# Development of SF-6Dv2 Health Utility Weights for the United States

Regina Rendas-Baum[1], Brendan Mulhern[2], Cary Thurm[1], Kristen McCausland[1], John E Brazier[3], Jakob B Bjorner[4]

[1]QualityMetric, Incorporated LLC, Johnston, RI, USA, [2]University of Technology Sydney, Sydney, NSW, Australia, [3]University of Sheffield, Sheffield, UK, [4]QualityMetric Incorporated, LLC, Virum, Denmark
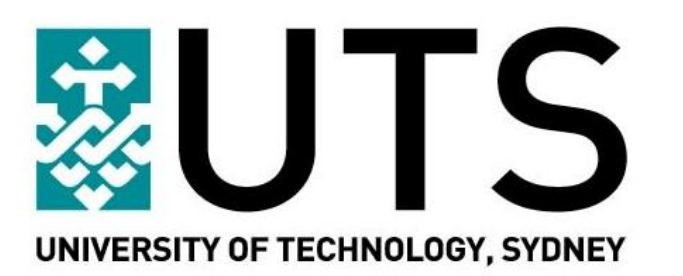
Background

- In 1998, researchers developed the SF-6D based on responses to 11 questions from the SF-36[1] with the resulting classification describing health states in 6 multi-level dimensions: physical functioning (PF), role limitations (RL), bodily pain (BP), vitality (VT), social functioning (SF), and mental health (MH)

- A new version of the SF-6D health state classification system (SF-6Dv2) was recently developed using a multi-step process[2] that included the following:
  - Evaluating item performance and differential item functioning (DIF) using item response theory (IRT) analyses and establishing dimensionality through factor analysis
  - Obtaining a simpler classification of physical function with clearer separation between levels;
  - Using negative wording to describe vitality, using pain severity rather than pain interference, and using a more detailed 5-level description of role limitations

- SF-6Dv2 value sets have been developed for the UK[3], China[4] and Australia[5]

Study Objectives

To estimate the SF-6Dv2 value set for the United States (US).

Methods

Discrete choice experiments (DCE) are a method by which stated preference data are obtained based on comparisons of different health states. By including survival in the model, the DCE estimates can be anchored on the utility scale (where 1 is equivalent to full health, and 0 is equivalent to death). This approach, named DCE with survival (DCE$_{TTO}$) was used in this study.

Discrete Choice Experiment Designs

- Design 1 (core): 8 choice tasks from a set of 304 based on a full factorial design that excluded combinations of two-way dimension levels that were very unlikely to occur; each task compared 2 health profiles with 1 of 4 duration levels (1, 4, 7, and 10 years); small priors for duration/dimension calculated as average of the corresponding model parameter in prior studies[3, 5]

- Design 2 (common): 2 choice tasks from a set of 76, based on pairs of health states commonly experienced by the general population. Common health states were identified from a US general population study[6] using a multivariate latent response model and Monte-Carlo simulation

- Design 3 (ternary): 2 choice tasks from a set of 76 that involved a 3rd choice of immediate death; for ternary tasks, severe health states were selected by restricting the levels to the 3 most severe in each dimension; respondents were asked to identify the best and the worst choice

Data Source and Collection Process

- A target sample of 3,800 participants was recruited from pre-existing panels to ensure ~100 participants evaluated each choice task. The survey was comprised of 3 components: 1) self complete SD-6Dv2 Health Utility Survey; 2) the primary survey component, consisting of the 15 TTO choice tasks; 3) questions about experience with certain health conditions. (Figure 1)

Figure 1. Data Collection Flow and Randomization Procedures
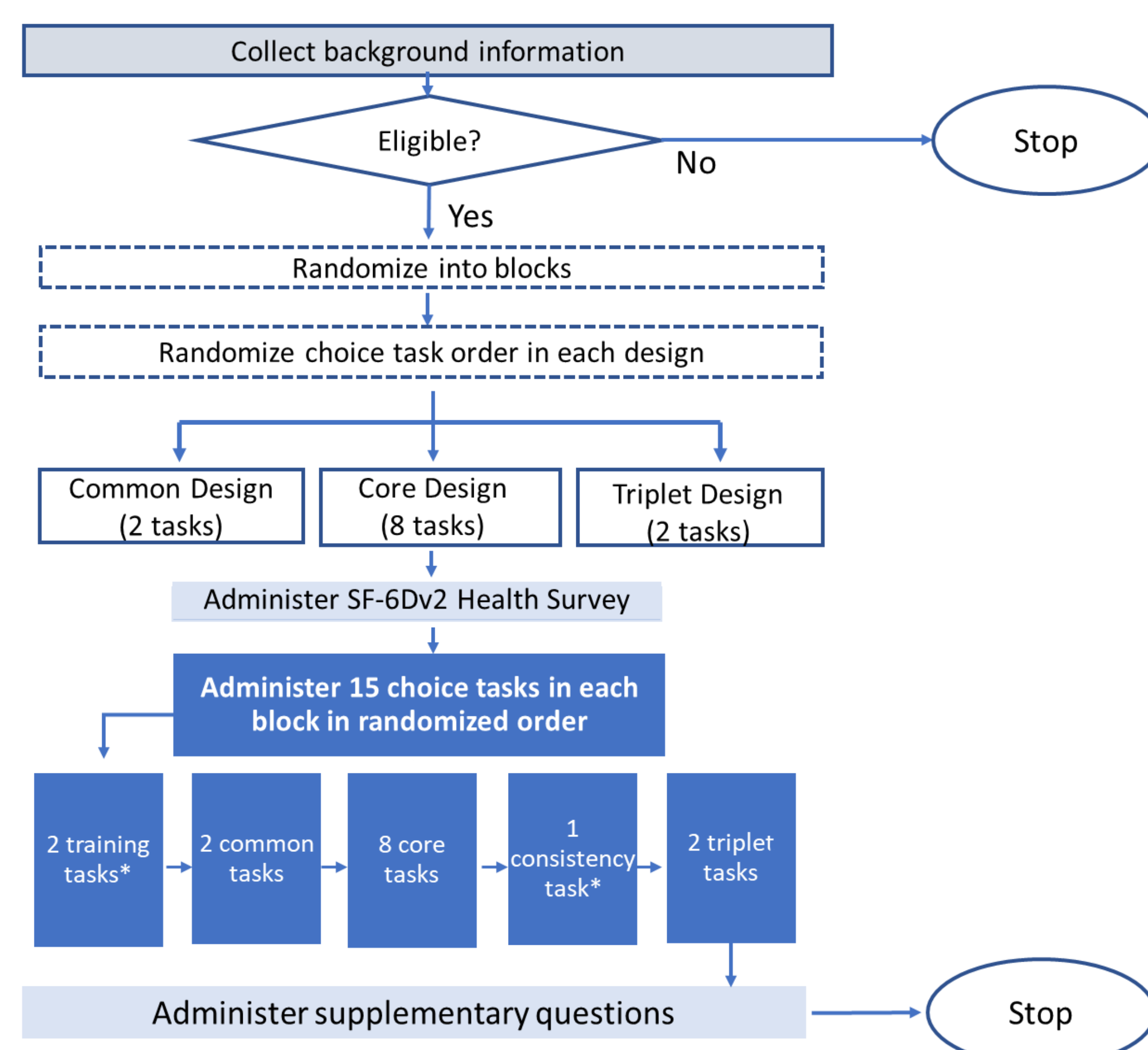
* Choice tasks that are the same across blocks. These tasks were not used for estimating utility weights.

Statistical Analysis

- Utility weights estimation was based on the conditional logit model (CLM)
- Estimates obtained using just data from design 1 (Model A), design 1+2 (Model B), design 1+3 (Model C), and design 1+2+3 (Model D)
- The following extensions to the base model were considered:
  a) Non-monotonicity of attribute levels remodeled to ensure logically ordered coefficients
  b) Coefficient added for health states where a dimension is at the worst level (*WORST*)
  c) Two-way interactions between the dimensions
- Models were compared by applying the estimates across subsets of respondents from a large independent study[6], grouped by self-reported medical condition diagnosis (a diverse set of 24 conditions was considered)

Poster presented at ISPOR Europe 2022; 6-9 November 2022, Vienna, Austria

Results

Sample Demographics

- The mean age of the sample (N=3,807) was 44.8 years; 27% were ≤ 30 and nearly 22% ≥ 65 or older
- Slightly more than half of the sample was White; nearly 25% identified as Black
- Approximately half of the sample reported at least one chronic condition (present or past)
- The percentage reporting the best health state (5.3%) was similar to what was reported in the Australian valuation study[5]; 0% reported the lowest score

Utility Weights Estimation

Monotonicity

- Model D (*core+common+ternary*) showed 3 disordered dimension levels, in comparison to models A, B and C, which had 4 or 5
- Non-monotonic coefficients occurred in dimensions and levels with lower disutility levels, with pain and mental health consistently ordered for all levels; this is analogous to what was reported in the Australian valuation study[5] and an improvement relative to the original version

WORST term

- The magnitude of this parameter was small and only statistically significant in model B; the direction of the interaction was against expectation in models B and D; based on these results, the WORST term was not included in subsequent models

Interactions between dimension levels

- While all interactions were statistically significant and improved model fit (AIC), the pain and mental health dimensions were associated with the highest inferential test statistics
- A simple interaction term ((BP≥5, MH≥4) was added to model D, after correcting for non-monotonicity, resulting in higher utility values for BP and MH at the most severe levels
- When applied to a large independent study, results indicated a high level of agreement between the two models - with and without interaction - with the differences between the two models varying between 0.003 and -0.007
- Despite the small magnitude in differences, exclusion of interactions resulted in slightly larger differences for most diseases, suggesting a greater sensitivity to the burden associated with health conditions

Figure 2. Comparison of Utility Weights for Two Alternative US Models and Weights for the UK, China, and Australia
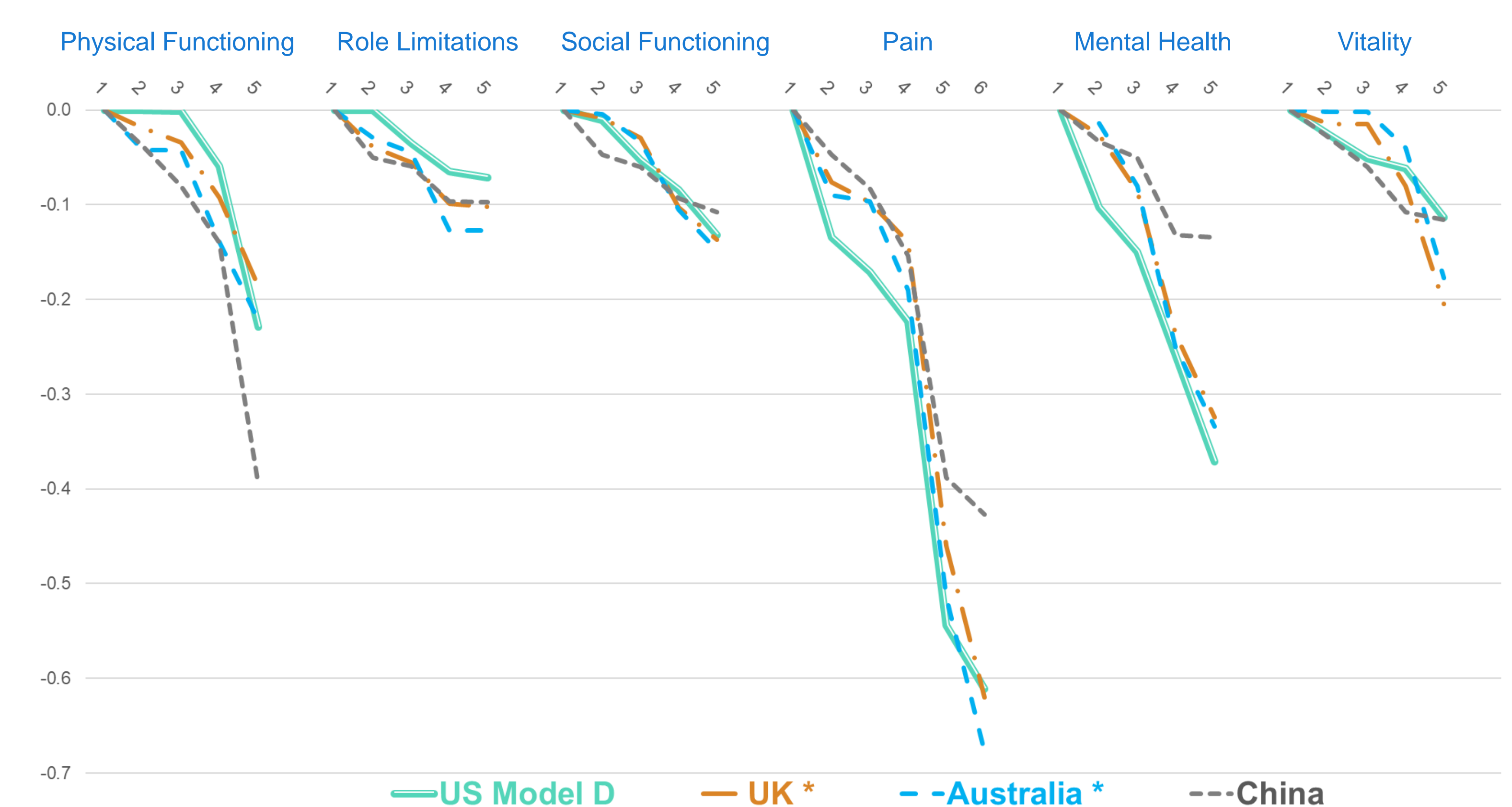
*Effects of WORST parameter reflected in the plotted estimates.

Conclusions

- In contrast to results from prior studies, a term triggered by selecting the worst health state on any dimension was not relevant to the US value set
- Compared to the UK, Chinese, and Australian weights, the US value set results in consistently numerically higher weights for physical functioning and role functioning
- The US tariffs suggest more sensitivity (greater disutility) to low levels of pain and mental distress, but less sensitivity (lower disutility) to impact on role and physical functioning compared to the UK and Australian tariffs
- Establishing US-based utility weights for the newly developed SF-6Dv2 is an important step towards greater adoption of use in health economic analyses in the US

References

1. Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol*. 1998;51(11):1115-1128. doi:10.1016/s0895-4356(98)00103-6
2. Brazier JE, Mulhern BJ, Bjorner JB, 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):557-565.
3. 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):566-573. doi:10.1097/MLR.0000000000001324
4. Wu J, Xie S, He X, 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):521-535.
5. Mulhern B, Norman R, Brazier J. Valuing SF-6Dv2 in Australia Using an International Protocol. *Pharmacoeconomics*. 2021. Oct;39(10):1151-1162
6. Bayliss M, Rendas-Baum R, White MK, Maruish M, Bjorner J, Tunis SL. Health-related quality of life (HRQL) for individuals with self-reported chronic physical and/or mental health conditions: panel survey of an adult sample in the United States. *Health Qual Life Outcomes*. 2012;10:154.