An EQ-5D-5L Value Set Based on Preferences of Patients with Heart Disease

Running title: An EQ-5D-5L value set using patient preferences

Mihir Gandhi[1,2,3*], Ru-San Tan[4], Shir Lynn Lim[5], Kim Rand[6,7], Carolyn S.P. Lam[4], Nan Luo[8], Yin Bun Cheung[2,3]

[1]Singapore Clinical Research Institute, Singapore

[2]Duke-NUS Medical School, Singapore

[3]Tampere University, Finland

[4]National Heart Centre Singapore, Singapore

[5]National University Heart Centre, Singapore

[6]Health Services Research Centre, Akershus University Hospital, Lørenskog, Norway

[7]Maths in Health B. V, Rotterdam, the Netherlands

[8]Saw Swee Hock School of Public Health, National University of Singapore, Singapore

*** Corresponding author:**

Mihir Gandhi, PhD

Department of Biostatistics, Singapore Clinical Research Institute

Centre for Quantitative Medicine, Duke-NUS Medical School

#02-01, Nanos, 31 Biopolis Way, Singapore 138669

Email: mihir.gandhi@scri.cris.sg

Phone: +65-6508 8309

Funding: The study was jointly funded by the Duke-NUS Medical School Signature Research Program (funded by the Agency for Science, Technology and Research (A*STAR), Singapore, and the Ministry of Health, Singapore; Account R-913-200-040-263) and the EuroQol Research Foundation (EQ Project 20170430).

Acknowledgments: We thank Tan Gretchen, Woo Kai Lee, and Koh Rui Xuan Audrey for their contribution to the data collection. The Clinical Trials Research Office, National Heart Centre Singapore, has provided administrative support.

Word count (maximum 4000): 3938

Number of tables & figures (maximum 6): 6

Abstract

Objectives: Several studies have shown that patients with heart disease value hypothetical health states differently from the general population. We aimed to develop a value set for the EQ-5D-5L based on the preferences of patients with heart disease using an international valuation protocol.

Methods: Patients with confirmed heart disease were recruited from two hospitals in Singapore. A total of 86 EQ-5D-5L health states (10 per patient) were valued using a composite time trade-off method according to the international valuation protocol for EQ-5D-5L. A 20-parameter linear model and an 8-parameter cross-attribute level effects model with and without an N45 term (indicating whether any health state dimension at level 4 or 5 existed) were estimated. Each model included patient-specific random intercepts, heteroscedastic error, and left-censored utility values at -1. Model performance was evaluated in terms of out-of-sample and in-sample predictive accuracy. The discriminative ability of the utility values was assessed using heart disease-related functional classes.

Results: A total of 539 patients were included in the analysis. The preferred model is a 20-parameter linear model with the N45 term. Predicted utility values ranged from -0.928 for the worst state to 1 for full health; the value for the second-best state was 0.982. Utility values demonstrated good discriminative ability in differentiating among patients of varied functional classes.

Conclusions: For the first time, an EQ-5D-5L value set was developed using the preferences of patients with heart disease. The value set could be used for patient-centric economic evaluation and treatment selection for patients with heart disease.

Highlights

- Patients with heart disease have health preferences that are systematically different from those of the general public.
- This study provides utility values for EQ-5D-5L health states based on the preferences of patients with heart disease.
- This value set is useful for clinical decision-making and for economic evaluation aiming to use patient preferences to inform treatment selection or reimbursement for patients with heart disease.

Introduction

Medical costs are escalating with population aging and advances in healthcare technologies[1], and these changes are placing pressure on national healthcare budgets. Health technology assessment (HTA) helps make efficient use of healthcare budgets. Using quality-adjusted life years (QALYs), HTA evaluates the costs of new treatment taking into consideration survival benefits and effects on health-related quality of life. QALYs are typically obtained from a generic preference-based instrument such as the EQ-5D that provides a utility value that is multiplied by the duration lived in a health state. Utility values are usually estimated by asking people to assign values to specific hypothetical health states that vary in severity from mild to extremely severe.

HTA methods and processes have been criticized for not being sufficiently "patient-centric"[2]. For example, several countries recommend that the reference case analysis be based on a societal perspective[3]. However, there are doubts about whether members of the general public who are relatively healthy can appreciate the health states that they are being asked to value. Therefore, there are arguments that support using the preferences of patients who have experienced health states of varying severity[4]. The Dental and Pharmaceutical Benefits Agency in Sweden recommends that the preferences of persons who have experienced the particular health condition that is being assessed be used in economic evaluations[5]. For medical technology evaluations, the Agency for Care Effectiveness in Singapore recommends the use of preferences based both on patients with the condition and on the general public[6,7].

The choice of using patient or general public preferences depends on the purpose and context of the evaluation. General public values are desirable when the values are used to inform decisions that allocate societal resources, while patient values may be more appropriate when making treatment decisions guided by patient health preferences. Patient preferences are essential for patient-centric healthcare decisions. In many counties, patient preferences are also crucial for economic evaluations, as patients themselves bear the majority of healthcare costs[8].

Empirical studies have shown systematic differences in the valuing of hypothetical health states by the general public and by patients with certain health conditions, such as heart disease[9-11], that are not explained by differences in sociodemographic characteristics. Pickard et al.[11] and Gandhi et al.[10] showed that patients with heart disease give higher values than the general public for the 3-level EQ-5D (EQ-5D-3L) health states. Differences in values given by patients with heart disease and healthy people were also reported for the 5-level EQ-5D (EQ-5D-5L) health states[9], along with the impact of these differences on utility gain estimates. These differences could occur for several reasons: variation in life experiences, uncertainty about life, adaptation to suboptimal health conditions and healthcare costs. These findings support the use of utility values based on patient preferences for patient-centric healthcare decision-making.

Cardiovascular diseases (CVDs), which include ischemic heart disease, stroke, peripheral arterial disease, heart failure, and several other cardiac and vascular conditions, contribute to more than 400 million new cases, 18 million deaths (31% of all mortality), and 36 million years of lived-with-disability per year worldwide[12]. Considering the disease prevalence, burden, and health preferences, a utility value set based on preferences of patients with heart disease will potentially have a significant impact on the evaluation of emerging therapies for CVD.

In this study, we aimed to develop a utility value set for health states defined by the EQ-5D-5L descriptive system using the preferences of patients with heart disease. The EQ-5D-5L is a new version of the widely used EQ-5D-3L and has demonstrated better measurement properties than the previous version[13].

Methods

Study design and participants

This was a cross-sectional study involving face-to-face interviews of patients with heart disease who were receiving treatment at the two largest cardiovascular tertiary hospitals in Singapore, a multiethnic Asian city-state. Consecutive patients were approached during their regular outpatient clinic visits.

The eligibility criteria for the study were (i) adult patient (21 years or older) with one or more types of clinically confirmed heart disease (ischemic heart disease, heart failure, heart rhythm disorder, valvular heart disease) and prior hospitalization for heart disease-related conditions; (ii) physically and mentally well enough to participate in a 30-minute interview; and (iii) able to read and communicate in either English or Chinese. The eligibility criterion of prior hospitalization was included to ensure that all of the study participants had experienced a severe health state. The Singapore resident population constitutes of 74% Chinese, 13% Malay, 9% Indian, and 3% others. More than 85% of Indians and Malay are literate in English[14]. Hence, the eligibility criteria for language covers all three major ethnic populations in Singapore. The diagnosis of heart disease was based on internationally accepted criteria as applied by the participants' managing cardiologists.

Informed consent was obtained from all participants. The study was approved by the ethics boards of the respective hospitals.

Valuation interview

The participants were interviewed in quiet areas in clinics. Each participant was interviewed by a trained interviewer in the language of their preference. The interviewer team comprised four bilingual interviewers who could fluently read and speak English and Chinese and had prior experience in conducting patient interviews. All the interviewers were trained in the valuation protocol and had conducted at least five practice interviews.

Each interview comprised two parts: the first involved self-administration of paper forms, and the second involved interviewer-guided computer-based valuation tasks. In the first part, participants self-reported their sociodemographic information and health profiles using the EQ-5D-5L along with a visual analog scale (EQ VAS) and the HeartQoL (heart disease-specific HRQoL instrument); they also reported their functional status using the New York Heart Association (NYHA) classification and the Canadian Cardiovascular Society (CCS) classification of angina. In the second part, participants valued EQ-5D-5L health states using the composite time trade-off (cTTO) module of the EuroQol Portable Valuation Technology (EQ-PVT) software version 1.7 running from a laptop. The interviewers followed a standard script in all interviews. In a previous study, the script and valuation tasks administered using very similar software were tested and shown to be well understood and accepted by local heart disease patients[9]. Protocol compliance was assessed using quality-control criteria developed by the EuroQol Group[15].

Detailed descriptions of the cTTO and the valuation protocol can be found elsewhere[16,17]. Briefly, the objective of the task was to identify the point of preferential indifference between 10 years of life in the described target state followed by death and a shorter life ($x \leq 10$ years) in full health followed by death. With a defined utility value of 1 for full health, the utility value of the target state can be calculated as $x/10$. For states considered to be worse than death, a lead time of 10 years was added to both alternatives to elicit a negative utility value for the state. The utility value of a worse-than-death health state was calculated as (x-10)/10 such that the utility value of each health state is bounded by -1 and 1; 0 represents the value for the "dead" state.

Outcome measures

EQ-5D-5L

EQ-5D-5L is a generic, multi-attribute utility-based instrument. It contains five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and a visual analog scale (EQ VAS) of the overall health status[18]. It describes each dimension at five levels of severity (broadly corresponding to no problem, slight problems, moderate problems, severe problems, and extreme problems). Thus, it can describe 3125 possible health states. This study used validated Singapore English and Chinese language versions of the EQ-5D-5L[19,20]. EQ-5D-5L has been psychometrically validated for a large number of diseases, including heart disease[13].

According to the valuation protocol[16], each participant valued a randomly selected set (called a block) of 10 hypothetical EQ-5D-5L health states. Each block included one very mild health state chosen from five prespecified health states (21111, 12111, 11211, 11121, 11112), the most severe health state (55555), and eight health states chosen from 80 prespecified health states among the remaining 3119 possible health states. Here, the health state "21111" indicated slight problems (level 2 severity) in the first dimension (mobility) and no problems (level 1 severity) in the remaining four dimensions. Other health states were defined similarly. The protocol contained a total of 10 unique blocks, consisting of 86 unique EQ-5D-5L health states.

HeartQoL

HeartQoL is a heart disease-specific HRQoL instrument[21]. It comprises 14 items with four response levels that range from “not bothered” to “bothered a lot”. It provides a global score based on the mean values of the responses. The score ranges from 0 (worst HRQoL) to 3 (best HRQoL). HeartQoL has been validated in more than 22 countries. Our study used its official English and Chinese translated versions.

New York Heart Association (NYHA) and Canadian Cardiovascular Society (CCS) functional classifications

The NYHA and CCS classifications are widely used clinical tools that measure cardiac functional capacity and the severity of exertional angina, respectively[22,23]. They classify patients into classes I, II, III, and IV based on limitations due to symptoms (shortness of breath or angina) at various levels of physical activity. A higher class indicates a worse functional capacity. In this study, participants self-evaluated their NYHA and CCS classes based on structured definitions of these classification systems.

Statistical methods

Sample size

The sample size required to achieve the desired precision of fixed-effect coefficients of health state descriptors in a statistical model estimating utility values using a 20-parameter linear random-effects model was determined. Determination of the sample size was performed using the methodologies proposed by Gandhi et al.[24] for the EQ-5D-5L value set studies. A sample size of 400 participants was required to estimate the coefficients with a precision level (95% confidence interval) of ±0.05, considering 0.05 as the minimum important difference (MID) for EQ-5D-5L utility values. The other parameters required for the sample size calculation—a residual variance of 0.4 and a design effect of 0.5—were estimated from the

EQ-5D-5L value set study in the general Singaporean population[24]. We anticipated that data from 20% of the participants might not be usable (e.g., dropouts or nontraders) and accordingly planned a sample size of 500 participants.

Exclusion of logically inconsistent data

The study data of participants who met one or more of the following criteria were excluded from the analysis: (i) valued the all-worst state ("55555") same as full health; (ii) used less than 10% of the utility scale; and (iii) gave the same values to all ten health states. These response patterns indicate that participants were either inattentive or insensitive to health states. These criteria were suggested by Dewitt et al.[25], who studied these and several other criteria as well as their impact on utility values.

Model development

Various model specifications were explored, and the utility values of the resulting models were examined; only the most appropriate models are reported here. In all the models, we defined the dependent variable as disutility (i.e., 1 – utility value) for a given health state. Two core models (a 20-parameter linear random-effect model and an 8-parameter cross-attribute level effects (CALE) model (a nonlinear random-effect constrained model)) and their variants were extensively tested for performance. Because each participant valued ten health states, participant-specific random-effect intercepts were considered in all the models to account for intraparticipant correlation. We used the regular dummy coding scheme for health state descriptors in the model illustrations and main tables presented in this article because this scheme is widely used. Model results using the backward difference coding scheme are provided in Appendix 1.

The linear model can be presented as follows:

Linear model (Model 1): $y = \alpha + \sum_{l} \sum_{d} \beta_{dl} X_{dl} + v + e$

$$= \alpha + \beta_{MO2}X_{MO2} + \beta_{MO3}X_{MO3} + \beta_{MO4}X_{MO4} + \beta_{MO5}X_{MO5} + \beta_{SC2}X_{SC2} + \beta_{SC3}X_{SC3} + \beta_{SC4}X_{SC4} + \beta_{SC5}X_{SC5} + \beta_{UA2}X_{UA2} + \beta_{UA3}X_{UA3} + \beta_{UA4}X_{UA4} + \beta_{UA5}X_{UA5} + \beta_{PD2}X_{PD2} + \beta_{PD3}X_{PD3} + \beta_{PD4}X_{PD4} + \beta_{PD5}X_{PD5} + \beta_{AD2}X_{AD2} + \beta_{AD3}X_{AD3} + \beta_{AD4}X_{AD5} + v + e$$

where y represents disutility; α, intercept; X_{dl}, fixed-effect indicator variable for the presence of problems on dimension d at level l; β_{dl}, coefficient for the estimated disutility of having problems on dimension d at level l; v, participant-specific random-effect intercept; and e, a heteroscedastic error.

As a preliminary analysis showed nonmonotonicity in coefficients of a few dimensions, each of the β_{dl} coefficients was constrained to have a value greater than or equal to its previous level coefficient β_{dl-1}. The model also left-censored the utility values at -1 because participants could hypothetically value a health state lower than -1. Right censoring at 1 was not considered, as 1 is the theoretical upper bound for the utility value of full health. Furthermore, because the observed variance of the utility values increased with increasing severity of the health states, heteroscedasticity of the error term was modeled using the log link of a linear regression model with an intercept and 20 indicator variables X_{dl}.

An alternative to the linear model is a nonlinear CALE model. It includes a single coefficient per dimension (β_{MO}, β_{SC}, β_{UA}, β_{PD}, and β_{AD}) representing the disutility of having problems at level 5 and one coefficient for each of levels 2, 3, and 4 (L_2, L_3, L_4), all of which are multiplied by the respective dimensional coefficients. Here, L_l (l = 2, 3, 4) should be interpreted as the ratio of disutility at level l to that at level 5 with disutility at level 5 set to 1. The model assumes that these ratios are constant across all dimensions. Empirically, it has been found that the constraint imposed by the multiplicative CALE model makes it more efficient, less susceptible to overfitting, and reduces the risk of nonmonotonicity compared to the linear model.[26]

The model can be presented as follows:

CALE model (Model 2): $y = \alpha + \sum_{l} \left(\sum_{d} \beta_d X_{dl} \right) L_l + v + e$

$$
\begin{aligned}
= \alpha &+ (\beta_{MO} X_{MO2} + \beta_{SC} X_{SC2} + \beta_{UA} X_{UA2} + \beta_{PD} X_{PD2} + \beta_{AD} X_{AD2}) L_2 \\
&+ (\beta_{MO} X_{MO3} + \beta_{SC} X_{SC3} + \beta_{UA} X_{UA3} + \beta_{PD} X_{PD3} + \beta_{AD} X_{AD3}) L_3 \\
&+ (\beta_{MO} X_{MO4} + \beta_{SC} X_{SC4} + \beta_{UA} X_{UA4} + \beta_{PD} X_{PD4} + \beta_{AD} X_{AD4}) L_4 \\
&+ (\beta_{MO} X_{MO5} + \beta_{SC} X_{SC5} + \beta_{UA} X_{UA5} + \beta_{PD} X_{PD5} + \beta_{AD} X_{AD5}) + v + e
\end{aligned}
$$

where α, X_{dl}, v, *and* e are the same as defined for the linear model.

Like the linear model, the CALE model also left-censored the utility values at -1, and the heteroscedasticity of the error term was modeled using a linear regression model with the log link. Due to the nonlinear nature of the CALE model, standard errors were calculated using bootstrap sampling (1000 participant-level samples).

Each core model was also tested with an additional term N45 as a fixed effect. The N45 term was defined as an indicator variable for health states having at least one dimension at either level 4 or 5. It is similar to the N3 term used with the EQ-5D-3L value set in the United Kingdom to represent additional disutility due to severe health conditions in any of the dimensions[27].

Model selection

The predictive accuracy of the models was evaluated in terms of mean absolute error (MAE), root mean square error (RMSE), and Lin's concordance coefficient between the predicted and mean values of the observed values of the health state. Lower MAE and RMSE and higher concordance coefficient indicate better predictive accuracy. The out-of-sample fit was evaluated in cross-validation samples. Cross-validation was performed by fitting the models to a subset of the dataset prepared by excluding one of the 10 blocks of health states and assessing the predictive accuracy in the excluded block[26]. In-sample fit was assessed using the full dataset. If multiple models performed similarly in predictive accuracy, the model using the least number of fixed-effect parameters and achieving the lowest Bayesian information criterion (BIC) was considered the preferred model (i.e., model parsimony).

Rescaling

The predicted utility value for full health ("11111") may not be 1 because of the nonzero intercept in the preferred model. We rescaled all the predicted utility values by dividing them with 1 – intercept to obtain a value of 1 for full health and proportionally adjusted values of the other health states[28].

All models were fitted using the xreg package[29] for R software[30].

Model validation

The preferred model was assessed for the known-groups discriminative ability of its predicted utility values (rescaled). Mean utility values based on the participants' own EQ-5D-5L health states were estimated for each of the NYHA and CCS functional classes as well as for the EQ VAS and HeartQoL global score classes. EQ VAS and HeartQoL global scores were divided into three classes using their first (Q1) and third (Q3) quartiles (class I: ≥Q3, class II: Q1-Q3, class III: ≤Q1). Lower class represents better health or functional capacity. Mean utility values were expected to decrease as class increased. Mean utility values across the classes and differences between two individual classes were compared using analysis of variance (ANOVA) and the two-sample t-test, respectively.

Results

A total of 1166 potential participants were approached for this study. Of these, 64% were willing to participate, and 78% of those met the eligibility criteria. Of the recruited participants (N = 582), 43 were excluded from analysis: three were recruited twice during the quality control review, three did not complete the interview, and 37 provided logically inconsistent values (Supplemental Figure 1). Therefore, 539 participants were included in the analysis. Table 1 shows the sociodemographic and health characteristics of the participants who were included in the analysis. The study recruited older, more men, Malay and Indian participants than are found in the Singapore population[14]. There were no systematic differences in baseline characteristics between participants who were included in the analysis and those who were excluded (Supplemental Table 1).

Table 2 compares the performances of all four models using both the full and cross-validation datasets. Among linear models, the model with the N45 term had lower MAE and RMSE and higher concordance coefficients with both the full and cross-validation datasets compared to the model without the N45 term. The coefficient of the N45 term was also statistically significant (p-value <0.001) in the model (Supplemental Table 2). As in the linear models, inclusion of the N45 term improved the performance of the CALE model, and the coefficient for the N45 term was statistically significant (p-value <0.001) (Supplemental Table 2). However, the predictive accuracy (MAE, RMSE, concordance coefficient) of the CALE model was lower than that of the linear model with the N45 term. Hence, the linear model with the N45 term (Model 1 + N45 term) was selected as the preferred model for developing the value set. Figure 1 shows the predicted utility values obtained using the linear and CALE models with the N45 term as a function of the directly valued health states of the participants. Supplemental Table 3 summarizes results of all models using the backward difference coding scheme.

Table 3 shows coefficients for the preferred model. The largest and smallest utility decrements at level 5 were for mobility and anxiety/depression dimensions, respectively. This was also the case for the CALE model with the N45 term, indicating that mobility and anxiety/depression dimensions have the highest and lowest impacts, respectively, on disutility values (Supplemental Table 2).

Utility values predicted by the preferred model were rescaled by dividing each value by 1 – intercept = 1 – 0.135 = 0.865 for the value set. The value set has values of 1, 0.982 and -0.928 for full health, second-best ('11112') and the worst state ('55555'), respectively. Figure 2 shows the originally predicted and rescaled utility values for the preferred model. An example that demonstrates how to use the coefficients of the preferred model to calculate the utility values can be found in Supplemental Table 4. The utility values for all 3125 health states are available in Appendix 2.

The mean utility values based on the preferred model using the participants' own health states ranked in the expected high-to-low direction for participants with increasing NYHA and CCS classes I to III/IV (p-value <0.001; Table 4). Differences in mean utility value between two consecutive classes were also statistically significant (p-value <0.01), with the majority of mean differences ≥ 0.05 (MID). Similar results were observed for EQ VAS and HeartQoL classes (Table 4).

Discussion

A value set for EQ-5D-5L using heart disease patient preferences was developed according to a standardized international protocol. This is the first EQ-5D-5L value set developed using patient preferences exclusively. For patients with heart disease, it can inform patient-centric economic evaluations and clinical decision-making, and be used to evaluate differences in the outcomes of such decisions derived using societal versus patient preferences.

Mobility is the most relevant EQ-5D dimension, and anxiety/depression the least, in terms of impact on utility value decrement according to preferences of heart disease patients. This differs from the preferences of the Singapore general population, which considers usual activities the most relevant dimension of the EQ-5D-3L value set, and pain/discomfort the least[31]. Such differences are expected. Mobility is essential to life in Singapore, where most people work past official retirement and commute on public transport. Heart disease often significantly limits physical activity such as walking, which explains patients' preference for avoiding this dimension. That anxiety/depression is the least important dimension might be related to mental adaptation to disease, since most heart diseases are chronic. Previous studies have also found differences between the preferences of heart disease patients versus the general population[9-11]. Our results provide granularity regarding the dimensions in which the preferences differ.

We chose a 20-parameter linear model with the N45 term as the final model for developing a value set. We observed logical inconsistencies in the initial version of the model and had to constrain coefficients to achieve monotonicity. Such logical inconsistencies have also been observed in several countries' value sets for the EQ-5D-5L[26,32-34]. This could be due to the complexity of the model, which might predispose patients to overfitting to random variance. Nevertheless, the constrained linear model still provided better predictive accuracy than the constrained 8-parameter CALE models, possibly because the assumption of a constant ratio of level parameters across the dimensions was not fully satisfied in our study data.

There were demographic differences between the study and general populations. The former included more elderly individuals, men, individuals with lower educational levels and had a higher representation of Malay and Indian ethnicities than the latter. These characteristics are known risk factors for heart disease in Singapore[35]. Hence, a higher representation of these characteristics in the patient sample is not unexpected and supports the sample's face validity.

There have been attempts to develop disease-specific value sets for health dimensions affected by specific diseases or their treatments. For example, a cancer-specific QLU-C10D descriptive system has been developed from the EORTC C30 HRQoL measure[36], and its value sets have been or are being developed in several countries[37,38]. We chose the EQ-5D-5L for our study for two reasons: (i) to our knowledge, no heart disease-specific descriptive system that can be used to develop a preference-based value set is currently available, and (ii) the EQ-5D-5L and its former version the EQ-5D-3L have demonstrated acceptable measurement properties in heart disease patients and are widely used for economic evaluation using societal preferences. A value set based on patient preferences for the same descriptive system can facilitate comparisons of economic evaluations based on patients' and societal perspectives.

Our study has some limitations. As it would have been difficult to conduct cognitively demanding valuation tasks among hospitalized patients, we could only approach patients in outpatient clinics. However, we enriched the sample by recruiting those with prior hospitalizations. Recruiting participants with heart disease among the general population would be ideal but is logistically challenging. We believed that recruiting participants from hospital outpatient clinics would help us sample the target population (patients with documented clinical diagnoses based on hospital records) without screening a large number of "generally healthy" candidates. Notably, the profiles of the study patients are comparable to heart disease patients recruited from the Singapore general population in Gandhi et al.[10], which suggests our study findings can be generalizable. The requirement for basic literacy to complete the valuation tasks could have selectively excluded some participants, especially the elderly, who would otherwise have qualified but is common in most valuation studies. We reported preferences of patients in Singapore but patient preferences can vary among countries, possibly due to differences in culture and healthcare systems. The appropriateness of this value set should therefore be evaluated before adoption in other countries.

Conclusions

We have developed a time trade-off-based EQ-5D-5L value set using the preferences of patients with heart disease that enables patient-centric health technology assessments and clinical decision-making for treatment selection.

References

1. Nations U. World Population Ageing 2017-Highlights. Department of Economic and Social Affairs Web site. https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.pdf. Published 2017. Accessed January 2021.
2. Drummond M, Torbica A, Tarricone R. Should health technology assessment be more patient centric? If so, how? 2020.
3. Organization WH. 2015 global survey on health technology assessment by national authorities. In:2015.

4. Helgesson G, Ernstsson O, Åström M, Burström K. Whom should we ask? A systematic literature review of the arguments regarding the most accurate source of information for valuation of health states. *Quality of Life Research.* 2020:1-18.
5. (TLV) TDaPBA. General guidelines for economic evaluations from the Dental and Pharmaceutical Benefits Agency (TLVAR 2017:1). https://www.tlv.se/download/18.467926b615d084471ac3230c/1510316374332/TLVAR_2017_1.pdf. Accessed January 2021.
6. ACE. ACE's Drug Evaluation Methods and Process Guide (Version 2) http://www.ace-hta.gov.sg/public-data/our-process-and-methods/ACE%20methods%20and%20process%20guide%20for%20drug%20evaluation%20(20%20Dec%202019).pdf. Accessed January 2021.
7. ACE. ACE's Medical Technologies Methods and Process Guide (Version 1). http://www.ace-hta.gov.sg/public-data/our-process-and-methods/ACE%20methods%20and%20process%20guide%20for%20medical%20technologies%20evaluation%20(1%20Oct%202018).pdf. Accessed January 2021.
8. WHO. World Health Statistics. WHO Press. http://apps.who.int/iris/bitstream/10665/112738/1/9789240692671_eng.pdf. Published 2014. Accessed January 2021.
9. Gandhi M, San Tan R, Ng R, et al. Comparison of health state values derived from patients and individuals from the general population. *Quality of Life Research.* 2017;26(12):3353-3363.
10. Gandhi M, Thumboo J, Luo N, Wee H-L, Cheung Y-B. Do chronic disease patients value generic health states differently from individuals with no chronic disease? A case of a multicultural Asian population. *Health and quality of life outcomes.* 2015;13(1):1.

11. Pickard AS, Tawk R, Shaw JW. The effect of chronic conditions on stated preferences for health. *The European Journal of Health Economics.* 2013;14(4):697-702.

12. Roth GA, Johnson C, Abajobir A, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *Journal of the American College of Cardiology.* 2017;70(1):1-25.

13. Janssen M, Pickard AS, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of Life Research.* 2013;22(7):1717-1727.

14. DOS. Census of population 2010 statistical release 1. Demographic characteristics, education, language and religion. Singapore Department of Statistics. https://www.singstat.gov.sg/docs/default-source/default-document-library/publications/publications_and_papers/cop2010/census_2010_release1/cop2010sr1.pdf. Published 2010. Accessed January 2021.

15. Ramos-Goñi JM, Oppe M, Slaap B, Busschbach JJ, Stolk E. Quality control process for EQ-5D-5L valuation studies. *Value in Health.* 2017;20(3):466-473.

16. Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health.* 2014;17(4):445-453.

17. Oppe M, Rand-Hendriksen K, Shah K, Ramos-Goñi JM, Luo N. EuroQol protocols for time trade-off valuation of health outcomes. *PharmacoEconomics.* 2016;34(10):993-1004.

18. Foundation ER. EQ-5D-5L User Guide. http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-5L_UserGuide_2015.pdf. Published 2015. Accessed January 2021.

19. Luo N, Wang Y, How CH, Tay EG, Thumboo J, Herdman M. Interpretation and use of the 5-level EQ-5D response labels varied with survey language among Asians in Singapore. *Journal of Clinical Epidemiology.* 2015;68(10):1195-1204.

20. Luo N, Wang Y, How C, et al. Cross-cultural measurement equivalence of the EQ-5D-5L items for English-speaking Asians in Singapore. *Quality of Life Research.* 2015;24(6):1565-1574.

21. Oldridge N, Höfer S, McGee H, et al. The HeartQoL: Part I. Development of a new core health-related quality of life questionnaire for patients with ischemic heart disease. *European journal of preventive cardiology.* 2014;21(1):90-97.

22. Dolgin M, Fox A, Gorlin R, Levin R. New York Heart Association. Criteria Committee. In: *Nomenclature and criteria for diagnosis of diseases of the heart and great vessels.* Litte, Brown & Co.; 1994.

23. Campeau L. Grading of angina pectoris. *Circulation.* 1976;54(3):522-523.

24. Mihir Gandhi YX, Nan Luo, Yin Bun Cheung. Sample size determination for EQ-5D-5L value set studies: an exploration of four approaches. *The 2nd EuroQol Academic Meeting.* 2017.

25. Dewitt B, Fischhoff B, Davis AL, Broomell SB, Roberts MS, Hanmer J. Exclusion Criteria as Measurements II: Effects on Utility Functions. *Medical Decision Making.* 2019;39(6):704-716.

26. Rand-Hendriksen K, Ramos-Goñi JM, Augestad LA, Luo N. Less is more: cross-validation testing of simplified nonlinear regression model specifications for EQ-5D-5L health state values. *Value in Health.* 2017;20(7):945-952.

27. Dolan P. Modeling valuations for EuroQol health states. *Medical Care.* 1997;35(11):1095-1108.

28. Luo N, Liu G, Li M, Guan H, Jin X, Rand-Hendriksen K. Estimating an EQ-5D-5L value set for China. *Value in Health.* 2017;20(4):662-669.

29. Rand K. xreg: Flexible, multi-data.frame maximum likelihood-based regression functions in R. https://github.com/intelligentaccident/xreg. Published 2018. Accessed January 2021.

30. Team RC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org. Published 2016. Accessed January 2021.

31. Luo N, Wang P, Thumboo J, Lim Y-W, Vrijhoef HJ. Valuation of EQ-5D-3L health states in Singapore: modeling of time trade-off values for 80 empirically observed health states. *Pharmacoeconomics.* 2014;32(5):495-507.

32. Xie F, Pullenayegum E, Gaebel K, et al. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Medical care.* 2016;54(1):98.

33. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5 D-5 L value set for E ngland. *Health economics.* 2018;27(1):7-22.

34. Versteegh MM, Vermeulen KM, Evers SM, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. *Value in health.* 2016;19(4):343-352.

35. Picco L, Subramaniam M, Abdin E, Vaingankar JA, Chong SA. Prevalence and correlates of heart disease among adults in Singapore. *Asian journal of psychiatry.* 2016;19:37-43.

36. King MT, Costa D, Aaronson N, et al. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. *Quality of Life Research.* 2016;25(3):625-636.

37. King MT, Viney R, Pickard AS, 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):225-238.

38. Norman R, Mercieca-Bebber R, Rowen D, et al. UK utility weights for the EORTC QLU-C10D. *Health economics.* 2019;28(12):1385-1401.

Table 1: Patient characteristics

Characteristics	Our study N = 539	General population*
Age (years), Mean (SD)	57.8 (11.3)	
21 – 40, n (%)	39 (7.2)	40.9%
41 – 60, n (%)	258 (47.9)	42.5%
>60, n (%)	242 (44.9)	16.6%
Men, n (%)	386 (71.6)	49.1%
Ethnicity, n (%)		
Chinese	309 (57.3)	75.8%
Malay	114 (21.2)	12.1%
Indian	87 (16.1)	8.8%
Others	29 (5.4)	3.2%
Educational level, n (%)		
Primary (6 years) or les	98 (18.2)	8.2%
Secondary (up to 11 years)	250 (46.4)	34.2%
Diploma, university or higher	191 (35.4)	56.6%
Married, n (%)	397 (73.7)	65.8%
Monthly household income <S$4000, n (%)	273 (50.7)	37.5%
Employed, n (%)	304 (56.4)	
Heart disease diagnosis†, n (%)		
Ischemic heart disease	425 (78.9)	
Heart rhythm disorder	156 (28.9)	
Heart failure	147 (27.3)	
Valvular heart disease	94 (17.4)	
Other heart problems	31 (5.8)	
Number of comorbidities, n (%)		
0	60 (11.1)	
1 – 2	211 (39.2)	
3 - 4	223 (41.4)	
>4	45 (8.4)	
NYHA functional classification		
I	249 (46.2)	
II	239 (44.4)	
III-IV	51 (9.5)	
CCS functional classification for angina		
I	423 (78.5)	
II	98 (18.2)	
III-IV	18 (3.3)	
EQ VAS, Mean (SD)	77.2 (15.0)	
HeartQoL global score, Mean (SD)	2.35 (0.55)	

NYHA, New York Heart Association; CCS, Canadian Cardiovascular Society; EQ VAS, EQ visual analog scale; SD, Standard deviation.

* General population o20 – 79 years of age based on Singapore census 2010.[14]

† A patient may have multiple heart disease diagnoses; hence, he/she may be counted under more than one diagnosis.

Table 2: Comparison of model performance

	Linear models		CALE models	
	Model 1	Model 1 + N45 (Preferred model)	Model 2	Model 2 + N45
Mean absolute error				
Full dataset	0.089	0.076	0.097	0.087
Cross-validation dataset	0.103	0.087	0.105	0.093
Root mean square error				
Full dataset	0.109	0.097	0.120	0.110
Cross-validation dataset	0.128	0.110	0.131	0.118
Concordance coefficient				
Full dataset	0.974	0.979	0.968	0.973
Cross-validation dataset	0.964	0.973	0.962	0.969
Number of fixed-effect parameters	21	22	9	10
BIC based on the full dataset	8538.1	8481.7	8478.6	8430.7

Model 1, 20-parameter linear random-effect model. Model 2, 8-parameter cross-attribute level effects model. N45, indicator variable for states with at least one dimension at a severity level of either 4 or 5. See methods section for details.

CALE, cross-attribute level effects; BIC, Bayesian information criterion.

Table 3: Coefficients (SEs) of the preferred model (Model 1 + N45) based on the full dataset

	Coefficient	SE
Intercept	0.135	0.030
MO2	0.052	0.023
MO3	0.114	0.024
MO4	0.230	0.027
MO5	0.354	0.026
SC2	0.106	0.023
SC3	0.213	0.026
SC4	0.285	0.026
SC5	0.342	0.023
UA2	0.062	0.024
UA3	0.139	0.024
UA4	0.201	0.027
UA5	0.221	0.026
PD2	0.048	0.021
PD3	0.048	0.027
PD4	0.276	0.025
PD5	0.296	0.028
AD2	0.016	0.024
AD3	0.114	0.026
AD4	0.142	0.025
AD5	0.210	0.023
N45	0.246	0.031

Model 1, 20-parameter linear random-effect model (see methods section for details). MO2 to MO5, SC2 to SC5, UA2 to UA5, PD2 to PD5, and AD2 to AD5 represent indicator variables for severity levels 2 to 5 with reference to level 1 for mobility, self-care, usual activities, pain/discomfort, and anxiety/depression dimensions, respectively. N45 represents an indicator variable for health states with at least one dimension at level 4 or 5. SE, Standard error.

Table 4: Known-group validity of rescaled utility values derived from the preferred model (Model 1 + N45)

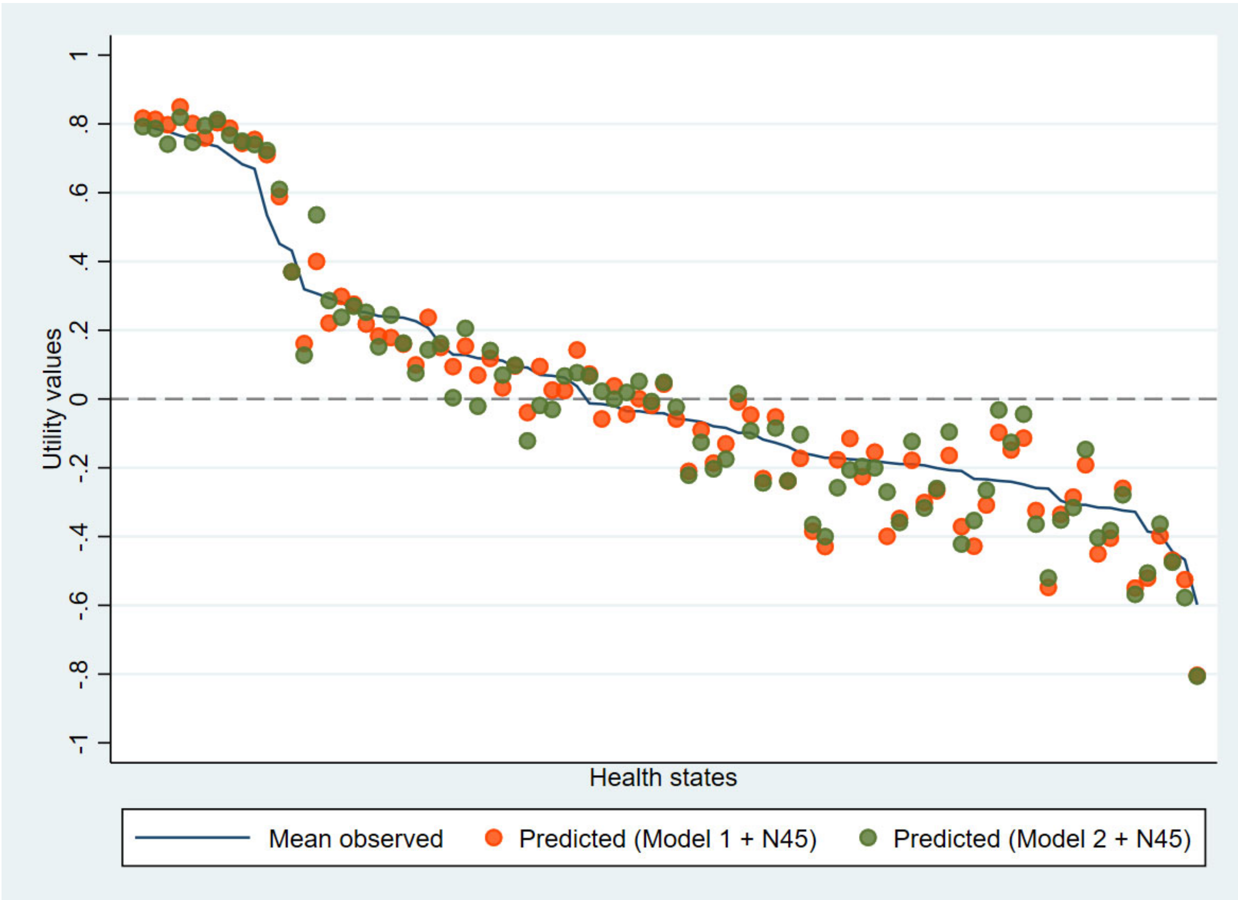
	NYHA		CCS		EQ VAS		HeartQoL Global	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Class I	249	0.961 (0.086)	423	0.948 (0.103)	152	0.972 (0.052)	154	0.985 (0.034)
Class II	239	0.922 (0.121)	98	0.845 (0.239)	187	0.928 (0.159)	186	0.951 (0.073)
Class III/IV	51	0.676 (0.372)	18	0.582 (0.401)	200	0.865 (0.219)	199	0.832 (0.247)
ANOVA *p-value*		<0.001		<0.001		<0.001		<0.001
		Mean (95% CI)		Mean (95% CI)		Mean (95% CI)		Mean (95% CI)
Diff (I – II)		0.039 ** (0.020 – 0.058)		0.102 ** (0.072 – 0.133)		0.043 * (0.017 – 0.070)		0.034 ** (0.022 – 0.047)
Diff (II – III/IV)		0.246 ** (0.188 – 0.304)		0.264 ** (0.127 – 0.400)		0.063 * (0.024 – 0.101)		0.119 ** (0.082 – 0.156)
Diff (I – III/IV)		0.285 ** (0.233 – 0.337)		0.366 ** (0.305 – 0.427)		0.106 ** (0.071 – 0.142)		0.153 ** (0.114 – 0.193)

NYHA, New York Heart Association functional classification; CCS, Canadian Cardiovascular Society functional classification for angina; EQ VAS, EQ Visual analog scale; Diff, Difference; SD, Standard deviation; CI, Confidence interval; ANOVA, Analysis of variance.

Classes I, II and III/IV for EQ-VAS and HeartQoL Global represent ≥ the third quartile (Q3), first quartile (Q1) to third quartile (Q3), and ≤ first quartile (Q1) of their values, respectively.

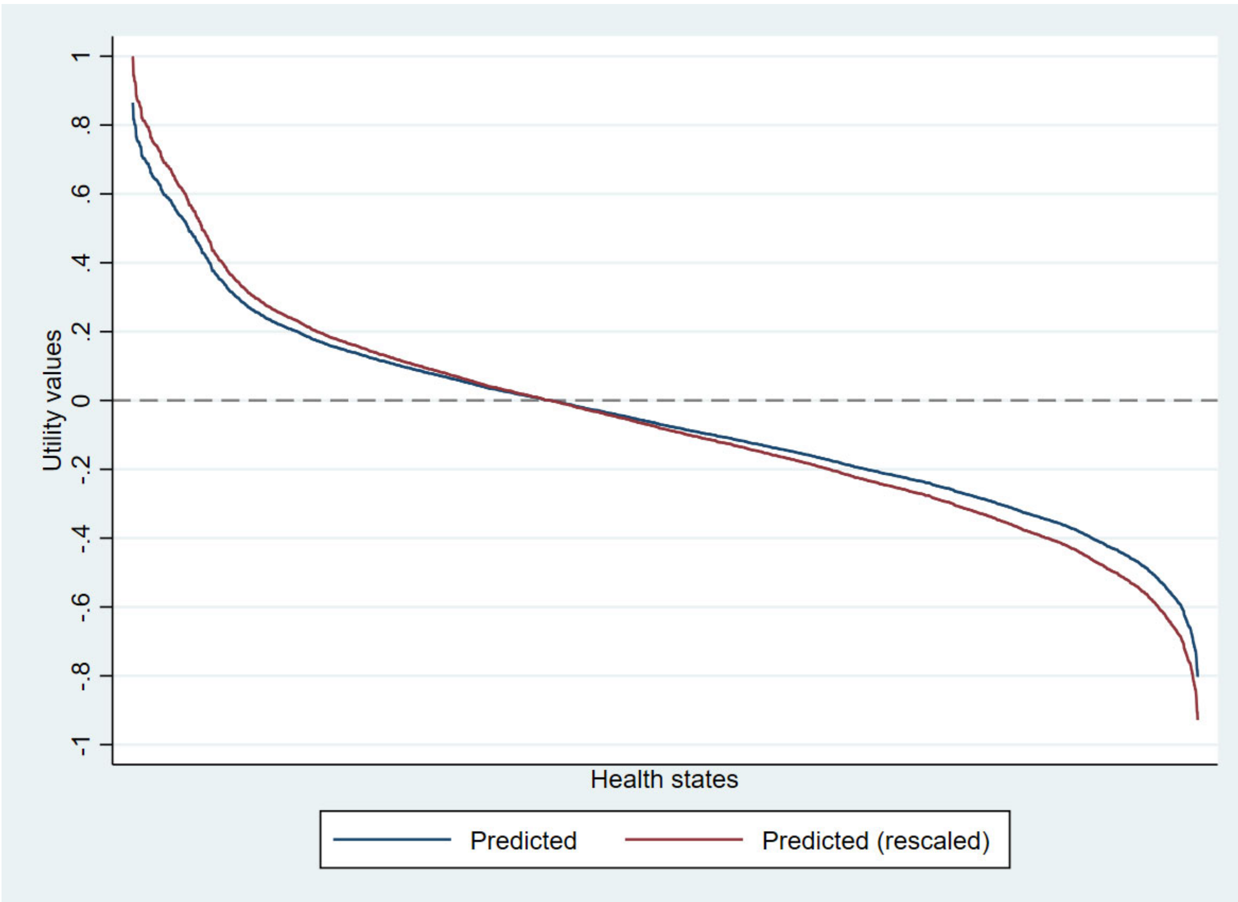
Two-sample t-test *p-value*: * < 0.01, ** <0.001.

Figure 1: Observed and predicted utility values for directly valued health states



Utility values are sorted based on observed mean values.

Figure 2: Predicted utility values using the preferred model (Model 1 + N45) for all possible health states



Utility values are sorted based on predicted values.

Appendix 1

Supplemental Table 1: Comparison of the characteristics of patients included in and excluded from the analysis

Characteristics	Included patients N = 539	Excluded patients N = 37*	p-value
Age (years), Mean (SD)	57.8 (11.3)	56.8 (14.1)	0.636
21 – 40, n (%)	39 (7.2)	6 (16.2)	0.155
41 – 60, n (%)	258 (47.9)	15 (40.5)	
>60, n (%)	242 (44.9)	16 (43.2)	
Men, n (%)	386 (71.6)	30 (81.1)	0.258
Ethnicity, n (%)			0.231
Chinese	309 (57.3)	19 (51.4)	
Malay	114 (21.2)	5 (13.5)	
Indian	87 (16.1)	10 (27.0)	
Others	29 (5.4)	3 (8.1)	
Educational level, n (%)			0.242
Primary (6 years) or les	98 (18.2)	3 (8.1)	
Secondary (up to 11 years)	250 (46.4)	21 (56.8)	
Diploma, university or higher	191 (35.4)	13 (35.1)	
Married, n (%)	397 (73.7)	26 (70.3)	0.701
Monthly household income <S$4000, n (%)	273 (50.7)	19 (51.4)	>0.999
Employed, n (%)	304 (56.4)	20 (54.1)	0.864
Heart disease diagnosis†, n (%)			
Ischemic heart disease	425 (78.9)	31 (83.8)	0.675
Heart failure	147 (27.3)	10 (27.0)	>0.999
Heart rhythm disorder	156 (28.9)	11 (29.7)	>0.999
Valvular heart disease	94 (17.4)	3 (8.1)	0.176
Other heart problems	31 (5.8)	2 (5.4)	>0.999
Number of comorbidities, n (%)			0.760
0	60 (11.1)	2 (5.4)	
1 - 2	211 (39.2)	17 (46.0)	
3 - 4	223 (41.4)	15 (40.5)	
>4	45 (8.4)	3 (8.1)	
NYHA functional classification			0.878
I	249 (46.2)	19 (51.4)	
II	239 (44.4)	15 (40.5)	
III-IV	51 (9.5)	3 (8.1)	
CCS functional classification for angina			>0.999
I	423 (78.5)	30 (81.1)	
II	98 (18.2)	6 (16.2)	
III-IV	18 (3.3)	1 (2.7)	
EQ VAS, Mean (SD)	77.2 (15.0)	77.6 (13.0)	0.873

HeartQoL Global score, Mean (SD)	2.35 (0.55)	2.34 (0.59)	0.952

* Complete data on patient characteristics for 6 patients who did not complete the interview or did not meet eligibility criteria were not available. These patients were not included in the "Excluded patients" column.

† A patient may have multiple heart diagnoses and hence may be counted under more than diagnosis.

NYHA, New York Heart Association; CCS, Canadian Cardiovascular Society; EQ VAS, EQ Visual analog scale; SD, Standard deviation.

Supplemental Table 2: Summary of models using regular dummies for health state descriptors

	Linear models		CALE models	
	Model 1	Model 1 + N45	Model 2	Model 2 + N45
	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
		Nonlinear parameters		
Intercept	-	-	0.128 (0.017)	0.135 (0.017)
MO	-	-	0.403 (0.016)	0.356 (0.016)
SC	-	-	0.368 (0.016)	0.316 (0.017)
UA	-	-	0.310 (0.016)	0.237 (0.018)
PD	-	-	0.363 (0.015)	0.331 (0.015)
AD	-	-	0.270 (0.015)	0.207 (0.017)
L1	-	-	0	0
L2	-	-	0.303 (0.023)	0.220 (0.029)
L3	-	-	0.524 (0.019)	0.433 (0.026)
L4	-	-	0.885 (0.017)	0.834 (0.020)
L5	-	-	1	1
N45	-	-	-	0.225 (0.026)
		Linear parameters		
Intercept	0.123 (0.030)	0.135 (0.030)	0.128 (0.017)	0.135 (0.017)
MO1	0	0	0	0
MO2	0.112 (0.022)	0.052 (0.023)	0.122 (0.011)	0.078 (0.011)
MO3	0.189 (0.024)	0.114 (0.024)	0.211 (0.010)	0.154 (0.011)
MO4	0.315 (0.027)	0.230 (0.027)	0.356 (0.013)	0.297 (0.013)
MO5	0.412 (0.026)	0.354 (0.026)	0.403 (0.007)	0.356 (0.008)
SC1	0	0	0	0
SC2	0.155 (0.022)	0.106 (0.023)	0.112 (0.011)	0.069 (0.012)
SC3	0.248 (0.026)	0.213 (0.026)	0.193 (0.009)	0.137 (0.010)
SC4	0.337 (0.026)	0.285 (0.026)	0.326 (0.011)	0.264 (0.011)
SC5	0.389 (0.023)	0.342 (0.023)	0.368 (0.007)	0.316 (0.007)
UA1	0	0	0	0
UA2	0.117 (0.023)	0.062 (0.024)	0.094 (0.010)	0.052 (0.010)
UA3	0.206 (0.024)	0.139 (0.024)	0.162 (0.007)	0.102 (0.007)
UA4	0.285 (0.027)	0.201 (0.027)	0.274 (0.009)	0.197 (0.009)
UA5	0.285 (0.026)	0.221 (0.026)	0.310 (0.006)	0.237 (0.005)
PD1	0	0	0	0
PD2	0.094 (0.021)	0.048 (0.021)	0.110 (0.009)	0.073 (0.011)
PD3	0.101 (0.028)	0.048 (0.027)	0.190 (0.010)	0.143 (0.010)
PD4	0.316 (0.025)	0.276 (0.025)	0.322 (0.011)	0.276 (0.012)
PD5	0.356 (0.028)	0.296 (0.028)	0.363 (0.006)	0.331 (0.007)
AD1	0	0	0	0
AD2	0.046 (0.024)	0.016 (0.024)	0.082 (0.007)	0.045 (0.008)
AD3	0.164 (0.026)	0.114 (0.026)	0.141 (0.008)	0.090 (0.008)

AD4	0.229 (0.025)	0.142 (0.025)	0.239 (0.008)	0.173 (0.008)
AD5	0.268 (0.023)	0.210 (0.023)	0.270 (0.005)	0.207 (0.005)
N45	-	0.246 (0.031)	-	0.225 (0.026)

Model 1, 20-parameter linear random-effect model (see methods section for details). MO1 to MO5, SC1 to SC5, UA1 to UA5, PD1 to PD5, and AD1 to AD5 represent indicator variables for severity levels 1 to 5 for the mobility, self-care, usual activities, pain/discomfort, and anxiety/depression dimensions, respectively.

Model 2, 8-parameter cross-attribute level effects model (see methods section for details). MO, SC, UA, PD, and AD represent indicator variables for the mobility, self-care, usual activities, pain/discomfort, and anxiety/depression dimensions, respectively. L1 to L5 represent indicator variables for severity levels 1 to 5, respectively.

N45 represents an indicator variable for health states with at least one dimension at level 4 or 5.

CALE, cross-attribute level effects model; SE, standard error. SEs for cross-attribute level effects models were based on 1000 bootstrap samples (participant-level sampling).

Supplemental Table 3: Summary of models using backward difference dummies for health state descriptors

	Linear models		CALE models	
	Model 1	Model 1 + N45	Model 2	Model 2 + N45
	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Intercept	0.123 (0.030)	0.135 (0.030)	0.128 (0.017)	0.135 (0.017)
MO1	0	0	0	0
MO2 – MO1	0.112 (0.022)	0.052 (0.023)	0.122 (0.011)	0.078 (0.011)
MO3 – MO2	0.077 (0.024)	0.062 (0.024)	0.089 (0.010)	0.076 (0.011)
MO4 – MO3	0.126 (0.027)	0.116 (0.027)	0.145 (0.013)	0.143 (0.013)
MO5 – MO4	0.097 (0.026)	0.124 (0.026)	0.046 (0.007)	0.059 (0.008)
SC1	0	0	0	0
SC2 – SC1	0.155 (0.022)	0.106 (0.023)	0.112 (0.011)	0.069 (0.012)
SC3 – SC2	0.094 (0.026)	0.106 (0.026)	0.081 (0.009)	0.068 (0.010)
SC4 – SC3	0.088 (0.026)	0.072 (0.026)	0.133 (0.011)	0.128 (0.011)
SC5 – SC4	0.052 (0.023)	0.057 (0.023)	0.042 (0.007)	0.053 (0.007)
UA1	0	0	0	0
UA2 – UA1	0.117 (0.023)	0.062 (0.024)	0.094 (0.010)	0.052 (0.010)
UA3 – UA2	0.089 (0.024)	0.077 (0.024)	0.068 (0.007)	0.051 (0.007)
UA4 – UA3	0.079 (0.027)	0.062 (0.027)	0.112 (0.009)	0.096 (0.009)
UA5 – UA4	0.000 (0.026)	0.020 (0.026)	0.036 (0.006)	0.040 (0.005)
PD1	0	0	0	0
PD2 – PD1	0.094 (0.021)	0.048 (0.021)	0.110 (0.009)	0.073 (0.011)
PD3 – PD2	0.007 (0.028)	0.000 (0.027)	0.080 (0.010)	0.071 (0.010)
PD4 – PD3	0.216 (0.025)	0.228 (0.025)	0.131 (0.011)	0.134 (0.012)
PD5 – PD4	0.040 (0.028)	0.019 (0.028)	0.042 (0.006)	0.055 (0.007)
AD1	0	0	0	0
AD2 – AD1	0.046 (0.024)	0.016 (0.024)	0.082 (0.007)	0.045 (0.008)
AD3 – AD2	0.118 (0.026)	0.098 (0.026)	0.060 (0.008)	0.044 (0.008)
AD4 – AD3	0.065 (0.025)	0.029 (0.025)	0.097 (0.008)	0.083 (0.008)
AD5 – AD4	0.039 (0.023)	0.068 (0.023)	0.031 (0.005)	0.035 (0.005)
N45	-	0.246 (0.031)	-	0.225 (0.026)

Model 1, 20-parameter linear random-effect model (see methods section for details). MO1 to MO5, SC1 to SC5, UA1 to UA5, PD1 to PD5, and AD1 to AD5 represent indicator variables for severity levels 1 to 5 for the mobility, self-care, usual activities, pain/discomfort, and anxiety/depression dimensions, respectively.

Model 2, 8-parameter cross-attribute level effects model (see methods section for details). MO, SC, UA, PD, and AD represent indicator variables for the mobility, self-care, usual activities, pain/discomfort, and anxiety/depression dimensions, respectively. L1 to L5 represent indicator variables for severity levels 1 to 5, respectively.

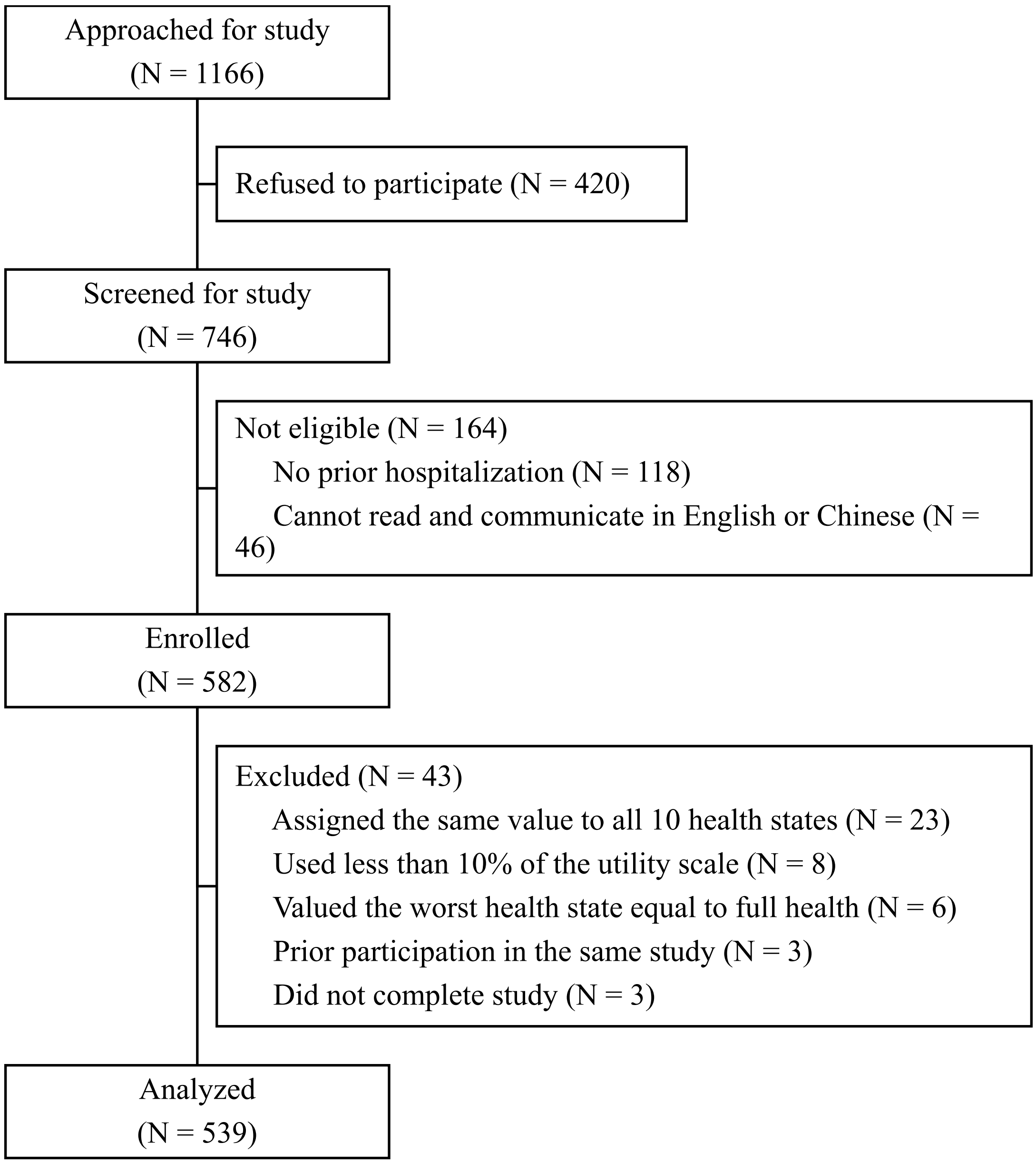
N45 represents an indicator variable for health states with at least one dimension at level 4 or 5.

CALE, cross-attribute level effects model; SE, standard error. SEs for cross-attribute level effects models were based on 1000 bootstrap samples (participant-level sampling).

Supplemental Table 4: Illustration showing use of the estimated coefficients from the preferred model (Model 1 + N45) to calculate utility values for EQ-5D-5L health states

	Coefficient	Health state '12221'	Health state '24315'
Intercept	0.135	0.135	0.135
Mobility level 1 (MO1)	0	0	
Mobility level 2 (MO2)	0.052		0.052
Mobility level 3 (MO3)	0.114		
Mobility level 4 (MO4)	0.230		
Mobility level 5 (MO5)	0.354		
Self-care level 1 (SC1)	0		
Self-care level 2 (SC2)	0.106	0.106	
Self-care level 3 (SC3)	0.213		
Self-care level 4 (SC4)	0.285		0.285
Self-care level 5 (SC5)	0.342		
Usual activities level 1 (UA1)	0		
Usual activities level 2 (UA2)	0.062	0.062	
Usual activities level 3 (UA3)	0.139		0.139
Usual activities level 4 (UA4)	0.201		
Usual activities level 5 (UA5)	0.221		
Pain/discomfort level 1 (PD1)	0		0
Pain/discomfort level 2 (PD2)	0.048	0.048	
Pain/discomfort level 3 (PD3)	0.048		
Pain/discomfort level 4 (PD4)	0.276		
Pain/discomfort level 5 (PD5)	0.296		
Anxiety/depression level 1 (AD1)	0	0	
Anxiety/depression level 2 (AD2)	0.016		
Anxiety/depression level 3 (AD3)	0.114		
Anxiety/depression level 4 (AD4)	0.142		
Anxiety/depression level 5 (AD5)	0.210		0.210
Any dimension at level 4 or 5 (N45)	0.246		0.246
Sum of coefficients (Disutility)		0.351	1.067
Utility (1 – Disutility)		0.649	-0.067
Rescaled utility (Utility/0.865)		0.750	-0.077

Supplemental Figure 1: Participant flow chart



Appendix 2

Excel file listing the rescaled utility values of all 3125 health states.