Article:

Bae, E.-Y., Cho, J., Moon, J. et al. (2026) Deriving a Korean SF-6Dv2 value set using a discrete choice experiment with duration. PharmacoEconomics. ISSN: 1170-7690

https://doi.org/10.1007/s40273-026-01629-w

Deriving a Korean SF-6Dv2 Value Set Using a Discrete Choice Experiment with Duration

Author name(s)

Eun-Young Bae
College of Pharmacy, Gyeonsang National University, Jinju, South Korea
E-mail: eybae@gnu.ac.kr
ORCID: 0000-0001-8304-5578

Jahyun Cho
Graduate School of Public Health, Seoul National University, Seoul, South Korea
E-mail: jahyun04@snu.ac.kr
ORCID: 0000-0001-6008-3774

Jooho Moon
Department of Economics, Seoul National University, Seoul, South Korea
E-mail: joo0783@snu.ac.kr

Brendan Mulhern
Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, NSW, Australia
E-mail: Brendan.mulhern@chere.uts.edu.au

Donna Rowen
Sheffield Centre for Health and Related Research (SCHARR), School of Medicine and Population Health, University of Sheffield, Sheffield, UK
E-mail: d.rowen@sheffield.ac.uk

Jakob Bue Bjorner
IQVIA Inc, Copenhagen, Denmark; Department of Public Health, University of Copenhagen, Denmark; National Centre for the Working Environment, Copenhagen, Denmark
E-mail: jbjorner@qualitymetric.com
ORCID No. 0000-0001-7033-8224

Min-Woo Jo
Department of preventive medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea
E-mail: mdjominwoo@gmail.com
ORCID No. 0000-0002-4574-1318

Corresponding author

Eun-Young Bae, Ph.D.

College of Pharmacy, Gyeonsang National University, Jinju, South Korea

E-mail: eybae@gnu.ac.kr

Tel.: +82-55-772-2431

ORCID: 0000-0001-8304-5578

Statements and Declarations

Funding: This study was supported by the National Research Foundation of Korea grant funded by the Korea government (MSIT). (No. 2022R1A2C10122521361382116530103)

Competing Interests: Brendan Mulhern, Donna Rowen, and Jakob Bjorner were involved in the original development of the SF-6Dv2 classification system. Eun-Young Bae and Min-Woo Jo were involved in the development of the Korean translation of SF-6Dv2.

Data availability: The predicted values for all possible health states, along with their associated standard errors, are available from the corresponding author upon reasonable request.

Ethics Approval: Ethical approval was obtained from the Institutional Review Board of Gyeongsang National University (no. GIRB-A24-NY-0088).

Consent to Participate: All participants provided informed consent electronically after reviewing an IRB-approved information sheet prior to their participation.

Consent to Publication: Not applicable.

Code availability: Available on request from the corresponding author.

Acknowledgements

This research was funded by National Research Foundation of Korea grant, and we are grateful for their financial support, which made this study possible

Abstract

Objectives: This study aimed to derive a Korean value set for the SF-6Dv2.

Methods: Following an adapted international protocol, an online discrete choice experiment with duration (DCE_{TTO}) was conducted among 3,800 Korean adults representative of the general population. The experimental design included three components: choice sets from a Core Module, choice sets pairing commonly occurring health states (Common Module), and triplet choice sets with death options (Triplet Module). Preferences were analyzed using conditional and mixed logit models

Results: Model 9 (based on the Core Module) was selected as the preferred model for its superior logical consistency, including monotonicity of coefficients and avoidance of counterintuitive results. The pain dimension had the largest utility decrement at its worst level (−0.496), followed by physical functioning (−0.480), mental health (−0.391), social functioning (−0.329), vitality (−0.178), and role limitations (−0.138). The utility for the worst possible health state (556555) was estimated at −1.012, with 31.8% of all states valued as worse than dead. Significant preference heterogeneity was observed across dimensions, indicating substantial individual-level variation in health state valuations. While including death comparisons (Triplet Module) produced less negative worst-state utilities, it was not recommended as the primary value set due to its non-efficient experimental design and logical monotonicity issues necessitating greater level collapsing.

Conclusion: We derived the first Korean SF-6Dv2 value set. This enables quality-adjusted life year calculations from both SF-36v2 and SF-6Dv2 data, supporting economic evaluations in Korea.

Key points

- This study establishes the first Korean population-based value set for the SF-6Dv2, enabling health state utility estimation from both standalone SF-6Dv2 and existing SF-36v2 datasets to support health technology assessment in Korea.
- While pain and physical functioning are the most influential dimensions, the Korean population assigns higher relative importance to mental health and social functioning compared to other East Asian populations, reflecting distinct cultural preference patterns.
- The study ensured methodological rigor by following the international DCE_{TTO} protocol and implementing strategic refinements, such as pre-task training and linguistic clarifications, to minimize logical inconsistencies in health state valuations.

1. Introduction

The quality-adjusted life year (QALY) is a standard measure for quantifying health benefits by multiplying the utility level assigned to a given health state by the life expectancy in that state. For QALY to be used as an appropriate measure in resource allocation decisions, consistent and valid utility weights must be applied. Many countries recommend generic preference-based measures (GPBMs), such as the EQ-5D, SF-6D, and HUI, as utility measurement instruments [1]. However, different tools often generate different index scores for the same patient groups, as the health state descriptive systems and the methods used to derive value sets differ between instruments [1].

Korea's pharmacoeconomic evaluation guidelines recommend using utility values measured from GPBMs that reflect the preferences of the Korean general population [2]. However, to date, such value sets have only been established for the EQ-5D-3L and EQ-5D-5L [3]. Since each GPBM captures different health dimensions, or describes dimensions in different ways, and no single tool covers all health aspects, it is advantageous to develop country-specific value sets for a range of these instruments [4].

SF-6Dv2 is a GPBM derived from the SF-36v2 [5]. The SF-36 is one of the most extensively used descriptive systems reflecting health-related quality of life and is widely adopted in clinical research. In Korea, all versions of SF-36 have been translated and validated, but no Korean value sets for the SF-6D or SF-6Dv2 currently exist. The development of a Korean SF-6Dv2 value set would enable utility values for QALY calculations to be derived from clinical research data.

SF-6Dv2 was developed to address the weaknesses of version 1 [5]. The index score range of SF-6Dv1 differed significantly from other GPBMs, potentially due to issues with the descriptive system and differences in evaluation methods (two-stage standard gamble vs. time trade-off (TTO)) [1, 6]. Like SF-6Dv1, SF-6Dv2 has six dimensions (physical functioning (PF), role limitation (RL), social functioning (SF), pain (PA), mental health (MH), and vitality (VT)). However, there were some changes in the selected items comprising the descriptive system [5]. All dimensions except PA have five levels, while PA has six levels [1, 6]. In contrast with the previous version, the new version uses a discrete choice experiment (DCE) with duration as an additional attribute for valuation (DCE_{TTO}). By including duration, DCE_{TTO} offers the advantage that the derived scores can be used directly as weights for calculating QALY without the need for other external anchors [7-11].

To date, value sets for SF-6Dv2 using DCE_{TTO} have been published for the UK [6], China [12], Australia [13], Quebec [14], Iran [15], New Zealand [16], and Japan [17]. While the UK, Australia, Quebec, Iran, and Japan used DCE_{TTO} and applied an international protocol (with some country-specific adaptations) New Zealand utilized an adaptive DCE approach called PAPRIKA (Potentially All Pairwise RanKings of all possible Alternatives) [16]. China applied both TTO and DCE_{TTO}, ultimately adopting the results derived from TTO [12, 18], and the most recent Quebec study employed a hybrid modeling approach across multiple methods [19]. Variations were also observed in the survey administration; China and Iran conducted face-to-face interviews, whereas the other countries collected data via online surveys.

When comparing anchored DCE_{TTO} coefficients, the relative importance of health dimensions varies across countries, with significant differences particularly in the relative importance of PF and MH. The UK and Australia showed similar patterns, whereas China and Japan demonstrated distinctly different preferences in the relative importance of these dimensions.

It remains unclear why the importance assigned to each dimension varies. This variance may be attributed to factors such as cultural differences, varying survey methods (face-to-face interviews vs. unsupervised online surveys) [20, 21], model discrepancies, or the interrelation of these factors. A recent study from Japan [17], which followed the same protocol as the UK and Australian studies, strongly suggests that these differences may be attributed to cultural backgrounds. Interestingly, Japan showed different patterns in the ranking of dimensional importance and the extent of utility decrements compared to China. This contrast highlights that each country requires its own value sets to accurately reflect local population preferences [12, 17].

The necessity of country-specific value sets has been well-established through international research. Comprehensive reviews of EQ-5D-5L value sets have demonstrated substantial cross-national variation in health state valuations, even when using identical methods [22, 23]. These differences have been attributed to cultural values, societal norms, healthcare system characteristics, and population health experiences, providing empirical justification for requiring locally derived value sets in health technology assessment guidelines.

This study aimed to derive a value set for the SF-6Dv2 reflecting the preferences of the Korean general population for use in health technology assessment. By employing the international DCE_{TTO} protocol refined for the U.S. study, this approach enables comparison with international value sets while generating a locally relevant Korean value set.

2. Methods

2.1. Survey Design

This study applied a web-based survey (mobile access allowed) to align with previous studies. This approach facilitates large scale data collection in a cost-efficient way. Given 75,000 possible combinations of attribute levels and duration, and over 2.8 billion possible comparison pairs, a full factorial design was impractical. Moreover, because preferences for health states and their durations are not independent, their interaction must be estimated in DCE_{TTO} [7, 8]. Therefore, this study adopted the fractional factorial design developed for the US value set study, which built on the original UK design [24]. While the core DCE_{TTO} approach remains identical to the UK design, the US protocol's primary refinement is the inclusion of a supplementary module for common health states to improve estimation precision.

The survey was structured into three sequential modules (Core and two Supplementary Modules) to serve distinct methodological purposes (see Figure 1 for the survey flow diagram).

Core Module: This module served as the primary source of preference data for value set estimation. It consisted of 304 binary choice sets organized into 38 blocks of 8 choice sets each. The 304 choice sets were generated using NGene's Fedorov algorithm, from a restricted full factorial design that excluded specific two-way combinations of dimension levels unlikely to occur in practice: RL1-PA6, RL2-PA6, PF1-PA6, PF2-PA6, VT1-PA6, and VT1-MH5. The number of blocks was selected to provide sufficient replication of each choice set across the sample for stable parameter estimation while maintaining a manageable respondent burden per person.

Supplementary Module 1 (Common): Administered after the Core Module, this module consisted of 76 choice sets (two per block) featuring health states commonly reported in the general population. Based on the 200 most frequent health states identified in a previous survey [25], all possible pairwise combinations (200x199/2) were generated and assigned 16 duration pairs from four levels (1, 4, 7, and 10 years). The Fedorov algorithm selected 76 D-efficient choice sets from this candidate pool. This module was included to improve estimation precision for the most frequently reported health states and assess preference stability.

Supplementary Module 2 (Triplet): Administered as the final task, this module included 76 triplet choice sets (two per block) where immediate death was presented as a third option alongside two health states with duration. Choice sets were restricted to severe health states, with dimension levels limited to the three most severe: levels 4-6 for Pain and levels 3-5 for all other dimensions. Using NGene's Fedorov algorithm, 76 choice sets were selected from pairs of these severe states combined with duration levels, with immediate death then appended as the third option. Respondents were asked to identify both the best and worst option among the three alternatives. This module provides information on states considered worse than dead (WTD), supplementing the duration-based anchoring of the DCE_{TTO}.

Quality control (QC) measures: Each block also included one consistency check question that repeated practice question 2, which was presented as the 9th question after the 8 Core Module choice sets. This was implemented to assess response consistency, but was not part of the experimental design itself. Practice question 2 was selected for repetition, rather than one of the eight Core Module choice sets, because its large utility differential makes it a useful indicator of basic task comprehension and preference stability.

2.2. Participant Recruitment and Screening

This study used a web-based survey accessible via mobile devices, which facilitates large-scale, cost-efficient data collection. Participants aged 19 and older (the legal adult age in Korea) were recruited through Macromill Embrain's general population panel [26]. We targeted 3,800 respondents with quotas for gender, age, and region to ensure at least 100 responses per choice set, aligning with previous studies [6, 13, 24]. All participants received an IRB-approved information sheet, provided informed consent before proceeding, and were compensated with points from the survey company upon completion.

2.3. Survey Procedure

Through demographically balanced randomization, participants were assigned to one of 38 blocks. They first completed the SF-6Dv2 questionnaires to evaluate their health status and then proceeded to the training exercises.

The first training task was a ranking exercise in which respondents were presented with one dimension at a time, with its levels shown in random order. They were asked to arrange these levels from the best to the worst. If respondents ordered the levels incorrectly, they were shown the correct ordering to help them understand both the severity hierarchy and the descriptive distinctions between levels. The second training task involved a pairwise comparison between two different levels within the same dimension. Corrections were provided if a respondent selected a strictly inferior state. In the pilot survey, this training session was found to improve coefficient consistency compared to previous surveys that lacked such exercises [25].

Upon completing the training, respondents received detailed instructions on the composition of the choice sets-specifically how health states are paired with durations-and how to perform the choice tasks. To optimize comprehension, the presentation format (e.g., bold-underlined text and upfront duration placement) was selected based on a preliminary qualitative study [27]. To ensure participants were comfortable with this layout and the task requirements, two practice questions followed. The first featured a dominant-dominated pair, where Health State A had higher functioning levels than Health State B in four dimensions and equal levels in the others, with identical durations. The second practice question presented a choice set with a substantial utility difference based on the UK SF-6Dv2 value set [6], although it was not designed as a strictly dominant-dominated pair.

After the practice questions, respondents were asked if they had compared all dimensions; those who answered “no” received a message encouraging them to consider all dimensions in subsequent tasks. Such attention reminders have been shown to increase the proportion of respondents who engage in full attribute attendance [28] , which is particularly important given the complexity of the SF-6Dv2 descriptive system [28-32].

Next, respondents were presented with eight choice sets from their allocated block (Core Module). To mitigate order effects, the sequence of the six dimensions (excluding duration) was randomized between blocks; however, for any given respondent, this attribute order remained consistent across all choice sets to minimize cognitive burden [33, 34]. All alternatives were unlabeled and consistently presented as ‘Health State A’ and ‘Health State B’ to ensure respondents focused solely on the health state descriptions. Following the first eight sets, practice question 2 was repeated to check the choice consistency.

The respondents then completed two choice sets composed of common health states (Common Module) and two triplet choice sets that included an “immediate death” alternative (Triplet Module). To ensure respondents focused on the zero-duration aspect of the death option, we included an additional explanation stating that all health states would conclude in peaceful, painless death, thereby preventing other factors associated with the process of dying from influencing their decisions.

After completing all choice sets, respondents answered questions assessing their comprehension of the tasks and the difficulty of making choices. They also identified which dimensions most and least influenced their decisions, and rated their level of consideration for each dimension. Finally, respondents provided demographic information.

To ensure data quality, a minimum viewing time of 5 seconds per choice set (10 seconds for instructions) was enforced. We also identified potentially problematic response patterns, such as consistently selecting the same position (left or right) or choosing the dominated option in practice question 1. Although respondents exhibiting these patterns were retained in the primary analysis to maintain sample representativeness, we conducted sensitivity analyses excluding them to verify the robustness of our findings (see Table 2 in Appendix 1). Inconsistent responses on the repeated question were not classified as irrational, acknowledging that preferences may evolve as respondents gain experience with the choice tasks [35-40].

2.4. Pilot Survey

Multiple pilot surveys were conducted prior to the main study. An analysis of data from a preliminary survey with an identical design identified unexpected level-disordering within the RL dimension [25]. Interviews with a convenience sample indicated that the Korean translation of RL levels was interpreted inconsistently. Specifically, some respondents perceived the phrase "accomplish less than you would like" as having more leisure time rather than as a health-related restriction. To address this ambiguity and ensure a health-related interpretation, we added the phrase "due to health issues" to each RL level, despite the disadvantage of lengthening the phrases. Furthermore, training exercises, including level-ordering and pairwise comparisons, were implemented to enhance respondents' understanding of the distinctions between severity levels (see Section 2.3 and Appendix 2). A subsequent pilot study with 380 participants confirmed that these exercises effectively mitigated the disordering issues.

Finally, after launching the main survey, we temporarily paused data collection once the first 380 responses (10% of the target sample) were received. This interim analysis was conducted to verify question comprehension and data quality before proceeding with the remainder of the survey.

2.5. Ethical Considerations

This study received approval from the Gyeongsang National University's Institutional Review Board (no. GIRB-A24-NY-0088).

2.6. Analysis

2.6.1. Modelling framework

We used conditional logit regression to model respondents' choices. The DCE_{TTO} model expresses individual *i*'s utility for health profile *j* (x_{ij}) as a multiplicative function of health states *x* and life years *t* [7]:

$$\mu_{ij} = \beta_1 t_{ij} + \beta_2' x_{ij} \cdot t_{ij} + \varepsilon_{ij}$$

The health profile *x* is represented by dummy variables indicating each dimension-level combination, with level 1 (no problems) serving as the reference category for all dimensions. The coefficient β_1 represents the value of living one year in perfect health (expected positive), while $\beta_2{}'$ represents the disutility of living in state x_j for one year (expected negative) compared to the baseline state with the best functioning level in all dimensions [6]. The error term ε_{ij} is random.

In the DCE_{TTO} framework, value for χ_j is anchored to the standard 0 – 1 scale (where dead = 0, full health = 1) by dividing each dimension-level coefficient ($\beta_2{}'$) by the duration coefficient (β_1). This ratio ($\beta_2{}'/\beta_1$) represents the utility decrements associated with each dimension-level relative to full health [6, 7]. This normalization ensures that the resulting value set is directly comparable to TTO-derived utility.

$$V_j{}^{DCE} = 1 + \frac{\widehat{\beta_2{}'}}{\widehat{\beta_1}} x_j$$

2.6.2. Model Specifications and Selection Criteria

We tested multiple model specifications starting with the Core Module as the primary analysis, followed by models incorporating Supplementary Modules. Each specification included baseline models (with duration and dimension-level × duration interactions) and extended models (incorporating WORST × duration interactions, where WORST is a dummy variable indicating health states with at least one dimension at its lowest level). It is hypothesized that the presence of the 'worst' functioning level in any dimension results in an additional, non-linear decrement in utility, consistent with findings from the UK [6], Australia [13], and Iranian SF-6Dv2 valuation studies [15]. Constrained versions were also developed by merging adjacent levels where coefficient ordering was reversed. Specifically, where an increase in severity level was associated with an increase rather than a decrease in utility, the disordered level was merged with its adjacent level and a common coefficient was estimated for the merged pair.

In total, 16 conditional logit models were estimated across four data combinations (Core only; Core + Common; Core + Triplet; Core + Common + Triplet) and two specifications (with or without WORST interaction term), with constrained versions developed for each. Mixed logit models were additionally estimated to characterize preference heterogeneity and are reported as supplementary evidence. The full model inventory is summarized in Table 1 in Appendix 1.

Candidate models were evaluated based on three criteria to identify the preferred value set: (1) the extent of level collapsing required to achieve monotonicity, with fewer collapses preferred to preserve the granularity and sensitivity of the estimates; (2) the direction of the WORST interaction term, where models with positive coefficients were excluded to avoid counterintuitive results—specifically, instances where health states with at least one dimension at its most severe level could receive higher utility than states without such severity; and (3) statistical fit within each data combination, with the BIC serving as the primary criterion to prioritize parsimony. To formally assess whether data from different modules could be pooled without adjustment, a preference parameter equality test was conducted following the approach of Swait and Louviere [41]. The evaluation results for all 16 models are summarized in Table 1 in Appendix 1.

2.6.3. Preference Heterogeneity and Sensitivity Analyses

To assess preference heterogeneity, a mixed logit model was estimated in which all dimension-level × duration interaction terms and the duration coefficients were specified as random parameters following a normal distribution, assumed to be independent across parameters [42]. Significant standard deviations were interpreted as evidence of preference heterogeneity. Results are provided in Tables 5 and 6 in Appendix 1.

Sensitivity analyses were performed by comparing results across completion time groups, by excluding respondents who failed QC checks, and by education level (see Table 2 and 3 in Appendix 1). All quality indicators—including dominant option selection, balanced A/B choices, completion time, and consistency on the repeated choice set—were used as diagnostic tools to characterize data quality and inform sensitivity analyses rather than as exclusion criteria. The preferred value set was therefore estimated on the full analytic sample. The reporting of this study follows the Discrete Choice Experiment Reporting Standards (DIRECT) checklist, and the detailed description of each item is provided in Appendix 3.

3. Results

3.1. Sample Characteristics and Data Quality

A total of 3,800 respondents completed the survey between October and November 2024 and were included in the analysis (Table 1). The gender and age distribution matched the Korean general population due to recruitment quotas [43]. However, approximately 80% of respondents had a college education or higher, typical of online samples.

Over half of respondents (53.61%, N=2,037) reported at least one long-term health condition. On the SF-6Dv2, 6.92% (N=263) reported full health (111111), and the most frequent impaired state was 112211 (3.29%, N=125). No respondents reported the worst possible health state (556555) (Table 1).

The median completion time was 15.1 minutes (mean 22.1, SD 42.2), with fewer than 0.02% completing in under

5 minutes.

Several response quality issues were identified: 2.47% selected the dominated option in rationality checks, 35.68% provided inconsistent responses to the repeated question (practice question 2 and question 9), 1.3% always selected the same position (straight-liners).

When rating task difficulty, most respondents (42.58%) characterized understanding the choice sets as "moderate", while 38.24% reported difficulty making choices and 23.84% found it easy. Regarding which dimension most influenced their decisions, respondents most frequently cited PA (36.5%), followed by PF (20.8%), duration (16.3%), and MH (12.0%) (Table 1).

3.2. Model Estimation

We estimated 16 models using different dataset combinations and model specifications (Table 2 and 3). Models 1–8 are unconstrained models estimated across different datasets, with or without a WORST interaction term (Table 2), and Models 9-16 are constrained versions of Models 1-8 respectively (Table 3). Anchored coefficients for candidate value sets are shown in Table 4.

Sensitivity analyses results excluding respondents with QC issues (dominated selections, straight-liners, inconsistent responses) showed similar patterns to the main analysis (Table 2 in Appendix 1). Additional sensitivity analyses by education level confirmed that the preferred value set was robust to these sources of variation. Inconsistent respondents and those with a high school education or below displayed fewer disordered dimension-level coefficients than their respective counterparts (Table 3 in Appendix 1).

The analysis revealed ordering reversals in several dimension-level coefficients Ordering reversals were observed in all models, most frequently in the RL and VT dimensions. The number of reversed level pairs ranged from two (Models 1, 4) to four (Model 2, 7), with the full details summarized in Table 2.

Model 1 (Core) and Model 4 (Core + Common) had the fewest disordering coefficients among all specifications. However, the constrained version of Model 4 (Model 12) produced positive WORST interaction coefficients, which could lead to counterintuitive results—worse health states yielding higher utilities. The same issue occurred in Model 16 (Core + Common + Triplet).

Model 12 (Core + Common) produced the widest utility range, with the worst state valued at−1.019, while Model 15 (Core + Common + Triplet) produced the least negative worst-state utility (−0.646). Including Triplet Module (triplet choice sets with the death option) decreased the estimated impact of all dimensions except PA. Consequently, combined dataset analyses produced less negative worst-state utilities (e.g., −0.804 in Model 13) compared to Core Module alone (−1.012 in Model 2) (Table 4 in Appendix 1). This pattern of less negative worst-state utilities in combined analyses aligns with findings from international studies [6, 13, 14, 17, 20].

The proportion of health states worse than dead varied substantially across models, ranging from 16.4% (Model 14, Core + Triplet) to 36.9% (Model 11, Core + Common) (Table 4 in Appendix 1). Model fit was similar with or without WORST interaction terms (Table 3).

3.3. Preferred model selection and value set

Model 9 (Core Module, without WORST interaction) was selected for the Korean SF-6Dv2 value set based on its superior logical properties: it required the fewest level collapses and avoided counterintuitive WORST interaction effects that observed in Models 12 and 16. Model 9 showed a comparable model fit to Model 10 within the same dataset combination.

The selected model showed PA as having the largest utility range from the best to the worst level (−0.496), followed by PF (−0.480), MH (−0.391), SF (−0.329), VT (−0.178), and RL (−0.138) (Table 4). The steepest single-level utility decline occurred between PA levels 4 and 5 (from −0.149 to −0.469), with minimal additional decline from level 5 to 6 (−0.469 to −0.496). PF showed the next steepest decline between levels 4 and 5 (from −0.210 to −0.480). SF and MH demonstrated relatively linear utility decreases across severity levels. Notably, utility decrements for mild impairment (level 2) were negligible across most dimensions.

Model 14 (Core + Triplet), which incorporates direct death-comparison tasks, is presented as a supplementary value set for sensitivity analyses, as these tasks provide additional anchoring information at the dead = 0 boundary. However, a preference parameter equality test rejected parameter equality between the two modules (LR = 1752.87, df = 13, $p < 0.001$), indicating that parameter estimates across two modules are not directly comparable. Its use should therefore be interpreted with this limitation in mind.

The distribution of predicted utilities across all 18,750 SF-6Dv2 health states for both models is shown in Figure 2. Model 9 produced a wider utility range (worst state: −1.012; WTD: 31.7%), whereas Model 14 produced a narrower range (worst state: −0.751; WTD: 16.4%). With the exception of RL2, PA2, and VT2, Model 14 yielded less negative coefficients across all dimension levels compared to Model 9.

3.4. Heterogeneity Assessment

Mixed logit analyses were conducted across all model specifications; results are summarized in Tables 5 and 6 in Appendix. The mixed logit models corresponding to the preferred conditional logit specification (Core Module only, without WORST interaction term; Appendix Table 5, Model 1) revealed significant standard deviations for severe impairment levels (PF5, RL4-5, PA5-6, VT5, SF5, MH5), mild impairment levels in several dimensions (PF2, RL2, PA2, SF2, MH2), and the duration coefficient. These findings indicate substantial heterogeneity in the relative importance respondents assigned to health dimension levels and survival duration.

Notwithstanding this heterogeneity, the conditional logit model was selected for the recommended value set, as it provides mean population preferences suitable for health technology assessment and ensures comparability with international SF-6Dv2 valuation studies.

3.5. International Comparison

To enable meaningful comparisons across countries with different utility ranges (e.g., Korea's worst state: −1.012 vs. China's: −0.277), we calculated Relative Attribute Importance (RAI). This represents each dimension's utility range as a proportion of the total value set range (Table 5).

Korean preferences revealed distinctive patterns. PA was identified as the most important dimension (RAI: 24.7%), closely followed by PF (23.9%) and MH (19.4%). While Korea, Japan, and China all prioritized PA and PF—reflecting a shared East Asian pattern—Korea assigned substantially higher importance to MH (19.4%) and SF (16.4%) than Japan (MH: 12.4%, SF: 6.7%) or China (MH: 10.5%, SF: 8.5%). In contrast, Western countries (UK, Australia, New Zealand) and Iran ranked MH as the second most important dimension after PA, but assigned considerably lower importance to PF (UK: 11.8%, Australia: 13.2%) compared to East Asian nations.

Regarding specific health states, Korea assigned the most negative value to the worst state (556555: −1.012), whereas China's value was the least negative (−0.277). For moderate impairment (333333), Korea (0.345) and Japan (0.415) yielded substantially lower values than the UK (0.684) and Australia (0.702).

4. Discussion

This study derived a Korean value set for the SF-6Dv2 using the DCE_{TTO}, eliciting preferences from a general population sample. Among the candidate specifications, Model 9 (Core Module only) was selected as the preferred value set based on three jointly applied criteria: the extent of level collapsing required to achieve monotonicity, the direction of the WORST interaction term, and statistical fit within each dataset.

Level collapsing assigns identical utility decrements to merged severity levels, eliminating the value set's ability to distinguish between the health states they describe; in Model 9, this affects only the distinction between RL levels 1/2 and 4/5, which is considered a minor limitation given the small magnitude of the reversals involved.

The selected Model 9 showed PA as having the greatest impact on utility, followed by PF, MH, SF, VT, and RL. The worst health state (556555) utility was estimated at −1.012, with 31.8% of all possible health states valued as worse than dead.

Model 14, which incorporates Triplet Module data produced a narrower utility range and retains information from explicit death comparisons. However, its use as the primary value set is not recommended: the Triplet Module was not derived using an efficient experimental design, and a preference parameter equality test confirmed systematically different parameter estimates between the two modules, indicating that pooling without scale adjustment is not fully justified. The observed difference between Model 9 and Model 14 may reflect differences in response scale arising from the greater cognitive demand of death-comparison tasks, framing effects whereby health states are evaluated differently when immediate death is presented as an explicit alternative, or a combination of both. Model 14 is therefore recommended for sensitivity analyses, with this limitation explicitly acknowledged.

The mixed logit analysis revealed substantial preference heterogeneity, particularly for severe health states and duration. The observed preference heterogeneity underscores the inherent uncertainty surrounding utility estimates. While the conditional logit-based value set provides the population mean utility estimates required for cost-utility analysis, the heterogeneity findings imply that this mean may not adequately represent preferences in specific subpopulations.

4.1. Divergence from the EQ-5D-5L: Shifting Preferences or Methodological Artifacts?

The Korean EQ-5D-5L valuation study, which used composite time-trade off (cTTO) in face-to-face interviews [44], estimated the worst health state utility was at −0.066-substantially less negative than our estimate of -1.012. Particularly noteworthy is the difference in mental health valuation. In the current SF-6Dv2 study, mental health emerged as the third most important dimension with substantial utility decrements, whereas the Korean EQ-5D-5L study revealed relatively small decrements for the anxiety/depression dimension [44]. Given that both instruments assess similar mental health concepts (anxiety and depression), this substantial difference warrants further investigation to understand whether it reflects methodological differences, temporal changes in societal preferences, or other factors.

Nevertheless, the validity of our findings is supported by the concordance between indirectly derived and directly stated preferences: when respondents were asked to identify the single most important dimension in their decisions, the ranking based on response frequencies aligned perfectly with the dimension importance derived from the choice data analysis, suggesting that our study appropriately captured Korean population preferences. However, the substantial differences in the utility ranges compared to EQ-5D-5L pose challenges for cross-instrument consistency in decision-making processes. Regarding cost-effectiveness analysis, a wider utility range produces larger QALY gains for interventions preventing transitions into severe states, compared with analyses using narrower value sets such as the Korean EQ-5D-5L tariff. Analysts should therefore exercise caution when comparing cost-effectiveness ratios across studies using different instruments or valuation methods, as the choice of value set can materially affect conclusions.

4.2. Impact of Valuation Methodology on Utility Ranges: DCE_{TTO} vs. cTTO

Most SF-6Dv2 valuation studies have used DCE_{TTO}, except China and New Zealand [6, 12, 13, 16, 17]. The Chinese study employed both DCE_{TTO} and cTTO, finding substantial differences: utility ranges of 1 to −0.277 with cTTO versus 1 to −0.535 with DCE_{TTO} [12].

The two methodologies differ structurally in how duration is presented. In DCE_{TTO}, duration is one attribute alongside multiple health dimensions, while cTTO involves direct trade-offs between duration and health states. This structural difference likely contributes to the divergent results. Our findings are consistent with this interpretation: although pooling across modules was not fully justified, incorporating data from direct death-comparison tasks produced a narrower utility range than the Core Module alone. Recent research suggests that DCE with duration incorporating non-linear time preferences produces results similar to cTTO following the EuroQol protocol [45].

4.3. Cross-cultural preference patterns and Korean distinctiveness

Cross-national comparisons revealed notable differences between cultural zones, suggesting that cultural values significantly influence health state preferences. While countries in the Anglo-American (UK, Australia, New Zealand) and East Asian (Korea, China, Japan) zones exhibited substantial differences in dimension importance rankings, all agreed that PA was the most critical dimension. However, the second most crucial dimension for Korea, China, and Japan was PF, whereas it was MH for the UK, Australia, and New Zealand. These differences in dimension ordering between Anglo-American and other countries have been confirmed in previous research systematically reviewing SF-6D studies [46]. This pattern is also evident in the EQ-5D-5L value sets, with East Asian countries generally placing higher importance on physical health dimensions [22].

Nevertheless, variations existed even among countries belonging to the same cultural zone. Notably, Korea demonstrated a unique pattern in valuing MH and SF dimensions, displaying consistent and distinct utility decrements for both, whereas China and Japan exhibited much lower utility decrements for these dimensions despite being in the same East Asian zone. While SF was not regarded as particularly important in most other countries, Korea consistently demonstrated high utility decrements for both SF and MH. This distinctive pattern likely reflects Korea's cultural emphasis on social relationships and heightened awareness of mental health issues, particularly given Korea's high suicide rates [47, 48]. These differences in value systems across cultural zones and between countries within the same zone highlight the importance of developing country-specific value sets.

4.4. Methodological Refinements and Study Limitations

A key methodological strength of this study lies in the successful mitigation of respondent confusion regarding dimension-level ordering. Following the international protocol for Korean SF-6Dv2 valuation, our previous work identified significant directional confusion regarding the RL dimension [25]. Because the DCE format inevitably condenses the full descriptive system, critical contextual information can be lost, leading to misinterpretations of functional levels. To address this, we implemented two strategic refinements: (1) a level-ranking exercise prior to the trade-off tasks to reinforce the understanding of level directionality, and (2) the addition of the phrase "due to health issues" to enhance contextual clarity. These measures substantially reduced the directional confusion observed in the previous survey. Our findings demonstrate that such issues can be effectively resolved through preparatory steps, offering important implications for future valuation studies conducted in diverse linguistic and cultural contexts.

Despite these refinements, several limitations should be acknowledged. First, more than one-third of respondents provided inconsistent answers on the repeated choice set, which might initially raise concerns regarding data quality. It should be noted, however, that this item was a practice question designed primarily for training purposes—not part of the formal experimental design—and presented again only after respondents had completed eight core tasks. This interval between the two presentations, during which respondents accumulated task experience, means that any divergence may reflect task-learning effects, context effects, fatigue, or random noise, rather than a persistent tendency toward inattentive or careless responding. Two observations further mitigate these concerns: the preferred value set remained robust even after exclusion of these individuals, and, contrary to expectations, inconsistent respondents did not exhibit poor response quality; if anything, they showed superior monotonicity relative to their consistent counterparts. In this regard, the repeated practice item is better understood as a limited diagnostic of basic task comprehension and response stability rather than a definitive quality-control measure. Taken together, these findings suggest that inconsistent respondents should not be treated as simply inattentive or irrational.

Beyond the repeated item, a broader data quality consideration concerns response engagement. While a 5-second minimum viewing time was implemented to exclude clearly careless responses, this threshold alone does not guarantee thoughtful consideration of all attributes. However, sensitivity analyses comparing model estimates across different total completion time groups yielded consistent results (Table 2 in Appendix 1), supporting the robustness of our findings.

Second, we did not employ additional probabilistic modeling to formally partition potential random or disengaged responses from the main analysis. Although such an approach could further isolate data signals from noise, the stability of our results across various quality-control subgroups suggests that their impact on the final value set would be minimal. Future research may explore these modeling specifications to further refine response reliability in large-scale online panels.

Third, the online survey format resulted in an overrepresentation of highly educated individuals, which may limit the generalizability of our findings. Nevertheless, sensitivity analysis by education level (college-educated vs. others) showed consistent preference patterns and dimension rankings between groups (see Table 3 in Appendix 1), suggesting no material bias in the estimated preference structure.

Finally, a further limitation concerns the pooled model analyses. Although a preference parameter equality test indicated systematic differences between the Core and Triplet Modules, the sources of these differences cannot be statistically disentangled: observed differences may reflect scale or task-format effects, potential structural shifts in preferences, or order-related effects such as learning, fatigue, or contextual adaptation, given that the

Common and Triplet Modules were administered after the Core Module. These constraints reinforce the rationale for selecting the Core Module-only model as the preferred value set.

5. Conclusion

Based on an adapted international protocol, we successfully developed Korean value sets for SF-6Dv2 health states. Similar to findings in other countries, the DCE_{TTO} method performed relatively well in Korea; although disordering was observed in some dimension levels, the coefficients generally maintained logically consistent ordering. In line with other East Asian countries, PA and PF were identified as the most influential dimensions. However, the Korean value set is characterized by notably greater impact of the MH and SF dimensions compared to other East Asian value sets. With the development of this Korean SF-6Dv2 value set, deriving utility values from the standalone SF-6Dv2 questionnaire [49], as well as health-related quality of life data measured through SF-36v2, is now possible, enabling QALY calculations for economic evaluations.

Author contributions

Author Contributions: Concept and design: EB, JC, MJ, BM, DR, and JB. Data collection: EB, JC, and JM. Statistical analysis: JM and EB. Analysis and interpretation of data: EB, JC, MJ, BM, DR, and JB. Writing – original draft: EB. Writing – review & editing: EB, JC, MJ, BM, DR, and JB. Funding acquisition: EB. Supervision: EB. All authors read and approved the final manuscript.

References

Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. 2nd ed. Oxford: Oxford University Press; 2017.

2. Bae EY, Hong J, Bae S, Hahn S, An H, Hwang EJ, et al. Korean guidelines for pharmacoeconomic evaluations: updates in the third version. Appl Health Econ Health Policy. 2022;**20**(4):467-77.
3. Lee TJ, Bae EY, Hong J, et al. The development of second version of economic evaluation guidelines for pharmaceuticals in Korea. Seoul: Seoul National University·Gyeonsang National University·Health Insurance Review and Assessment Service; 2020.
4. Olsen JA, Misajon R. A conceptual map of health-related quality of life dimensions: key lessons for a new instrument. Qual Life Res. 2020;**29**(3):733-43.
5. Brazier JE, Mulhern BJ, Bjorner JB, Rowen D, Alonso J, Vilagut G. et al. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. Med Care. 2020;**58**(6):557-65.
6. Mulhern BJ, Bansback N, Norman R, Brazier J. Valuing the SF-6Dv2 classification system in the United Kingdom using a discrete-choice experiment with duration. Med Care. 2020;**58**(6):566-73.
7. 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.
8. Mulhern B, Norman R, Shah K, Bansback N, Longworth L, Viney R. How should discrete choice experiments with duration choice sets be presented for the valuation of health states? Med Decis Making. 2018;**38**(3):306-18.
9. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. Health Econ. 2018;**27**(1):7-22.
10. 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.
11. Wang H, Rowen DL, Brazier JE, Jiang L. Discrete choice experiments in health state valuation: a systematic review of progress and new trends. Appl Health Econ Health Policy. 2023;**21**(3):405-18.
12. Wu J, Xie S, He X, Chen G, Bai G, Feng D, et al. Valuation of SF-6Dv2 health states in China using time trade-off and discrete-choice experiment with a duration dimension. Pharmacoeconomics. 2021;**39**(5):521-35.
13. Mulhern B, Norman R, Brazier J. Valuing SF-6Dv2 in Australia using an international protocol. Pharmacoeconomics. 2021;**39**(10):1151-62.
14. Ameri H, Poder TG. Valuing SF-6Dv2 using a discrete choice experiment in a general population in Quebec, Canada. Int J Health Policy Manag. 2024;**13**:8404.

15. Daroudi R, Zeraati H, Poder TG, et al. Valuing the SF-6Dv2 in the capital of Iran using a discrete choice experiment with duration. Qual Life Res. 2024;**33**(7):1853-63.
16. Sullivan T, McCarty G, Ombler F, et al. Creating an SF-6Dv2 social value set for New Zealand. Soc Sci Med. 2024;**354**:117073.
17. Shiroiwa T, Yamamoto Y, Murata T, et al. Valuation survey for SF-6Dv2 in Japan based on the international protocol. Qual Life Res. 2024. https://doi.org/10.1007/s11136-024-03830-w
18. Xie S, Wu J, He X, et al. Do discrete choice experiments approaches perform better than time trade-off in eliciting health state utilities? Evidence from SF-6Dv2 in China. Value Health. 2020;**23**(10):1391-9.
19. Poder TG, Ameri H. A new SF-6Dv2 value set based on a hybrid model using SG, cTTO, and DCE data. Soc Sci Med. 2025;**366:117632**. DOI: 10.1016/j.socscimed.2024.117632
20. Shiroiwa T, Fukuda T. Are web-based valuation surveys for preference-based measures as reliable as face-to-face surveys? TTO, DCE and DCE with duration. Appl Health Econ Health Policy. 2024;**22**(3):391-400.
21. Rowen D, Brazier J, Keetharuth A, Tsuchiya A, Mukuria C. Comparison of modes of administration and alternative formats for eliciting societal preferences for burden of illness. Appl Health Econ Health Policy. 2016;**14**(1):89-104.
22. Roudijk B, Ludwig K, Devlin N. EQ-5D-5L value set summaries. In: Devlin N, Roudijk B, Ludwig K, editors. Value sets for EQ-5D-5L: a compendium, comparative review & user guide, Cham: Springer;2022. p.55-212.
23. Olsen JA, Lamu AN, Cairns J. In search of a common currency: a comparison of seven EQ-5D-5L value sets. Health Econ. 2018;**27**(1):39-49.
24. Rendas-Baum R, et al. Development of SF-6Dv2 health utility weights in the United States. Value Health. 2022;**25**(12):S328.
25. Bae EY, Jo MW, Cho J. SF-6Dv2 valuation study in South Korea: uncovering level inconsistencies and next steps for research. Value Health. 2024;**27**(12):S536.
26. Embrain. Infra. [Website]. https://embrain.com/kor/company/infra. Accessed 14 Aug 2025.
27. Cho, J., et al., *Optimizing presentation formats for SF-6Dv2 health status valuation: a qualitative study using cognitive interviews.* Qual Life Res. 2025;**34**(10):3017-27.
28. Kassie GT, Zeleke F, Birhanu MY, Scarpa R. Would a simple attention-reminder in discrete choice experiments affect heuristics, preferences, and willingness to pay for livestock market facilities? PLoS One. 2022;**17**(7):e0270917. DOI: 10.1371/journal.pone.0270917
29. Doherty E, Hobbins A, Whitehurst DGT, O'Neill C. An exploration on attribute non-attendance using discrete choice experiment data from the Irish EQ-5D-5L national valuation study. Pharmacoecon Open. 2021;**5**(2):237-44.
30. Flynn TN, Bilger M, Malhotra C, Finkelstein EA. Are efficient designs used in discrete choice

experiments too difficult for some respondents? A case study eliciting preferences for end-of-life care. Pharmacoeconomics. 2016;**34**(3):273-84.
31. Heidenreich S, Watson V, Ryan M, Phimister E. Decision heuristic or preference? Attribute non-attendance in discrete choice problems. Health Econ. 2018;**27**(1):157-71.
32. Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. Effect of level overlap and color coding on attribute non-attendance in discrete choice experiments. Value Health. 2018;**21**(7):767-71.
33. Mulhern B, Norman R, Lorgelly P, et al. Is dimension order important when valuing health states using discrete choice experiments including duration? Pharmacoeconomics. 2017;**35**(4):439-51.
34. Norman R, Kemmler G, Viney R, et al. Order of presentation of dimensions does not systematically bias utility weights from a discrete choice experiment. Value Health. 2016;**19**(8):1033-8.
35. Ryan M, Watson V, Entwistle V. Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. Health Econ. 2009;**18**(3):321-36.
36. Skedgel C, Wailoo A, Akehurst R. Societal preferences for distributive justice in the allocation of health care resources: a latent class discrete choice experiment. Med Decis Making. 2015;**35**(1):94-105.
37. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ. 2006;**15**(8):797-811.
38. Miguel FS, Ryan M, Amaya-Amaya M. 'Irrational' stated preferences: a quantitative and qualitative investigation. Health Econ. 2005;**14**(3):307-22.
39. Tervonen T, Schmidt-Ott T, Marsh K, et al. Assessing rationality in discrete choice experiments in health: an investigation into the use of dominance tests. Value Health. 2018;**21**(10):1192-7.
40. McFadden D. Rationality for economists? J Risk Uncertain. 1999;**19**(1-3):73-105*rameter in the estimation and comparison of multinomial logit models.* Journal of Marketing Research, 1993. **30**(3): p. 305-14.
41. Swait JL, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. Journal of Marketing Research 1993;**30**(3):305-14.
42. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;**21**(2):145-72.
43. Statistics Korea (KOSTAT). Korean Statistical Information Service (KOSIS) 2025 [Website]. https://kosis.kr/. Accessed 28 Feb 2026.
44. Kim SH, Ahn J, Ock M, et al. The EQ-5D-5L valuation study in Korea. Qual Life Res. 2016;**25**(7):1845-52.
45. Roudijk B, Jonker MF, Bailey H, Pullenayegum E. A direct comparison between discrete choice

with duration and composite time trade-off methods: do they produce similar results? Value Health. 2024;**27**(9):1280-8

46. Wang L, Poder TG. A systematic review of SF-6D health state valuation studies. J Med Econ. 2023;**26**(1):584-93.
47. Park YK. The quality of life and interpersonal relationships among Korean adolescents. Korean J Edu Psychol. 2008;**22**(4):801-36.
48. Park SL, Kim Y, Moon J, Hwang H. High suicide rate: what is and is not the problem – South Korea's suicide rate from a national integration perspective. National Future Strategy Insight. 2021;22.
49. QualityMetric. SF-6Dv2 health utility survey. [Website]. https://www.qualitymetric.com/health-surveys/sf-6dv2-health-utility-survey/. Accessed 16 Oct 2025..

Fig 1. Survey procedure

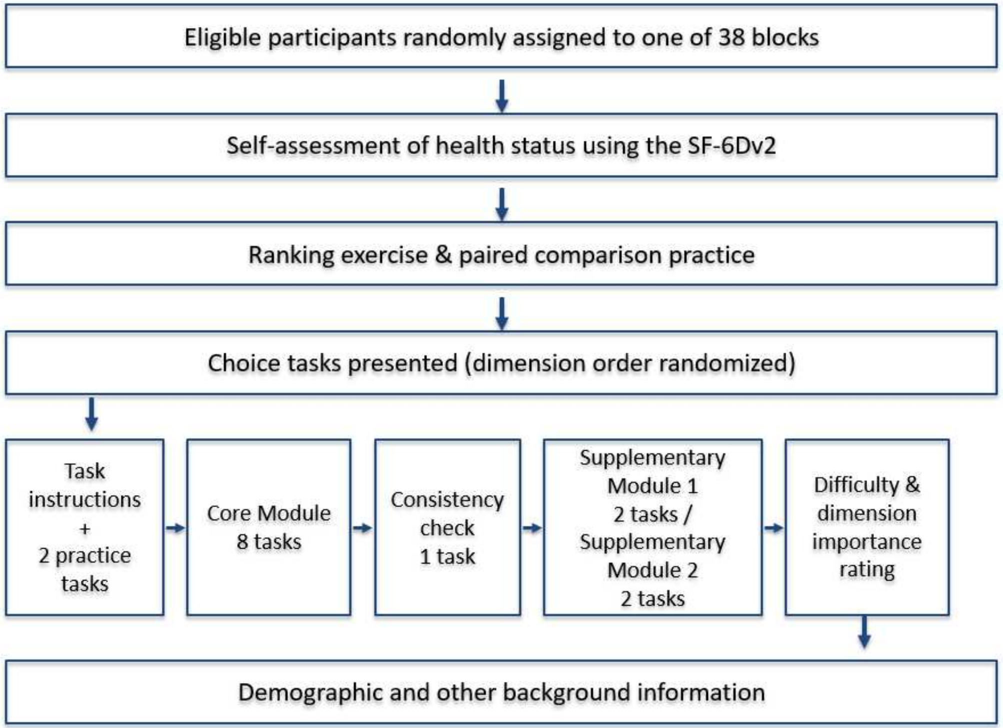
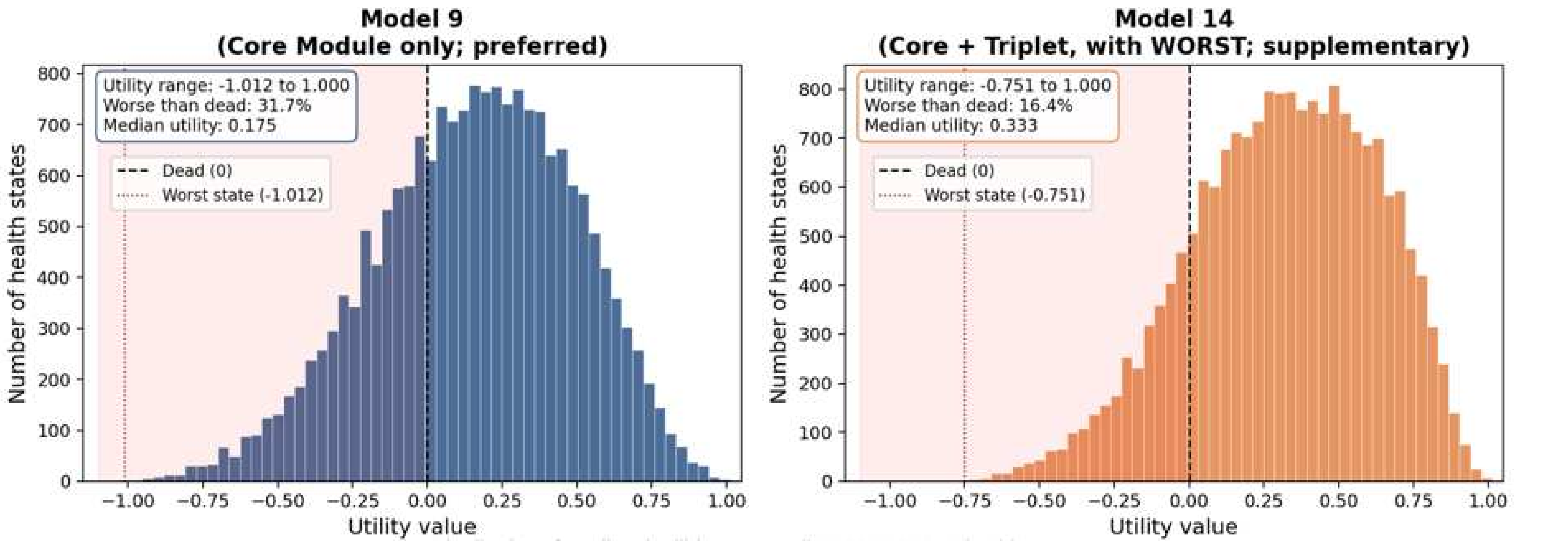


Fig 2. Distribution of SF-6Dv2 Utility Values: Preferred Model (Model 9) vs. Supplementary Model (Model 14)



Distribution of predicted utilities across all 18,750 SF-6Dv2 health states. Red Shading indicates states valued as worse than dead (utility <0)

Table 1. Participants' demographic characteristics

Characteristics	N	%	Korean general population (%)*
Female	1,867	49.13	50.45
Age, years/ 2024.12			
18-29	696	18.32	15.42
30-39	666	17.53	14.94
40-49	778	20.47	17.42
50-59	878	23.11	19.64
60+	782	20.58	32.58
Region			
Seoul	722	19.00	18.22
Gyeonggi-do/Inchon/Gangwon-do	1,379	36.29	35.60
Chungcheong area	380	9.99	10.85
Jeolla area (including Jeju)	401	10.56	10.95
Gyeongbuk area	356	9.37	9.56
Gyeongnam area	562	14.79	14.83
Education level			
Lower than elementary school	12	0.32	9.19
Middle school	21	0.55	7.56
High school	672	17.68	32.80
College or higher	3,086	81.21	50.45
Non-response	9	0.24	
Monthly household income (KW million)**			
Less than 2	288	7.58	NA
2-3.5	778	20.47	NA
3.5-5	890	23.42	NA
5-7.5	914	24.05	NA
More than 7.5	835	21.97	NA
Non-response	95	2.50	NA
Employment			
Permanent employee	1,917	50.45	NA
Temporary employee	555	14.61	NA
Self-employed without employees	249	6.55	NA
Self-employed with employees	108	2.84	NA
Unpaid family worker	271	7.13	NA
Non-employed person	700	18.42	NA
Have long-term condition	2037	53.61	NA
SF-6Dv2			
At ceiling	263	6.92	NA
At floor	0	0	NA
Second most frequent state (112211)	125	3.29	NA
Selection of the dominant option			
Yes	3,706	97.53	NA
No	94	2.47	NA
Consistency check			

Consistent	2,444	64.32	NA
Inconsistent	1,356	35.68	NA
Difficulty in understanding choice tasks			
Very easy	263	6.92	NA
Easy	868	22.84	NA
Moderate	1,618	42.58	NA
Difficult	977	25.71	NA
Very difficult	74	1.95	NA
Difficulty in making a choice			
Very easy	186	4.89	NA
Easy	720	18.95	NA
Moderate	1,441	37.92	NA
Difficult	1,328	34.95	NA
Very difficult	125	3.29	NA
Balanced selection			
Selected both A and B options	3,752	98.7	NA
Selected option A or option B only	48	1.3	NA
Lexicographic preference			
Selection based on multiple dimensions	2375	62.50	NA
Selection based on a single dimension	1425	37.50	NA
PF	239	6.29	NA
RL	75	1.97	NA
PA	434	11.42	NA
VT	135	3.55	NA
SF	176	4.63	NA
MH	184	4.84	NA
Duration	182	4.79	NA
Most influential dimension			
Duration	620	16.32	NA
PF	792	20.84	NA
RL	87	2.29	NA
PA	1,387	36.50	NA
VT	143	3.76	NA
SF	315	8.29	NA
MH	456	12.00	NA
Least influential dimension			
Duration	840	22.11	NA
PF	279	7.34	NA
RL	1,070	28.16	NA
PA	221	5.82	NA
VT	505	13.29	NA
SF	454	11.95	NA
MH	431	11.34	NA
Response time (min)			NA
mean (SD)	22.1(42.2)		NA
median (IQR)	15.1 (11.2~21.4)		NA

*Compiled by the authors from Statistics Korea (KOSTAT) via KOSIS (Korean Statistical Information Service), including the Population and Housing Census (2020), Household Income and Expenditure

Survey (2024-2025), and Registered Population Statistics (2025-2026) [41].
**Income quintile thresholds as of Q4 2024 are 2.11 (p20), 3.64(p40), 5.24(p60), 7.64(p80) (unit: million KRW)
NA not available; PF physical functioning; RL role limitation; PA pain; VT vitality; SF social functioning; MH mental health; SD standard deviation; IQR interquartile range

Table 2. Parameter estimates from unconstrained models by module combination

Parameter	Core				Core + Common[1)]				Core + Triplet[2)]				Core + Common + Triplet[3)]			
	Model 1		Model 2: M1 including WORST		Model 3		Model 4: M3 including WORST		Model 5		Model 6: M5 including WORST		Model 7		Model 8: M7 including WORST	
	Coef (p)[4)]	SE	Coef (p)	SE	Coef (p)	SE	Coef (p)	SE	Coef (p)	SE	Coef (p)	SE	Coef (p)	SE	Coef (p)	SE
PF2xLY	-0.012*	0.005	-0.011*	0.005	-0.026***	0.004	-0.028***	0.004	-0.003	0.005	-0.001	0.005	-0.022***	0.004	-0.022***	0.004
PF3xLY	-0.026***	0.005	-0.027***	0.005	**-0.016*****	0.004	**-0.017*****	0.004	-0.005	0.004	-0.005	0.004	**-0.002**	0.004	**-0.002**	0.004
PF4xLY	-0.084***	0.005	-0.085***	0.005	-0.072***	0.004	-0.071***	0.004	-0.045***	0.004	-0.046***	0.004	-0.037***	0.004	-0.037***	0.004
PF5xLY	-0.190***	0.005	-0.188***	0.006	-0.169***	0.005	-0.180***	0.005	-0.150***	0.005	-0.145***	0.005	-0.139***	0.004	-0.141***	0.004
RL2xLY	**0.005**	0.005	**0.006**	0.005	-0.014***	0.004	-0.015***	0.004	-0.003	0.005	-0.001	0.005	-0.023***	0.004	-0.024***	0.004
RL3xLY	-0.029***	0.005	-0.029***	0.005	-0.051***	0.004	-0.050***	0.004	-0.006	0.004	-0.005	0.004	-0.032***	0.004	-0.032***	0.004
RL4xLY	-0.056***	0.005	-0.057***	0.005	-0.056***	0.005	-0.054***	0.005	-0.028***	0.004	-0.029***	0.005	-0.033***	0.004	-0.033***	0.004
RL5xLY	**-0.050*****	0.005	**-0.047*****	0.005	**-0.046*****	0.004	-0.059***	0.005	**-0.027*****	0.004	**-0.020*****	0.004	**-0.031*****	0.004	-0.034***	0.004
PA2xLY	-0.014**	0.005	-0.013***	0.005	-0.030***	0.005	-0.033***	0.005	-0.022***	0.005	-0.018***	0.005	-0.031***	0.004	-0.032***	0.004
PA3xLY	-0.032***	0.005	-0.031***	0.005	-0.053***	0.005	-0.056***	0.005	-0.034***	0.005	-0.030***	0.005	-0.050***	0.004	-0.051***	0.004
PA4xLY	-0.059***	0.005	-0.058***	0.005	-0.078***	0.004	-0.080***	0.004	**-0.025*****	0.004	**-0.023*****	0.004	**-0.046*****	0.004	**-0.046*****	0.004
PA5xLY	-0.186***	0.005	-0.186***	0.005	-0.189***	0.005	-0.192***	0.005	-0.139***	0.004	-0.137***	0.004	-0.145***	0..004	-0.146***	0.004
PA6xLY	-0.197***	0.008	-0.195***	0.008	-0.202***	0.007	-0.211***	0.008	-0.165***	0.006	-0.158***	0.006	-0.171***	0.005	-0.173***	0.006
VT2xLY	-0.000	0.005	**0.001**	0.005	-0.025***	0.004	-0.028***	0.004	-0.018***	0.005	-0.014**	0.005	-0.033***	0.004	-0.034***	0.004
VT3xLY	-0.039***	0.005	-0.038***	0.005	-0.054***	0.004	-0.056***	0.004	**-0.010***	0.004	**-0.007**	0.004	-0.035***	0.004	-0.035***	0.004
VT4xLY	-0.040***	0.005	**-0.038*****	0.005	**-0.048*****	0.005	**-0.053*****	0.004	-0.033***	0.005	-0.029***	0.005	-0.042***	0.004	-0.043***	0.004
VT5xLY	-0.071***	0.005	-0.068***	0.006	-0.074***	0.005	-0.088***	0.005	-0.054***	0.004	-0.045***	0.005	-0.063***	0.004	-0.066***	0.005
SF2xLY	-0.025***	0.005	-0.025***	0.005	-0.040***	0.004	-0.041***	0.004	-0.017***	0.005	-0.016***	0.005	-0.042***	0.004	-0.043***	0.004
SF3xLY	-0.059***	0.005	-0.059***	0.005	-0.062***	0.004	-0.064***	0.004	-0.023***	0.004	-0.021***	0.004	**-0.039*****	0.004	**-0.039*****	0.004
SF4xLY	-0.087***	0.005	-0.088***	0.005	-0.080***	0.004	-0.078***	0.004	-0.062***	0.004	-0.064***	0.004	-0.065***	0.004	-0.064***	0.004
SF5xLY	-0.132***	0.005	-0.130***	0.005	-0.122***	0.004	-0.133***	0.005	-0.085***	0.004	-0.079***	0.004	-0.087***	0.004	-0.089***	0.004
MH2xLY	-0.033***	0.005	-0.033***	0.055	-0.047***	0.004	-0.045***	0.004	-0.023***	0.005	-0.024***	0.005	-0.045***	0.004	-0.045***	0.004
MH3xLY	-0.073***	0.005	-0.073***	0.005	-0.086***	0.004	-0.088***	0.004	-0.036***	0.004	-0.035***	0.004	-0.061***	0.004	-0.061***	0.004
MH4xLY	-0.104***	0.005	-0.104***	0.005	-0.096***	0.005	-0.096***	0.005	-0.070***	0.004	-0.070***	0.004	-0.074***	0.004	-0.074***	0.004
MH5xLY	-0.154***	0.005	-0.151***	0.006	-0.144***	0.005	-0.157***	0.005	-0.109***	0.005	-0.102***	0.005	-0.114***	0.004	-0.117***	0.004
LY	0.395***	0.009	0.396***	0.009	0.396***	0.007	0.395***	0.007	0.327***	0.008	0.330***	0.009	0.368***	0.007	0.368***	0.007
WORST_LY			-0.006	0.005			0.030***	0.005			-0.021***	0.004			0.006	0.004
Observations (n)	60,800		60,800		76,000		76,000		83,600		83,600		98,800		98,800	
Log Likelihood	-17,413		-17,412		-22,037		-22,020		-25,837		-25,826		-30,594		-30,593	
AIC	34,878		34,879		44,126		44,094		51,726		51,706		61,240		61,240	
BIC	35,112		35,122		44,367		44,343		51,969		51,958		61,487		61,496	
Disordered coefficient	RL1-2; RL4-5		RL1-2; RL4-5;		PF2-3; RL4-5;		PF2-3; VT3-4		RL4-5; PA3-4;		RL4-5; PA3-4;		PF2-3; RL4-5;		PF2-3; PA3-4;	

			VT1-2; VT3-4		VT3-4				VT2-3		VT2-3		PA3-4; SF2-3		SF2-3	

[1]Based on responses to Core and Supplementary 1 (Common) Module choice tasks; [2] Based on responses to Core and Supplementary 2 (Triplet) Module choice tasks ; [3] Based on responses to all Modules (Core + Supplementary 1 + Supplementary 2); [4]Coefficient and significance; [5]Coefficient and significance; SD standard deviation; SE standard error; LY life-years; PF physical functioning; RL role limitation; PA pain; VT vitality; SF social functioning; MH mental health; NA not available; WORST dummy variable indicating health states with at least one dimension at its lowest level; AIC Akaike Information Criterion; BIC Bayesian Information Criterion; * significant at 0.05; ** significant at 0.01; *** significant at 0.001

Table 3 Parameter estimates from constrained models

Parameter	Core				Core + Common[1)]				Core + Triplet[2)]				Core + Common + Triplet[3)]			
	Model 9 (M1c[4)])		Model 10 (M2c)		Model 11 (M3c)		Model 12 (M4c)		Model 13 (M5c)		Model 14 (M6c)		Model 15 (M7c)		Model 16 (M8c)	
	Coef (*p*)[5)]	SE	Coef (*p*)[4)]	SE	Coef (*p*)[4)]	SE	Coef (*p*)	SE	Coef (*p*)	SE	Coef (*p*)	SE	Coef (*p*)	SE	Coef (*p*)	SE
PF2xLY	-0.012*	0.005	-0.011*	0.005	-0.021***	0.004	-0.022***	0.004	-0.005	0.005	-0.003	0.005	-0.011**	0.003	-0.011***	0.003
PF3xLY	-0.026***	0.005	-0.026***	0.005	-0.021***	0.004	-0.022***	0.004	-0.005	0.004	-0.005	0.004	-0.011**	0.003	-0.011***	0.003
PF4xLY	-0.083***	0.005	-0.083***	0.005	-0.070***	0.004	-0.071***	0.004	-0.045***	0.004	-0.044***	0.004	-0.039***	0.004	-0.039***	0.004
PF5xLY	-0.190***	0.005	-0.190***	0.006	-0.169***	0.005	-0.180***	0.005	-0.151***	0.005	-0.145***	0.005	-0.140***	0.004	-0.142***	0.004
RL2xLY	0.000		0.000		-0.014***	0.004	-0.015***	0.004	-0.003	0.005	-0.001	0.005	-0.023***	0.004	-0.023***	0.004
RL3xLY	-0.032***	0.004	-0.032***	0.004	-0.050***	0.004	-0.049***	0.004	-0.006	0.004	-0.005	0.004	-0.030***	0.004	-0.031***	0.004
RL4xLY	-0.055***	0.004	-0.055***	0.004	-0.050***	0.004	-0.053***	0.005	-0.028***	0.004	-0.025***	0.004	-0.030***	0.004	-0.031***	0.004
RL5xLY	-0.055***	0.004	-0.055***	0.004	-0.050***	0.004	-0.058***	0.005	-0.028**	0.004	-0.025***	0.004	-0.030***	0.004	-0.032***	0.004
PA2xLY	-0.014**	0.005	-0.013***	0.005	-0.030***	0.005	-0.034***	0.005	-0.019***	0.005	-0.016**	0.005	-0.031***	0.004	-0.032***	0.004
PA3xLY	-0.031***	0.005	-0.031***	0.005	-0.053***	0.005	-0.057***	0.005	-0.027***	0.004	-0.025***	0.004	-0.048***	0.004	-0.048***	0.004
PA4xLY	-0.059***	0.005	-0.059***	0.005	-0.078***	0.004	-0.081***	0.004	-0.027***	0.004	-0.025***	0.004	-0.048***	0.004	-0.048***	0.004
PA5xLY	-0.186***	0.005	-0.186***	0.005	-0.188***	0.005	-0.193***	0.005	-0.137***	0.004	-0.135***	0.004	-0.145***	0..004	-0.146***	0.004
PA6xLY	-0.197***	0.008	-0.195***	0.008	-0.198***	0.007	-0.208***	0.008	-0.164***	0.006	-0.158***	0.006	-0.169***	0.005	-0.170***	0.006
VT2xLY	-0.000	0.005	0.000		-0.024***	0.004	-0.027***	0.004	-0.013**	0.004	-0.010*	0.004	-0.031***	0.004	-0.032***	0.004
VT3xLY	-0.039***	0.005	-0.039***	0.004	-0.051***	0.004	-0.054***	0.004	-0.013**	0.004	-0.010*	0.004	-0.032***	0.004	-0.033***	0.004
VT4xLY	-0.040***	0.005	-0.040***	0.004	-0.051***	0.004	-0.054***	0.004	-0.032***	0.005	-0.029***	0.005	-0.040***	0.004	-0.040***	0.004
VT5xLY	-0.071***	0.005	-0.071***	0.004	-0.074***	0.005	-0.087***	0.005	-0.054***	0.004	-0.046***	0.005	-0.061***	0.004	-0.063***	0.005
SF2xLY	-0.024***	0.005	-0.024***	0.005	-0.039***	0.004	-0.040***	0.004	-0.017***	0.005	-0.016***	0.005	-0.039***	0.003	-0.039***	0.003
SF3xLY	-0.058***	0.005	-0.058***	0.005	-0.062***	0.004	-0.064***	0.004	-0.023***	0.004	-0.022***	0.004	-0.039***	0.003	-0.039***	0.003
SF4xLY	-0.087***	0.005	-0.087***	0.005	-0.079***	0.004	-0.077***	0.004	-0.062***	0.004	-0.063***	0.004	-0.062***	0.004	-0.061***	0.004
SF5xLY	-0.130***	0.005	-0.130***	0.005	-0.121***	0.004	-0.132***	0.005	-0.085***	0.004	-0.079***	0.004	-0.085***	0.004	-0.087***	0.004
MH2xLY	-0.033***	0.005	-0.033***	0.005	-0.047***	0.004	-0.046***	0.004	-0.024***	0.005	-0.024***	0.005	-0.045***	0.004	-0.045***	0.004
MH3xLY	-0.074***	0.005	-0.074***	0.005	-0.087***	0.004	-0.088***	0.004	-0.036***	0.004	-0.036***	0.004	-0.061***	0.004	-0.061***	0.004
MH4xLY	-0.105***	0.005	-0.105***	0.005	-0.096***	0.005	-0.096***	0.005	-0.071***	0.004	-0.071***	0.004	-0.075***	0.004	-0.075***	0.004
MH5xLY	-0.155***	0.005	-0.154***	0.005	-0.146***	0.005	-0.158***	0.005	-0.111***	0.004	-0.105***	0.005	-0.115***	0.004	-0.117***	0.004
LY	0.396***	0.009	0.396***	0.009	0.394***	0.007	0.394***	0.007	0.328***	0.008	0.329***	0.008	0.365***	0.007	0.365***	0.007
WORST_LY			-0.001	0.005			0.029***	0.005			-0.018***	0.004			0.005	0.004
Observations (n)	60,800		60,800		76,000		76,000		83,600		83,600		98,800		98,800	
Log Likelihood	-17,415		-17,415		-22,043		-22,024		-25,841		-25,831		-30,609		-30,608	
AIC	34,877		34,877		44,132		44,098		51,728		51,710		61,262		61,265	
BIC	35,094		35,094		44,344		44,328		51,942		51,934		61,471		61,493	
Collapsed level	RL1-2; RL4-5		RL1-2; RL4-5; VT1-2;		PF2-3; RL4-5; VT3-4		PF2-3; VT3-4		RL4-5; PA3-4; VT2-3		RL4-5; PA3-4; VT2-3		PF2-3; RL4-5; PA3-4;		PF2-3; PA3-4; SF2-3	

			VT3-4									SF2-3			

Model 9-16 are constrained versions of Model 1-8 respectively.

[1]Based on responses to Core and Supplementary 1 (Common) Module choice tasks; [2] Based on responses to Core and Supplementary 2 (Triplet) Module choice tasks ; [3] Based on responses to all Modules (Core + Supplementary 1 + Supplementary 2); [4]Coefficient and significance; [5]Coefficient and significance; SD standard deviation; SE standard error; LY life-years; PF physical functioning; RL role limitation; PA pain; VT vitality; SF social functioning; MH mental health; NA not available; WORST dummy variable indicating health states with at least one dimension at its lowest level; AIC Akaike Information Criterion; BIC Bayesian Information Criterion; * significant at 0.05; ** significant at 0.01; *** significant at 0.001

Table 4 Anchored models

	Model 9 (Core)		Model 14 (Core + Triplet)	
	Coef	95% CI	Coef	95% CI
PF1	0.000	[0.000, 0.000]	0.000	[0.000, 0.000]
PF2	-0.029	[-0.057, -0.002]	-0.008	[-0.038, 0.023]
PF3	-0.065	[-0.093, -0.038]	-0.015	[-0.042, 0.012]
PF4	-0.210	[-0.243, -0.177]	-0.135	[-0.168, -0.102]
PF5	-0.480	[-0.527, -0.432]	-0.442	[-0.493, -0.391]
RL1	0.000	[0.000, 0.000]	0.000	[0.000, 0.000]
RL2	0.000	[0.000, 0.000]	-0.002	[-0.030, 0.026]
RL3	-0.080	[-0.105, -0.056]	-0.015	[-0.042, 0.012]
RL4	-0.138	[-0.162, -0.115]	-0.076	[-0.103, -0.048]
RL5	-0.138	[-0.061, -0.008]	-0.076	[-0.103, -0.048]
PA1	0.000	[0.000, 0.000]	0.000	[0.000, 0.000]
PA2	-0.034	[-0.061, -0.008]	-0.048	[-0.079, -0.016]
PA3	-0.079	[-0.109, -0.050]	-0.076	[-0.103, -0.049]
PA4	-0.149	[-0.179, -0.119]	-0.076	[-0.103, -0.049]
PA5	-0.469	[-0.514, -0.424]	-0.412	[-0.458, -0.365]
PA6	-0.496	[-0.555, -0.437]	-0.479	[-0.537, -0.421]
VT1	0.000	[0.000, 0.000]	0.000	[0.000, 0.000]
VT2	0.001	[-0.024, 0.025]	-0.030	[-0.057, -0.004]
VT3	-0.098	[-0.126, -0.069]	-0.030	[-0.057, -0.004]
VT4	-0.100	[-0.130, -0.071]	-0.088	[-0.120, -0.056]
VT5	-0.178	[-0.211, -0.146]	-0.139	[-0.175, -0.103]
SF1	0.000	[0.000, 0.000]	0.000	[0.000, 0.000]
SF2	-0.061	[-0.088, -0.034]	-0.047	[-0.077, -0.017]
SF3	-0.147	[-0.178, -0.117]	-0.066	[-0.095, -0.037]
SF4	-0.220	[-0.252, -0.187]	-0.193	[-0.228, -0.158]
SF5	-0.329	[-0.366, -0.292]	-0.241	[-0.279, -0.203]
MH1	0.000	[0.000, 0.000]	0.000	[0.000, 0.000]
MH2	-0.084	[-0.111, -0.057]	-0.074	[-0.105, -0.043]
MH3	-0.186	[-0.218, -0.154]	-0.110	[-0.141, -0.079]
MH4	-0.264	[-0.301, -0.228]	-0.215	[-0.253, -0.178]
MH5	-0.391	[-0.434, -0.348]	-0.320	[-0.364, -0.276]
WORST	NA		-0.054	[-0.081, -0.027]
Utility	1 to -1.012		1 to -0.751	

range		
WTD(%)	31.7	16.4
Collapsed levels	RL1-2; RL4-5	RL4-5; PA3-4; VT2-3

Coef coefficient; CI confidence interval; PF physical functioning; RL role limitation; PA pain; VT vitality; SF social functioning; MH mental health; NA not available; WORST dummy variable indicating health states with at least one dimension at its lowest level; WTD worse than death

Table 5. International Comparison

	Kore	Japan [17] 1)	China [12] 2)		UK [6] 3)	Australia [13] 4)	New Zealand [16]5), 8)	Quebec [14]6)	Iran [15]7)
			TTO	DCE_{TTO}					
PF									
PF1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PF2	-0.029	-0.085	-0.038	-0.027	-0.019	-0.042	-0.037	-0.103	-0.022
PF3	-0.065	-0.107	-0.080	-0.050	-0.034	-0.042	-0.075	-0.152	-0.111
PF4	-0.210	-0.327	-0.140	-0.145	-0.092	-0.138	-0.173	-0.225	-0.174
PF5	-0.480	-0.593	-0.395	-0.404	-0.186	-0.222	-0.271	-0.338	-0.332
RAI3)	**0.239**	**0.344**	**0.309**	**0.263**	**0.118**	**0.132**	**0.172**	**0.201**	**0.185**
RL									
RL1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RL2	0.000	-0.016	-0.050	-0.029	-0.039	-0.029	-0.028	-0.057	-0.076
RL3	-0.080	-0.124	-0.059	-0.030	-0.055	-0.045	-0.055	-0.079	-0.123
RL4	-0.138	-0.162	-0.096	-0.081	-0.099	-0.127	-0.118	-0.079	-0.170
RL5	-0.138	-0.162	-0.097	-0.112	-0.102	-0.127	-0.182	-0.228	-0.189
RAI	**0.069**	**0.094**	**0.076**	**0.073**	**0.065**	**0.075**	**0.116**	**0.135**	**0.105**
PA									
PA1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PA2	-0.034	-0.085	-0.047	-0.072	-0.076	-0.090	-0.066	-0.097	-0.069
PA3	-0.079	-0.115	-0.083	-0.079	-0.097	-0.096	-0.119	-0.137	-0.080
PA4	-0.149	-0.245	-0.154	-0.155	-0.139	-0.190	-0.172	-0.173	-0.111
PA5	-0.469	-0.497	-0.388	-0.466	-0.460	-0.514	-0.274	-0.253	-0.308
PA6	-0.496	-0.522	-0.427	-0.541	-0.620	-0.677	-0.376	-0.417	-0.439
RAI	**0.247**	**0.303**	**0.334**	**0.352**	**0.394**	**0.402**	**0.239**	**0.248**	**0.244**
VT									
VT1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
VT2	0.001	-0.019	-0.029	0.000	-0.015	-0.002	-0.034	-0.033	-0.020
VT3	-0.098	-0.071	-0.060	-0.083	-0.015	-0.002	-0.067	-0.064	-0.038
VT4	-0.100	-0.115	-0.108	-0.089	-0.080	-0.040	-0.151	-0.140	-0.121
VT5	-0.178	-0.115	-0.116	-0.167	-0.121	-0.098	-0.236	-0.201	-0.139
RAI	**0.088**	**0.067**	**0.091**	**0.109**	**0.077**	**0.058**	**0.150**	**0.119**	**0.077**
SF									
SF1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SF2	-0.061	-0.047	-0.047	-0.035	-0.008	-0.004	-0.022	-0.036	-0.027
SF3	-0.147	-0.066	-0.060	-0.038	-0.029	-0.033	-0.045	-0.070	-0.094
SF4	-0.220	-0.104	-0.093	-0.104	-0.103	-0.105	-0.106	-0.134	-0.179
SF5	-0.329	-0.115	-0.108	-0.114	-0.137	-0.148	-0.167	-0.246	-0.195
RAI	**0.164**	**0.067**	**0.085**	**0.074**	**0.087**	**0.088**	**0.106**	**0.146**	**0.109**
MH									
MH1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

MH2	-0.084	-0.060	-0.033	0.000	-0.026	-0.014	-0.060	-0.033	-0.043
MH3	-0.186	-0.102	-0.050	-0.010	-0.086	-0.080	-0.122	-0.112	-0.155
MH4	-0.264	-0.195	-0.132	-0.146	-0.236	-0.251	-0.231	-0.143	-0.292
MH5	-0.391	-0.214	-0.134	-0.197	-0.324	-0.334	-0.340	-0.253	-0.346
RAI	**0.194**	**0.124**	**0.105**	**0.128**	**0.206**	**0.198**	**0.216**	**0.150**	**0.193**
WORST					-0.084	-0.079			-0.156
RAI					**0.053**	**0.047**			**0.087**
% of worse than death	31.8	NA	4.9	8.5	15.1	19.6	9.6	13.0	33.5
Health states									
556555	-1.012	-0.722	-0.277	-0.535	-0.574	-0.685	-0.572	-0.683	-0.796
112211	0.967	0.896	0.924	0.928	0.909	0.908	0.9	0.87	0.911
333333	0.345	0.415	0.608	0.710	0.684	0.702	0.517	0.386	0.399

[1]Data for Japan were obtained from Shiroiwa et al. [17]; [2]Data for China were obtained from Wu et al. [12]; [3]Data for the UK were obtained from Mulhern et al. [6]; [4]Data for Australia were obtained from Mulhern et al. [13]; [5]Data for New Zealand were obtained from Sullivan et al. [16]; [6]Data for Quebec were obtained from Ameri et al. [14]; [7]Data for Iran were obtained from Daroudi et al. [15]; [8]Unlike other countries using the DCETTO protocol, the New Zealand value set was derived using an adaptive DCE approach, specifically the PAPRIKA (Potentially All Pairwise RanKings of all possible Alternatives) method.

RAI relative attribute importance; PF physical functioning; RL role limitation; PA pain; VT vitality; SF social functioning; MH mental health; [3]The proportion of the overall range that each dimension covers; RAI relative attribute importance; WORST dummy variable indicating health states with at least one dimension at its lowest level; NA not available

The U.S. study was not included in the list of published value sets because, at the time of writing, its findings had only been presented in abstract form and the full results were not yet available for citation.