Article:

Kuluratna, S., Chen, G., Norman, R. et al. (8 more authors) (2022) Developing an Australian utility value set for MacNew-7D health states. Quality of Life Research. ISSN 0962-9343

https://doi.org/10.1007/s11136-022-03325-6

Developing an Australian utility value set for MacNew-7D health states

Sanjeewa Kularatna[1], Gang Chen[2], Richard Norman[3], Clara Mukuria[4], Donna Rowen[4], Sameera Senanayake[1], Ruvini Hettiarachchi[1], Brendan Mulhern, Katie Fozzard[1], William Parsonage[1], Steven M McPhail[1,5]

Affiliations

1. Australian Centre for Health Services Innovation, Centre for Healthcare Transformation, School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia.
2. Centre for Health Economics, Monash Business School, Monash University, Melbourne, VIC 3145, Australia.
3. School of Population Health, Curtin University
4. School of Health and Related Research, University of Sheffield, UK
5. Centre for Health Economics Research and Evaluation, University of Technology, Australia
6. Digital Health and Informatics, Metro South Health, Brisbane, Qld 4102, Australia

Corresponding Author : Sanjeewa Kularatna

Corresponding author email: sanjeewa.kularatna@qut.edu.au

ORCID

Sanjeewa Kularatna: 0000-0001-5650-154X

Gang Chen: 0000-0002-8385-5965

Richard Norman**:** 0000-0002-3112-3893

Clara Mukuria

Donna Rowen

Sameera Senanayake: 0000-0002-5606-2046

Ruvini Hettiarachchi: 0000-0003-3501-2645

Brendan Mulhern 0000-0003-3656-8063

Katie Fozzard

William Parsonage 0000-0002-0223-5378

Steven McPhail 0000-0002-1463-662X

Abstract

Background

A new preference-based measure (MacNew-7D) has recently been developed to allow condition-specific data to be used to capture the quality of life in health economic evaluations in cardiology; however, a general population value set has not yet been developed. This study developed a population utility value set for the MacNew-7D heart disease-specific instrument.

Methods

The discrete choice experiments (DCE) technique was chosen as the preference elicitation method. The DCE asked respondents to compare two options and to state their preferences. The survey was conducted using an online panel of respondents, with quota sampling using age groups, sex and jurisdictions to achieve representativeness of the Australian population. The total design consisted of 200 choice sets, of which each respondent answered eight. Additionally, each respondent answered two quality control choice sets. The best-fitting models were selected on the basis of consistency, parsimony, and goodness of fit.

Results

In total, 1903 respondents were included in the analyses. The MacNew-7D utility value set ranged from -0.4456 to 1.000 for health states defined by the classification system. The best-fitting model retained all levels for five dimensions and collapsed one adjacent level for the other two dimensions. Findings were robust to sensitivity analyses related to the inclusion or exclusion of dominancy and repeat tasks.

Conclusion

Findings indicated that the MacNew-7D utility value set is likely suitable for estimating quality-adjusted life years derived from the MacNew heart disease health-related quality of life questionnaire. This value set was derived from an Australian population-based sample and may not be generalisable to dissimilar populations.

Introduction

Healthcare represents a significant proportion of total spending in many economies (1). Whether governments or insurance-based systems provide funding, healthcare is a scarce resource. This presents a constrained optimisation problem: healthcare systems seek to maximise expected benefits, subject to scarce resources. With growing and ageing populations, demand for treatments continues to rise, and so does interest in making efficient and effective allocation decisions. Heart disease is a particularly important case in point because of its disease and healthcare resource burden, with patients living longer with debilitating cardiac complications than ever before (2). The management of these conditions requires ever-increasing complex and expensive interventions. For example, in the US, the annual direct cost of managing heart disease was estimated at USD216 billion (3). Although the estimates of cost and disease burden may vary somewhat across industrialised countries, the high disease burden and healthcare resource use pattern is consistent and may even be worse in developing countries. In Australia, cardiovascular diseases comprised approximately 11% of all hospitalisations in 2017-18 (2). They were the underlying cause of 25% of all deaths (2). Aside from these potentially severe health costs, heart disease also constituted 8.9% of total disease expenditure in 2015-16, an estimated AUD10.4 billion (4).

There are recent examples from National Institute for Health and Care Excellence (NICE), Pharmaceutical Benefits Advisory Committee (PBAC) and Medical Services Advisory Committee (MSAC) (5), where utility played an essential role in measuring outcomes in heart disease-related decision-making. A new medication to treat symptomatic chronic heart failure with reduced ejection fraction (Sacubitril valsartan) showed a minimal utility benefit (0.011) measured with the EQ-5D (6). In the absence of a disease-specific instrument, such small increments of utility measured with generic instruments which do not consider all important dimensions of heart disease are relied on for critical decision-making. In 2021 authors from this research group published a comparison between a heart disease-specific quality of life instrument and the EQ-5D-3L in heart failure patients showing the disease-specific instrument having better sensitivity and responsiveness (7). The implication is that generic multi-attribute utility instruments (MAUIs) potentially underestimate any quality-of-life improvement; therefore, the value for money from a cardiac intervention may be undervalued. However, this cannot be established until a condition-specific measure is available.

MAUIs describe health states using as few attributes as possible that define health to generate a preference-based utility value set for a given population. This is achieved by eliciting preference for a subset of health states defined by the MAUI by the respondent group, followed by a utility algorithm that can generate a utility value for every health state defined by the MAUI. The utility values are used to calculate the Quality Adjusted Life Years (QALYs) used in economic evaluations in healthcare. Generic MAUIs such as EQ-5D, AQoL and SF-6D are popular and prominent in country-specific value sets. The generic nature of their attributes in describing the health-related quality of life allows them to be used in comparisons between major health programmes. Condition-specific preference-based measures can complement generic measures by including dimensions relevant to particular disease areas whilst still producing utility values. They are appropriate for economic evaluation where generic measures lack the responsiveness and sensitivity needed to capture critical dimensions of health for a given condition. These preference-based measures can be developed from new, or valuations can be imputed for existing descriptive systems.

A previous paper developed a heart disease-specific classification system (MacNew-7D) based on the MacNew Heart Disease Health-Related Quality of Life (MacNew) instrument, which has demonstrated responsiveness and sensitivity to changes in health-related quality of life for heart disease (8). The MacNew-7D classification system consists of seven dimensions with four levels of severity in each (8). In order to generate utility values from the MacNew-7D, a preference elicitation study is required. Preference elicitation aims to assess the value associated with different health states defined by classification systems such as the MacNew-7D to generate utility values on the 0 (dead) to 1 (full health) scale. This paper values the health states defined by the MacNew-7D using a representative Australian general population sample.

Methods

The valuation was conducted to value the MacNew-7D health states classification system (8) (**Table 1**). This was made up of seven health dimensions: physical restriction, social activity, vitality, frustration, self-confidence, shortness of breath and chest pain. Each dimension is broken into four ordinal levels of severity. In this case, we will denote state 1111111 (i.e. level 1 in each dimension) as full health and state 4444444 as the worst possible health state described by the MacNew-7D. This system allows for 16,284 (=4^7) possible states to be defined. Different methods can elicit preferences in health state valuations, including time trade-off (TTO) and discrete choice experiments (DCE). DCE involves choices over pairs or triplets of health states (presented as scenarios) and is based on random utility theory. Compared to other preference-elicitation techniques, it is easy for participants to understand without an interviewer, making it easier to administer online (9).

This study received ethical approval from the Queensland University of Technology Human Research Ethics Committee (Reference No. #2000000573). Survey respondents were sourced from an existing Australian online panel administered by Pureprofile (www.pureprofile.com). This panel was drawn from volunteers (aged 18 and above who are able to give consent and understand English) in the general population who were paid a small amount by the panel administrators for completion of the survey. Pureprofile has experience in conducting similar DCE studies (10, 11). The targeting respondents were recruited from September to November 2020 with quota sampling using age groups, sex and jurisdictions to represent the Australian population. We conducted two pilots (n=105 and n=305) before initiating the main survey (n=1598).

The DCE used two health 'scenarios' represented by their characteristics or 'attributes' and corresponding levels. Respondents were asked to choose which scenario they would prefer to live in **(Figure 1).** The scenario included duration in each of the two health profiles defined by the dimensions of the MacNew-7D. Duration is included to anchor the utility values on the utility-scale of 0 (dead) and 1 (full health) of the QALY (12-14).

Experimental design

To select the combinations of levels shown to respondents, we developed an efficient experimental design. The DCE created pairs of hypothetical health scenarios, each with different combinations of these attribute levels and a given duration from four duration levels; 1 year, 4 years, 7 years and 10 years (4^8 = 65,536 possible scenarios).

Table 1: Dimensions and levels of the MacNew-7D classification system.

	Dimension	Level	Description
1	Physical restriction	1	You are not limited at all by physical restrictions
		2	You are somewhat limited by physical restrictions
		3	You are moderately limited by physical restrictions
		4	You are extremely limited by physical restrictions
2	Activities	1	You are not excluded from doing things with other people
		2	You are excluded from doing things with other people hardly any of the time
		3	You are excluded from doing things with other people some of the time
		4	You are excluded from doing things with other people all of the time

3	Vitality	1	You are not worn out or low in energy
		2	You are worn out or low in energy hardly any of the time
		3	You are worn out or low in energy some of the time
		4	You are worn out or low in energy all of the time
4	Frustration	1	You are not frustrated, impatient, or angry
		2	You are frustrated, impatient, or angry hardly any of the time
		3	You are frustrated, impatient, or angry some of the time
		4	You are frustrated, impatient or angry all of the time
5	Self-confidence	1	You are not unsure or lacking in self confidence
		2	You are unsure and lacking in self-confidence hardly any of the time
		3	You are unsure and lacking in self-confidence some of the time
		4	You are unsure and lacking in self-confidence all of the time
6	Shortness of breath	1	You have no shortness of breath
		2	You have some shortness of breath
		3	You have moderate shortness of breath
		4	You have extreme shortness of breath
7	Chest pain	1	You have no chest pain
		2	You have chest pain hardly any of the time
		3	You have chest pain some of the time
		4	You have chest pain all of the time

A fractional factorial design was chosen to avoid a costly and impractical surveying process for all scenarios due to the large possible number of pairwise combinations ([65536*65535]/2 = 2,147,450,880). A popular D-efficient design (that maximised the determinant of the variance-covariance matrix of the estimated parameters) was used to create a fractional factorial design with a manageable 200 pairwise choice tasks using Ngene DCE design software (15). Ngene design codes with zero priors were used to retrieve the D-efficient design as there was no literature to identify the priors. Also, evidence suggests that the incorrect or inexact priors could create a bias (16). Additionally, to reduce potential cognitive burden, we imposed three overlapping attributes in each pairwise choice task (i.e. three attribute levels were identical between two options), and the remaining five attributes varied. This overlapping design was considered likely to improve respondent efficiency (17) though it may require broader choice tasks to estimate accurate models compared with non-overlapping designs (9). An example choice set is provided in **Figure 1**.

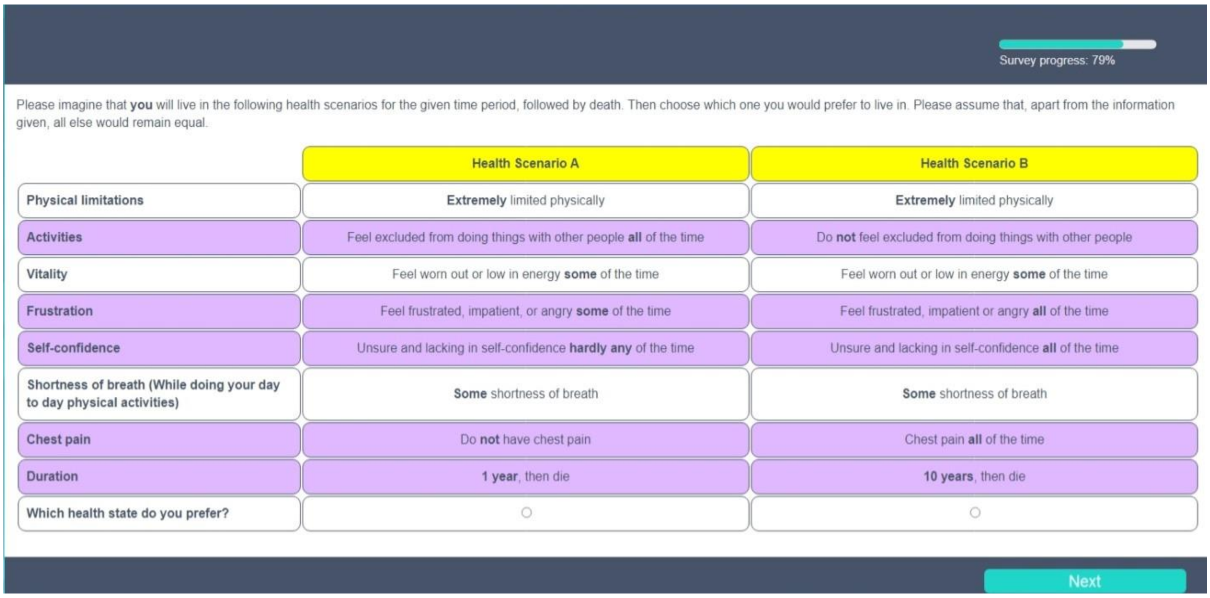

Figure 1: Example choice set

Choice task presentation

Previous research suggests that respondents can reasonably complete 10-15 pairwise choices without cognitive overload, depending on the complexity of the classification system (12, 18). Given that this DCE features 7 health dimensions plus duration, the researchers decided that 10 choice sets per respondent was appropriate (i.e. one block of eight plus one repeated and one dominant choice sets). This resulted in 25 blocks of eight choice sets, and therefore a total of 200 different pairwise choices being undertaken. Each participant completed one of these blocks of choice sets that was randomly allocated to them. The survey and analysis incorporated checks for stability, logical preferences, and engagement (19). One choice pair (the second task) was repeated for each respondent to check for stability in choices. A choice pair was also added where one scenario dominated as it was better than (or equal to) the other in all dimensions. This was to check for understanding of the task and to make sure that respondents were not making illogical decisions. Responses were examined for 'straight-lining', which describes a case in which individuals choose the same scenario (A or B) each time, and the distribution of completion times was analysed to ensure engagement with the task.

Some format-based methods can also be employed to make completing a DCE easier, such as highlighting words and colour-coding. Indeed, Jonker et al. (17) found a combination of colour-coding and level overlap to be even more effective in increasing consistency and reducing dropout rates. This DCE, therefore, incorporated colour-coding, with differing attributes highlighted in purple. The words which signified levels (e.g., not, somewhat, moderately, extremely) were also written in bold.

Data collection

Survey participants were given full information on the study background and objective as well as the content of the survey and asked to give their consent before the survey began. Survey responses were anonymised, and respondents were free to withdraw at any time. The first section of the survey explained the task and its purpose and offered an example choice pair. Respondents were told that the survey was about valuing health states, but not that it was particular to heart disease. They then answered some questions about their own health, including whether they had any long-term conditions (to assess whether they had a chronic heart condition) and to self-rate their health from 0-100 on a visual analogue scale (EQ-VAS). They were then asked to complete the EQ-5D (20) and MacNew-7D (8) questionnaires based on their own health. The second section comprised the choice tasks. After completing all choice-pairs, respondents were asked to rate the difficulty of a) understanding the task and b) making the choices. Finally, respondents were asked to provide demographic information (age, gender, postcode, education, marital status, employment status) to assess how representative the sample was of the general population.

We conducted two pilot studies prior to the data collection proper. Pilot 1 was conducted using a sample of 105. This sample was representative of the Australian general population and represented all state and territories. Pilot 2 used a sample of 305 representing the Australian general population. We implemented several changes for Pilot 2 data collection to mitigate shortcomings identified in the analysis of Pilot 1: improved explanation of the tasks for respondents, including two practice tasks followed by the dominant task; excluding respondents who completed the survey in less than 3 minutes; the addition of a pop up window which explained the choice task again (i.e. why the dominant health state is preferable) to participants who responded incorrectly to the dominant task, asking them to reconsider their choice; and delaying the appearance of the "next" button by 15 seconds, starting from the practice task and continuing throughout the remaining choice tasks. Additionally, the second-choice task was consistently repeated as the ninth task.

The Pilot 2 data were analysed, and the majority of estimated coefficients were found to be in logical order (i.e. utility decrements were increased from the best level to worst level for each attribute) and statistically significant, hence we proceeded to the final data collection. The final data collection (n=1598) had the same survey content as in pilot 2 data collection. Due to incompatibility with the rest of the data collection, Pilot 1 study data were discarded. Therefore, the final analysis included the Pilot 2 and the final data collection information. There is no clear guidance on calculating sample size for health state valuation using DCE, but there are suggestions of 20 respondents per choice set and a sample size of 1000-2000 in contemporary health state valuations (11, 21).

Analysis

The sample was compared with the Australian general population data from Australian Bureau of Statistics (4) and a comparison was made using Chi square tests to establish the representativeness of the study sample using, age, sex, state of residence and education among other characteristics. The quality of life of the sample was described using the MacNew-7D and the EQ-5D-5L responses. All statistical analyses were conducted using Stata 16. Discrete choice experiments are based upon random utility theory where utility is defined to be made up of a systematic and random component. We used the conditional logit model developed by McFadden (22) and used by Bansback et al. (12) to model health state valuation using DCE_{TTO} data. To allow for latent utilities to be anchored on the QALY scale, the following model specification including life duration as a continuous attribute and interactions between a health state and duration was adopted (12).

$$U_{ni} = \beta_0 duration_{ni} + \beta'_1 \boldsymbol{x}_{ni} * duration_{ni} + \varepsilon_{nj}$$

U_{ni} is the utility individual n gets from alternative I, and $\boldsymbol{x}_{ni}$ is a vector of 21 binary dummy variables$(x_{ni}^{12}, x_{ni}^{13}, x_{ni}^{14}, \ldots, x_{ni}^{74})$, representing each level of each of the seven health attributes. For example, x_{ni}^{12} indicates the physical dimension at level 2 (somewhat physically restricted). Level 1 is omitted for each attribute to avoid perfect collinearity, and therefore the 'baseline' represents perfect health state 1111111. Hence $\boldsymbol{\beta}_1$ is a vector $(\beta_1^{12}, \beta_1^{12}, \beta_1^{12}, \ldots, \beta_1^{12})$ estimating the utility associated with each attribute-level for $duration_{ni}$. The attribute-levels $\boldsymbol{x}_{ni}$ are not included independently of time (i.e. without interaction with duration) as this would imply that individuals can gain utility from a health state that they lived in for 0 years (12).

To anchor these estimated utilities onto the 0 to 1 (death to full health) scale used in QALY calculations, the trade-off between changes in health state and duration is examined:

$$MU_{ni} = \frac{\beta'_i x_{ni} * duration_{ni}}{\beta_0 duration_{ni}}$$

These estimates can be interpreted similarly to time trade off (TTO), as the average amount of life expectancy that respondents are willing to sacrifice for an improvement in the given health dimension.

In addition to the main analysis using the full sample, further analysis was carried out excluding those who gave different answers on the repeated question and incorrect answers to the dominated choice pair.

Main analysis was conducted using the full sample. Sensitivity analysis was conducted to check the robustness of the data by excluding the participants who did not answer repeat task and dominant task correctly. The significance of coefficients were also considered in model evaluation.

Logical consistency of coefficients

When the model was estimated using conditional logit, different combinations were attempted to achieve a parsimonious model. When a coefficient representing a worse level has higher value than a better level, it can be regarded as an inconsistent state. Any inconsistent coefficients were combined with immediate upper or lower level to achieve consistency. Only the best fitting models were selected, based on consistency, parsimony, and goodness of fit.

Finally, we reviewed the study using the 21-item Checklist for REporting VAluaTion StudiEs (CREATE) checklist for reporting valuation studies of multi-attribute utility-based instruments (23).

Results

As there were no survey changes following the second pilot, we combined the two samples from pilot 2 (n=305) and the final survey (n=1598), for the main analysis (n=1903). Complete survey respondents' information is included in the **Supplementary File 1**. The sample was similar to the characteristics of the Australian general population with age, sex and distribution of the state of residence (24) (Table 2). The sample was equally distributed through Australian state and territories jurisdictions proportional to population levels. The mean age of the sample was 47 years. There were 221 respondents who indicated they had diabetes and 95 with either coronary artery disease or previously had a stroke. Over 90% of respondents reported the DCE tasks were not difficult at all or somewhat difficult to understand

Table 2: Socio-demographic characteristics of the sample compared with the Australian general population (n= 1903)

Characteristic	Sample Number	%	Population value*	Statistic	P-value#
Age, mean (SD)	47.2 (17.8)		39.09		
Age (Years)					
18 – 34	557	29.3	31.3	$X^2 = 2.1057$	0.5507
35 – 54	668	35.1	33.4		
55 – 74	495	26.0	26.4		
> 75	171	9.0	8.9		
Prefer not to say	12	0.6			

Sex					
Male	940	49.4	49.3	$X^2 = 0.06$	0.8086
Female	956	50.2	50.7		
Prefer not to say	7	0.4			
State of residence					
New South Wales	605	31.8	31.80	$X^2 = 2.44$	0.9314
Victoria	488	25.6	26.08		
Queensland	385	20.2	20.12		
South Australia	135	7.1	6.89		
Western Australia	199	10.5	10.36		
Tasmania	40	2.1	2.10		
Australia Capital Territory	31	1.6	1.68		
Northern Territory	12	0.6	0.96		
Prefer not to say	8	0.4			
Highest level of Education, n (%)					
Grade 10	190	10.0	15.1^	$X^2 = 264.69$	< 0.001
Grade 12	284	14.9	18.0		
Certificate II-IV	311	16.3	17.6		
Diploma	283	14.9	13.3[$]		
Bachelor's degree	540	28.4	19.6		
Postgraduate Degree	250	13.1	8.1		
Other	30	1.6	6.6[@]		
Prefer not to say	15	0.8			
Marital Status, n (%)					
Single	498	26.1			
Married/De facto	1,164	61.1			
Divorced	142	7.5			
Widowed	68	3.6			
Other	19	1.0			
Prefer not to say	12	0.6			
Current employment status, n (%)					
Full-time employment	762	40.0			
Part-time employment	357	18.8			
Unemployed	174	9.1			
Disability pension	87	4.6			
Retired	389	20.4			
Other	106	5.6			
Prefer not to say	28	1.5			
Comorbidities					
Previous accidents	58	3.0			
Cancer	60	3.1			
Coronary heart disease	63	3.3			
Chronic kidney disease	29	1.5			
Diabetes	221	11.6			
Liver disease	21	1.1			
Stroke	32	1.7			
Difficulty in understanding, n (%)					
Not difficult at all	1147	60.3			
Somewhat difficult	584	30.7			
Difficult	102	5.4			
Extremely difficult	70	3.7			

Notes:

** Australian age and sex distribution (Australian Bureau of statistics, June 2020 from https://www.abs.gov.au/. The highest level of education (Education and Work, Australia, Australian Bureau of Statistics, May 2021 from https://www.abs.gov.au/).*

 # The chi-squared goodness-of-fit test was used to compare observed frequencies with population proportions

^Australian population data for Grade 10 and 11
$Australian population data for Advanced/graduate diploma
@Australian population data for highest education level below Grade 10

Responses to MacNew-7D and EQ-5D-5L

Mean values for the sample using the EQ-5D-5L and MacNew-7D values were 0.8 (SD 0.24) and 0.79 (SD 0.24), respectively, indicating a reasonably healthy sample compared with the general population norms (25). The majority (>75%) of the respondents had no problems in physical dimensions whilst nearly 50% reported concerns in non-physical health dimensions (anxiety/ depression, frustrated, impatient or angry and worn out or low in energy). Additionally, between 2% and 9% of respondents reported severe levels across the MacNew-7D dimensions. Shortness of breath while doing normal activities (37.5%) and chest pain (20%) were also identified in the sample.

Table 3: EQ-5D-5L & MacNew-7D

Dimension		Level 1 n (%)	Level 2 n (%)	Level 3 n (%)	Level 4 n (%)	Level 5 n (%)
EQ-5D-5L						
Mobility		1413 (74.3)	290 (15.2)	133 (7.0)	57 (3.0)	10 (0.5)
Self-care		1652 (86.8)	146 (7.7)	70 (3.7)	27 (1.4)	8 (0.4)
Usual activities		1398 (73.5)	282 (14.8)	164 (8.6)	43 (2.3)	16 (0.8)
Pain/discomfort		849 (44.6)	641 (33.7)	275 (14.5)	99 (5.2)	39 (2.0)
Anxiety/depression		995 (52.3)	481 (25.3)	273 (14.3)	97 (5.1)	57 (3.0)
Mac-New-7D						
Physical restriction		1192 (62.6)	476 (25.0)	158 (8.3)	77 (4.0)	
Excluded from doing things with other people		1206 (63.4)	302 (15.9)	317 (16.7)	78 (4.1)	
Worn out or low in energy		677 (35.6)	423 (22.2)	636 (33.4)	167 (8.8)	
Frustrated, impatient or angry		844 (44.4)	495 (26.0)	488 (25.6)	76 (4.0)	
Unsure and lacking in self-confidence		822 (43.2)	419 (22.0)	494 (26.0)	168 (8.8)	
Shortness of breath while doing your day to day physical activities		1189 (62.5)	514 (27.0)	143 (7.5)	57 (3.0)	
Chest pain		1523 (80.0)	226 (11.9)	120 (6.3)	34 (1.8)	

Note: Levels represent ordered increase in severity of health impact within each domain

Model results

Of the total sample, 234 (12%) did not answer dominant task correctly and 395 (21%) did not answer similarly to the repeat task **(Supplementary File 1)**. There was no substantial difference to the model in relation to inconsistency when these respondents were removed from the analysis. Therefore, the final primary analysis included the whole sample. We have presented the model estimations for the sample excluding dominant task and repeat task separately as a sensitivity analysis in Supplementary File 1.

The final models included the full sample (n=1903) and different collapsing methods to determine the best parsimonious model. Dimensions "worn out" and "self-confidence" were collapsed to look at how this affects the observed inconsistency (Table 4). Model 1 did not involve collapsing. With model 1, coefficients for "worn out" (level 3) and "self-confidence" (level 1) were not in the expected direction.

Model 2 collapsed level 3 and 4 of the "worn-out" attribute and level 2 and 3 of the "self-confidence" attributes (Table 4). Although the direction of the collapsed level 3_4 of the “worn out” attribute corrected, the severe level had lower value than the mild level (-0.0021 and -0.0110) indicating inconsistency. In addition, these coefficients were insignificant. With "self-confidence" attribute with collapsed levels direction changed for level 2_3 and consistency was observed. This can be excluded based inconsistency of the "worn-out" attribute.

Model 3 collapsed level 2 and 3 of the "worn-out" attribute and level 1 and 2 of the "self-confidence" attributes. The collapsed level 2_3 of the "worn-out" attribute did not change the direction showing inconsistency and the collapsed level 1_2 of the "self-confidence" attribute showing no difference from “no problem”. Again, although the collapsed "self-confidence" attribute showed consistency the "worn-out" attribute was inconsistent, excluding the results of this model.

Model 4 collapsed level 3 and 4 of the "worn-out" attribute and level 1 and 2 of the "self-confidence" attribute. This made the worn-out attribute consistent albeit with an insignificant level 2 coefficient. "Self-confidence" attribute also recorded consistency with significant coefficients. Based on this result model 4 was considered as the best model to develop utility value set for the MacNew-7D classification system (Table 4). The Australian MacNew-7D utility value set ranged from -0.4456 to 1.000 for the health states defined by the classification system. Of the 16384 possible health states, approximately 4% (n=662) were negative, indicating that they were worse than death health states. These “worse than death” health states include severe level (level 4) for at least three dimensions.

As an example, utility value for health state 2221132 is = 1-(0.1087+0.0972+0.0143+0.0000+0.0000+0.0751+0.0485) = 0.6562. The full health state (1111111) has a value of 1. The worst possible health state (4444444) has a value of -0.4116.

Further, utility decrements provided in Table 4 can be used to calculate utility values for the research studies conducted using the MacNew 27 items quality of life questionnaire. R codes and STATA codes to derive these utility values from the data set of MacNew 27 items scale are provided in Supplementary Files 2 and 3.

Table 4: Estimated coefficients for the Model comparison with collapsing attribute levels (Conditional logit, n=1903)

		Model 1*		**Model 2****		**Model 3*****		**Model 4******		**Anchored values for model 4 #**	
		Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Utility decrement	Coefficient95% CI
Duration		**0.4352**	0.0150	**0.4366**	0.0149	**0.4370**	0.0149	**0.4421**	0.0149		
Physical restriction x duration											
	2	**-0.0483**	0.0074	**-0.0443**	0.0073	**-0.0488**	0.0074	**-0.0480**	0.0074	-0.1087	(-0.0774 to -0.1400)
	3	**-0.0503**	0.0069	**-0.0520**	0.0068	**-0.0512**	0.0069	**-0.0544**	0.0068	-0.1232	(-0.0938 to -0.1525)
	4	**-0.1526**	0.0078	**-0.1582**	0.0077	**-0.1539**	0.0077	**-0.1589**	0.0078	-0.3595	(-0.3259 to -0.3930)
Exclusion of Activities x duration											
	2	**-0.0473**	0.0077	**-0.0426**	0.0076	**-0.0467**	0.0077	**-0.0430**	0.0076	-0.0972	(-0.0649 to -0.1295)
	3	**-0.0524**	0.0070	**-0.0462**	0.0068	**-0.0498**	0.0069	**-0.0466**	0.0069	-0.1055	(-0.0766 to -0.1344)
	4	**-0.0791**	0.0084	**-0.0834**	0.0082	**-0.0801**	0.0082	**-0.0829**	0.0082	-0.1875	(-0.1528 to -0.2222)
Worn out x duration											
	2	-0.0094	0.0068	-0.0021	0.0067	0.0009	0.0059	-0.0063	0.0068	-0.0143	(0.0155 to -0.0440)
	3	0.0087	0.0068	-0.0110	0.0059	0.0009	0.0059	**-0.0120**	0.0059	-0.0272	(-0.0017 to -0.0528)
	4	**-0.0382**	0.0069	-0.0110	0.0059	**-0.0361**	0.0068	**-0.0120**	0.0059	-0.0272	(-0.0017 to -0.0528)
Frustration x duration											
	2	**-0.0144**	0.0067	**-0.0157**	0.0067	**-0.0137**	0.0067	**-0.0149**	0.0067	-0.0337	(-0.0045 to -0.0628)
	3	**-0.0272**	0.0065	**-0.0277**	0.0064	**-0.0269**	0.0065	**-0.0255**	0.0064	-0.0578	(-0.0296 to -0.0859)
	4	**-0.0805**	0.0077	**-0.0810**	0.0074	**-0.0775**	0.0076	**-0.0764**	0.0075	-0.1727	(-0.1405 to -0.2049)
Self-confidence x duration											
	2	0.0094	0.0075	-0.0099	0.0064	0.0000		0.0000		0000	
	3	**-0.0265**	0.0072	-0.0099	0.0064	**-0.0286**	0.0061	**-0.0284**	0.0060	-0.0642	(-0.0379 to -0.0904)
	4	**-0.0432**	0.0067	**-0.0469**	0.0066	**-0.0485**	0.0054	**-0.0495**	0.0055	-0.1119	(-0.0895 to -0.1343)
Shortness of Breath x duration											
	2	-0.0150	0.0082	-0.0106	0.0080	-0.0150	0.0080	-0.0134	0.0080	-0.0304	(0.0047 to -0.0654)
	3	**-0.0290**	0.0079	**-0.0332**	0.0078	**-0.0303**	0.0078	**-0.0332**	0.0078	-0.0751	(-0.0418 to -0.1084)
	4	**-0.1156**	0.0078	**-0.1204**	0.0078	**-0.1176**	0.0077	**-0.1197**	0.0078	-0.2707	(-0.2383 to -0.3031)
Chest Pain x duration											
	2	**-0.0201**	0.0084	**-0.0195**	0.0083	**-0.0175**	0.0083	**-0.0214**	0.0084	-0.0485	(-0.0121 to -0.0849)
	3	**-0.0539**	0.0070	**-0.0534**	0.0069	**-0.0541**	0.0069	**-0.0556**	0.0069	-0.1257	(-0.0958 to -0.1555)
	4	**-0.1206**	0.0082	**-0.1228**	0.0081	**-0.1185**	0.0082	**-0.1247**	0.0081	-0.2821	(-0.2485 to -0.3157)
Log likelihood		-8556		-8592		-8560		-8582			
AIC		17156		17225		17161		17205			
BIC		17339		17392		17328		17371			

**Model 1: Full model without collapsing any attribute levels*

***Model 2: Collapsing levels 3 and 4 of the Worn-out attribute and levels 2 and 3 of the self-confidence attribute*

****Model 3: Collapsing levels 2 and 3 of the Worn-out attribute and levels 1 and 2 of the self-confidence attribute*

*******Model 4: Collapsing levels 3 and 4 of the Worn-out attribute and level 1 and 2 of the self-confidence attributes- the selected model***

#Anchored value = estimated coefficient for each level in dimension/ duration coefficient

Significant coefficients at $p<0.05$ level are in bold. AIC Akaike information criterion, BIC Bayesian information criterion, SE Standard error

Discussion

We report the development of the MacNew-7D utility value set, derived from a representative Australian general population sample. These values can be used to estimate heart disease specific utility values and will be useful in the cost-utility analysis (CUA) of heart disease-related health interventions. This utility set enables utility values to be directly estimated from the heart disease-specific MacNew quality of life questionnaire. The MacNew-7D utility values have potential to be more sensitive for heart-related problems such as chest pain, fatigue and social issues that are exacerbated by a chronic debilitating disease. However, further investigation of the use of this instrument among people with heart disease will be beneficial for confirming or refuting the relative sensitivity of utility sets derived from the MacNew-7D compared to generic instruments for estimating utility.

Although the MacNew-7D is intended for use among people with heart disease, members of the general population are appropriate for valuing health states. The general public may fund healthcare either directly or indirectly though taxation and insurance mechanisms. Their involvement in healthcare-related decision making is advocated in many publicly funded healthcare systems, including NICE and the US Public Health Service. Based on the public preferences for the MacNew health state values, the physical restriction dimension recorded the highest decrements - especially with increased severity - indicating the significant role generic quality of life components played in the valuation. However, heart disease-specific dimensions such as chest pain and shortness of breath also created substantial utility decrements. All three emotional dimensions ("worn out", "frustration", "self-confidence") contributed small utility decrements as well as the social dimension "exclusion of activities". These could be considered as generic and widely prevalent within the general population, therefore participants may not readily consider trade-offs based on these attributes due to personal adaptations related to their own health state exposure (26). However, although generic in nature, "physical restriction" was associated with significantly large disutility. The inclusion of "chest pain" and "breathlessness" attributes in the MacNew classification system may have created a negative trade off within respondents who were generally healthy with no heart-related issues. Having a sizeable significant coefficient for duration indicated respondents' tendency to want to live longer. We used the DCE with duration (DCE_{TTO}) for this health state valuation. Health state valuation has traditionally been conducted using face-to-face Time Trade-Off (TTO) and standard gamble preference elicitation methods. The development of ordinal techniques to value health states led to increased use of choice experiment methodology in eliciting preferences in health state valuation. Their online applicability and relative cost and time savings for research teams could be another cause of the popularity of quick adoption. However DCEs, which is the common choice methodology used in current valuations, cannot produce utility estimates along the full health (1) and death (0) scale due to their inability to anchor these values. Therefore additional measures have been adopted to anchor the full health and death utility scale, including duration (DCE_{TTO}), additional TTO data, or ranking. The TTO method to anchor the values still cannot be fully utilised online and the analysis constraints have not been fully resolved. The DCE_{TTO} method has been popular due to fewer methodological constraints, want of additional data and easy online adaptability.

This is the first of country-specific value set for the MacNew-7D. There is no other value set for the same attributes and levels to be compared with and there are no other heart disease-specific MAUI developed so far to warrant a comparison. However, compared with other disease-specific MAUIs, this instrument demonstrates validity to a certain extent. The cancer disease-specific QLQ-

C10-D has a lowermost value of -0.096 compared with the -0.4116 in this instrument (27). This lower utility value could be due to the outlook of the presented attributes, such as "physical restriction" (-0.3595), "shortness of breath" (-0.2707) and "chest pain" (-0.2821) which fetched higher disutility as compared with "physical functioning" (-0.250), "pain" (-0.155) and "emotional functioning" (-0.133) presented in cancer-specific health states. The worst value in the MacNew-7D is also much lower than the reported Canadian EQ-5D-5L worst health state (55555) of -0.148 (28). In contrast, the worst health state utility value of the Australian dementia-specific AD-5D was much closer to the MacNew-7D, with -0.68 (11). The distribution of the MacNew-7D utility values is not vastly different to other reported value sets, but recorded lower values for severe health states than generic measures expected, as the worse cases attributes in this valuation are deemed severe.

The MacNew Heart Failure questionnaire and its predecessor have been used to measure the quality of life in heart disease (29). It has been used not only in heart failure patients but also in other heart-related diseases. Its reliability, validity and psychometric properties have been validated many times over the last decade. The MacNew-7D classification system was developed using this quality-of-life measure, and the utility value set will establish its role as a heart disease-specific preference-based measure, that may contribute to future policy decision making in related resource allocations. The monotonicities observed are commonplace in regression models used in health state valuations(27, 30). Previously we described the development of the classification system(8), and this is the last stage of the development of a value set to complete the classification system. Further studies should be conducted in other jurisdictions to develop country-specific value sets and compare them with generic values. The ability of MacNew-7D in capturing heart disease-related specific characteristics extends across physical, emotional, and social dimensions defining the quality of life of such patients. It captures mobility restrictions experienced by patients, and the discomfort cause by an inability to engage with social, family, and recreational activities. The emotional toll of living with heart disease is displayed by the frustration, feeling worn out, and the level of self-confidence in engaging with the family, community, and activities. The need to depend on other people and the limitations on social and family activities frustrate people and lower their self-confidence. These essential but specific characteristics are difficult to pick up in a generic instrument such as the EQ-5D. The MacNew-7D incorporates these characteristics and will be sensitive enough to discern these changes in quality of life and convert them to utility changes in future data collections. We recommend analysis comparing generic utility measures with the MacNew-7D to determine effects on an incremental cost-effectiveness ratio (ICER) in future heart disease-related interventions.

Although MacNew-7D is a new instrument, the MacNew quality of life questionnaire has been used for many years in randomised control trials and other surveys. The availability of MacNew-7D will improve usability by providing quality of life measures and utility estimates using the same instrument. Past RCTs that used MacNew and were unable to estimate QALYs can now estimate these to progress economic evaluations and provide additional evidence for decision-makers. We recommend data collection using the MacNew questionnaire and the algorithm provided in the appendix to calculate a utility value for each instance. In the future, where utility values only are needed for a study, MacNew-7D classification could be used for data collection once further validity has been reported.

Strengths, limitations and future research

This study used a relatively large representative sample to conduct the DCE and evaluate references for health states. We excluded respondents who completed the survey in less than three minutes without downloading their responses or including them in the analysis to mitigate against risk associated with responses that were not appropriately considered. In addition, there were around 8% who refused to participate in the survey and we were not able to ascertain anything about their preferences. Also, online samples may have specific characteristics that differentiate them from the wider population, including higher educational levels or differing socio-economic characteristics. However, the large sample matched on the basis of population characteristics to the general population and geographical distribution helped mitigate risk of sampling bias. The reported level of inconsistencies related to the dominant and repeated tasks of this study (Supplementary File 1) were below the inconsistencies level reported in other DCE studies (19) and the low percentage of logical inconsistencies improved the confidence of the results. However, this study was for the general Australian population and may not be generalisable to dissimilar populations. Although quota sampling was used to generate a sample that was as representative of the Australian population as possible, the mean age of the sample was eight years higher than the mean age of 18+ population. . The use of conditional logit model used in this and other DCE studies also assumes homogeneity of preferences, which may mean inaccuracies are introduced with this approach (31). Although the solution to counteract this is to test for heterogeneity using mixed models, they are not-well suited for conversion into utility value sets. For this reason, conditional logit models remain the most widely used approach in this field (31).

Important future research includes analyses to create mapping value sets for quality of life questionnaires of Minnesota Living with Heart Failure Questionnaire (32), Seattle Angina Questionnaire (33), and The Kansas City Cardiomyopathy Questionnaire (34) using the MacNew-7D value set, which members of our research team have commenced. These values will also be compared with existing generic measures such as EQ-5D and SF-6D. Therefore, once further comparative and validation work is completed, the MacNew-7D value set has potential to reduce data collection time and effort, allowing the use of one quality of life data collection for both patient-reported outcomes and utility changes in clinical trials for heart disease or for real-world comparisons.

Acknowledgements

School of Health and Related Research, The University of Sheffield, hosted Dr. Kularatna and facilitated the exchange of ideas between academic groups.

Reference

1. OECD. Health at a Glance 2019: OECD Indicators. Paris: OECD; 2019.
2. Australian Institute of Health and Welfare. Cardiovascular disease. Canberra: AIHW; 2020.
3. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association. Circulation. 2021;143(8):e254-e743.
4. Statistics ABo. Census of Population and Housing (Usual residence data). 2016.
5. Medical Services Advisory Committee. 1387 - Cardiac Contractility Modulation (OPTIMIZER™ IVs). In: Health Do, editor. Australia2015
6. Health NIf, Excellence C. Sacubitril valsartan for treating symptomatic chronic heart failure with reduced ejection fraction. Technology appraisal. 2016;388(TA388).
7. Kularatna S, Byrnes J, Chan YK, Carrington MJ, Stewart S, Scuffham PA. Comparison of contemporaneous responses for EQ-5D-3L and Minnesota Living with Heart Failure; a case for disease specific multiattribute utility instrument in cardiovascular conditions. Int J Cardiol. 2017;227:172-6.
8. Kularatna S, Rowen D, Mukuria C, McPhail S, Chen G, Mulhern B, et al. Development of a preference-based heart disease-specific health state classification system using MacNew heart disease-related quality of life instrument. Qual Life Res. 2021.
9. Mulhern B, Norman R, Street DJ, Viney R. One Method, Many Methodological Choices: A Structured Review of Discrete-Choice Experiments for Health State Valuation. Pharmacoeconomics. 2019;37(1):29-43.

10. Norman R, Viney R, Aaronson NK, Brazier JE, Cella D, Costa DSJ, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. Quality of Life Research. 2016;25(3):637-49.
11. Comans TA, Nguyen K-H, Ratcliffe J, Rowen D, Mulhern B. Valuing the AD-5D Dementia Utility Instrument: An Estimation of a General Population Tariff. PharmacoEconomics. 2020;38(8):871-81.
12. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. J Health Econ. 2012;31(1):306-18.
13. Bansback N, Hole AR, Mulhern B, Tsuchiya A. Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues. Soc Sci Med. 2014;114:38-48.
14. Norman R, Mercieca-Bebber R, Rowen D, Brazier JE, Cella D, Pickard AS, et al. U.K. utility weights for the EORTC QLU-C10D. Health Econ. 2019;28(12):1385-401.
15. Choicemetrics. Ngene 1.1.1 User Manual & Reference Guide2012.
16. Johnson F, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health. 2013;16(1):3-13.
17. Jonker MF, Donkers B, de Bekker-Grob E, Stolk EA. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. Health Economics. 2019;28(3):350-63.
18. Coast J, Flynn TN, Salisbury C, Louviere J, Peters TJ. Maximising responses to discrete choice experiments: a randomised trial. Appl Health Econ Health Policy. 2006;5(4):249-60.
19. Johnson FR, Yang JC, Reed SD. The Internal Validity of Discrete Choice Experiment Data: A Testing Tool for Quantitative Assessments. Value Health. 2019;22(2):157-60.
20. EuroQoL. EQ-5D-5L About [Available from: https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/.
21. Hettiarachchi R, Kularatna S, Byrnes J, Mulhern B, Chen G, Scuffham PA. Valuation study for a preference-based quality of life measure for dental caries (Dental Caries Utility Index - DCUI) among Australian adolescents - study protocol. BMJ Open. 2020;10(10):e038626.
22. McFadden D. Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka P, editor. Frontiers in Econometrics. New York: Academic Press; 1974. p. 105-42.
23. Xie F, Pickard AS, Krabbe PF, Revicki D, Viney R, Devlin N, et al. A Checklist for Reporting Valuation Studies of Multi-Attribute Utility-Based Instruments (CREATE). Pharmacoeconomics. 2015;33(8):867-77.
24. Australian Bureau of Statistics. National, State and territory population. In: ABS, editor. Australia 2021
25. McCaffrey N, Kaambwa B, Currow DC, Ratcliffe J. Health-related quality of life measured using the EQ-5D-5L: South Australian population norms. Health Qual Life Outcomes. 2016;14(1):133.
26. McPhail S, Haines T. Response shift, recall bias and their effect on measuring change in health-related quality of life amongst older hospital patients. Health Qual Life Outcomes. 2010;8:65.
27. King MT, Viney R, Simon Pickard A, Rowen D, Aaronson NK, Brazier JE, et al. Australian Utility Weights for the EORTC QLU-C10D, a Multi-Attribute Utility Instrument Derived from the Cancer-Specific Quality of Life Questionnaire, EORTC QLQ-C30. Pharmacoeconomics. 2018;36(2):225-38.
28. Xie F, Pullenayegum E, Gaebel K, Bansback N, Bryan S, Ohinmaa A, et al. A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada. Med Care. 2016;54(1):98-105.
29. Höfer S, Lim L, Guyatt G, Oldridge N. The MacNew Heart Disease health-related quality of life instrument: a summary. Health Qual Life Outcomes. 2004;2:3.
30. Kularatna S, Whitty JA, Johnson NW, Jayasinghe R, Scuffham PA. Valuing EQ-5D health states for Sri Lanka. Qual Life Res. 2015;24(7):1785-93.

31. Hauber AB, González JM, Groothuis-Oudshoorn CGM, Prior T, Marshall DA, Cunningham C, et al. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value in Health. 2016;19(4):300-15.
32. Bilbao A, Escobar A, García-Perez L, Navarro G, Quirós R. The Minnesota living with heart failure questionnaire: comparison of different factor structures. Health and quality of life outcomes. 2016;14:23-.
33. Spertus JA, Winder JA, Dewhurst TA, Deyo RA, Prodzinski J, McDonell M, et al. Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. Journal of the American College of Cardiology. 1995;25(2):333-41.
34. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol. 2000;35(5):1245-55.

Statements and Declarations

Funding

This study was funded by an Australian Heart Foundation Vanguard grant (2016/101407) and Heart Foundation post-doctoral fellowship for Dr. Kularatna.

Centre for Healthcare Transformation, Queensland University of Technology, funded the DCE data collection.

SMM is supported by an NHMRC administered fellowship #1161138.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose

Author contribution

SK developed the idea, conducted the analysis and wrote the first draft of the paper. GC, RN, CM, DR, BM contributed to the plan of data collection, DCE design and contributed to the writing of the manuscript. SS,RH, KF did the data collection, assisted the data analysis and contributed to the manuscript. WP and SM assisted in developing the idea, data collection and writing the manuscript. All authors read and approved the final manuscript.

Ethics approval

This study received ethical approval from the Queensland University of Technology Human Research Ethics Committee (Reference No. #2000000573).

Consent to participate

Survey respondents were sourced from an existing Australian online panel administered by Pureprofile (www.pureprofile.com). This panel was drawn from volunteers (aged 18 and above, be able to give consent and understand English) in the general population who were paid a small amount by the panel administrators for completion of the survey.