. On the assessment of preferences for health and duration: maximal endurable time and better than dead preferences. Med. Care 45 (9), 835–841.

StatsNZ-Tatauranga Aotearoa, 2014. Disability Survey: 2013. https://www.stats.govt.nz/information-releases/disability-survey-2013.

StatsNZ-Tatauranga Aotearoa, 2018. 2018 Census. https://www.stats.govt.nz/2018-census/.

StatsNZ-Tatauranga Aotearoa, 2022. Labour Market Statistics: June 2022 Quarter. https://www.stats.govt.nz/information-releases/labour-market-statistics-june-2022-quarter/.

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. https://doi.org/10.1016/j.jval.2016.12.008.

Sullivan, T., Hansen, P., Ombler, F., Derrett, S., Devlin, N., 2020. A new tool for creating personal and social EQ-5D-5L value sets, including valuing 'dead'. Soc. Sci. Med. 246, 112707 https://doi.org/10.1016/j.socscimed.2019.112707.

Sullivan, T., McCarty, G., Wyeth, E., Turner, R.M., Derrett, S., 2023. Describing the health-related quality of life of Māori adults in Aotearoa me Te Waipounamu (New Zealand). Qual. Life Res. 32 (7), 2117–2126. https://doi.org/10.1007/s11136-023-03399-w.

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D.L., Pulcini, C., Kahlmeter, G., Kluytmas, J., Carmeli, Y., Ouellette, M., Outterson, K., Patel, J., Cavaleri, M., Cox, E.M., Houchens, C.R., Grayson, M.L., Hansen, P., Singh, N., 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. https://doi.org/10.1016/S1473-3099(17)30753-3.

The Pharmaceutical Benefits Advisory Committee, 2015. Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). Commonwealth of Australia, Canberra, Australia. https://pbac.pbs.gov.au/.

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. https://doi.org/10.1002/hec.2953.

Wang, L., Poder, T.G., 2023. A systematic review of SF-6D health state valuation studies. J. Med. Econ. 26 (1), 584–593. https://doi.org/10.1080/13696998.2023.2195753.

Wang, H., Rowen, D.L., Brazier, J.E., Jiang, L., 2023. Discrete choice experiments in health state valuation: a systematic review of progress and new trends. Appl. Health Econ. Health Pol. 21 (3), 405–418. https://doi.org/10.1007/s40258-023-00794-9.

Ware, J.E., Sherbourne, D., 1992. The MOS 36-item short-form health survey (SF-36): I. Conceptual Framework and Item Selection. Med. Care 30 (6), 473–483.

Wu, J., Xie, S., He, X., Chen, G., Bai, G., Feng, D., Hu, M., Jiang, J., Wang, XiaohuiWu, Hongyan, W., Wu, Q., Brazier, J.B., 2021. Valuation of SF-6Dv2 health states in China using time trade-off and discrete-choice experiment with a duration dimension. Pharmacoeconomics 39, 521–535. https://doi.org/10.1007/s40273-020-00997-1.

Xie, S., Wu, J., He, X., Chen, G., Brazier, J.E., 2020. Do discrete choice experiments approaches perform better than time trade-off in eliciting health state utilities? Evidence from SF-6Dv2 in China. Value Health 23 (10), 1391–1399. https://doi.org/10.1016/j.jval.2020.06.010.