*39th EuroQol Plenary 2022 - Paper presentation*

# Anchoring DCE latent utilities using the TTOs collected from the same versus different respondents

Shitong Xie, PhD [1,2,#], Jing Wu, PhD [2,3,#], Feng Xie, PhD [1,4,*]

[1] Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada

[2] School of Pharmaceutical Science and Technology, Tianjin University, Tianjin, China

[3] Center for Social Science Survey and Data, Tianjin University, Tianjin, China

[4] Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, ON, Canada

[#] Shitong Xie and Jing Wu contributed equally to this paper

[*] **Corresponding author:** Feng Xie, PhD, Professor, Email: fengxie@mcmaster.ca

## ABSTRACT

**Objectives:** To compare anchored DCE utility values using own vs. other's TTO responses in the valuation of SF-6Dv2.

**Methods:** A representative sample of the general population in Tianjin, China, was recruited. During face-to-face interviews, DCE and TTO data were collected from a randomly selected half of the respondents (own TTO sample), while only TTO data were collected from the other half (other's TTO sample). Conditional logit model was used to estimate DCE latent utilities. Three anchoring methods were used: (1) using the observed TTO value for the worst state; (2) using the modeled TTO value for the worst state; and (3) mapping DCE values onto TTO. For each method, prediction accuracy using intraclass correlation coefficient (ICC), mean absolute difference (MAD), root mean squared difference (RMSD), and the number of states with prediction errors > 0.05 and > 0.1 compared with the mean observed TTO values, were compared between the anchoring results using the own vs. other's TTO data.

**Results:** Comparable demographic characteristics were observed between the own TTO sample (N = 252) and the other's TTO sample (N = 251). The mean (SD) observed TTO value for the worst state was -0.259 (0.591) for the own TTO sample and -0.236 (0.616) for the other's TTO sample. Bland–Altman plot showed a comparable distribution above and below zero across the severity spectrum between the two samples. The estimated TTO values were -0.223 for the own TTO sample and -0.237 for the other's TTO sample. Anchoring with own TTOs consistently showed a better prediction accuracy than using other's TTOs across the three anchoring methods in terms of the ICC (0.835-0.873 vs. 0.771-0.804), MAD (0.127-0.181 vs. 0.146-0.203), RMSD (0.164-0.237 vs. 0.192-0.270), and the number of states with prediction error > 0.05 (21-25 vs. 21-26) and > 0.1 (61-72 vs. 64-75).

**Conclusion:** There are systematic differences in anchored DCE utility values between using the own TTO data and other's TTO data.



**Keywords:** Time trade-off; discrete choice experiment; latent utility; anchor

## INTRODUCTION

Preference-based measures (PBMs) are widely used to generate health utility values for calculating quality-adjusted life years (QALYs) (1). Health utility values are cardinal preferences anchored at 0 for dead and 1 for full health with negative values indicating states worse than dead (1, 2). Among the most widely used generic PBMs are the EQ-5D and the Short Form six-dimension (SF-6D) (1, 2). These instruments use a scoring function developed based on preferences from a representative sample of the general public to assign health utility values to the health states they describe (1).

Health utility values can be elicited using established methods such as standard gamble and time trade-off (TTO) (1-5). However, there are concerns about these approaches because they are likely to be affected by factors other than respondents' preferences for the state, such as time preference and aversion to losses for TTO (6-8). Furthermore, these approaches are cognitively complex, and respondents might have some difficulty in understanding and completing the task, particularly those in vulnerable groups such as the elderly or children (7). One of the most recent developments in utility elicitation is the adoption of the discrete choice experiment (DCE), especially for online surveys (9-11).

Compared with the iterative process of identifying the indifference point between two options in SG and TTO, DCE is usually regarded as a promising alternative since it only requires respondents to make ordinal preference among alternative health states (12, 13). However, modeling on DCE data generates latent utilities. So they need to be rescaled to the health utility scale.(7). Several studies have attempted to anchor DCE values onto the QALY scale using external data, for example, using the TTO values of the worst state as the anchor, or mapping DCE values onto TTO values (7, 14). Several variants of the DCE, e.g., DCE with "immediate death" and DCE with duration, are also explored to anchor latent utility values without using external data (13, 15).

The international valuation protocol for the EQ-5D-Y-3L, a child-friendly version of the EQ-5D-3L (16), recommends a combination of the use of DCE and TTO (17). DCE responses are used to obtain the relative importance of the different dimensions/levels (17). The composite

TTO responses are employed for anchoring purposes (17). In this protocol, DCE responses are collected through online survey while TTO responses are elicited through face-to-face interviews. However, there is no evidence showing the impact of using TTOs from a different sample for anchoring. This study aimed to compare anchored DCE values using own or other's TTO responses. This methodological investigation was part of the SF-6Dv2 valuation study.

## METHODS

We used the data collected alongside the pilot valuation study of the SF-6Dv2 conducted between June and August 2018. The detailed information on the pilot study can be found elsewhere (8). Informed consent was obtained from all respondents. The protocol of this study was approved by the Institutional Review Board of School of Pharmaceutical Science and Technology, Tianjin University, China (No. 20180615).

### *Instrument*

The SF-6Dv2 has six dimensions: physical functioning (PF), role limitation (RL), social functioning (SF), pain (PN), mental health (MH), and vitality (VT). All dimensions have five response levels except for the pain dimension, which has six response levels. The SF-6Dv2 describes a total of 18,750 health states. A detailed description of the SF-6Dv2 can be found elsewhere (18, 19).

### *Elicitation Tasks Design*

Both composite TTO and DCE elicitation approaches were employed in this study (8). The composite TTO approach (hereafter TTO) uses conventional TTO to elicit health states perceived as "better than dead" and lead-time TTO to health states perceived as "worse than dead", respectively (4, 20, 21). 115 health states were selected for TTO tasks, including the six mildest health states, the worst state, and 108 states selected based on the orthogonal design using SAS Studio. The 108 states were split into 18 blocks of six health states. The worst state and randomly selected one of the six mildest states were then added to each block. Each respondent was randomly assigned to complete one block of eight TTO tasks for valuation.

The DCE task presents a pair of health states (labeled state A and state B) described by the SF-6Dv2. 150 pairs of states were selected based on the balanced overlap method. The statistical efficiency was maximized regarding the D-efficiency using Lighthouse Studio 9.6.0 (Sawtooth Software, Inc) (21-23). The 150 pairs of states were then assigned to 15 blocks of 10 pairs. Each respondent was randomly assigned to complete one block for valuation.

The plausibility of combinations of levels of dimensions was also incorporated into the health state selection process for both TTO and DCE approaches. To balance the experimental design that can explore any possible combination of dimensions and levels in a health state and one that ensures plausibility is retained, only one implausible combination (level 1 in RL with level 6 in PN) was excluded prior to the health state selection, following previous literature in this study (24).

***Sampling and Data collection***

Adult respondents (target N = 500) from the general public were recruited in Tianjin, China. Tianjin city is one of the four municipalities in China, with more than 15 million permanent population. A quota sampling method was used to recruit a representative sample stratified in terms of age group, sex, education level, and area of residence (urban/rural) of the general population in Tianjin (8).

Face-to-face interviews were used for the data collection (8). The interview started with the respondent completing the eligibility questions, and then describing his/her own health state using the SF-6Dv2. Half of the respondents (hereafter own TTO sample) were then randomly assigned to 8 TTO tasks and 10 DCE tasks, while the other half (hereafter other's TTO sample) were assigned to 8 TTO tasks and 10 DCE with duration ($DCE_{TTO}$) tasks. The order of TTO and DCE/$DCE_{TTO}$ tasks within each respondent were randomized. After valuation tasks, the respondents answered social-demographic questions. A detailed description of sampling and data collection was published elsewhere (8). We used the data of TTO and DCE tasks to conduct this study (Figure 1).

***Data analysis***

Descriptive analyses (mean, standard deviation [SD], frequency, and proportion) were firstly conducted to present the respondents' characteristics. The utility value for the self-reported SF-6Dv2 health state was calculated using the China value set (21). The characteristics between the own TTO sample and other's TTO samples were compared using t-test for continuous variables and chi-square test for categorical variables. The Bland-Altman plot was used to assess the agreement of the observed TTO values between the two samples. The difference in mean TTO values for each elicited state between the two samples was also evaluated using the Mann-Whitney test.

TTO and DCE data were modeled with the same model specification chosen for the pilot valuation study of the SF-6Dv2 (8). In brief, the TTO data of each sample were analyzed using the fixed effects model (Eq. 1):

$$y_i = \alpha + \sum_d \sum_l \beta_{dl} x_{dl} + \epsilon \quad (1)$$

Where $y_i$ is the disutility value given by the respondent $i$; α is the intercept; $x_{dl}$ are 25 dummy variables indicating the health state described by SF-6Dv2 dimension $d$ at level $l$, except the first level of each dimension (for reference) with the corresponding coefficient $\beta_{dl}$, and $\varepsilon$ is the error term.

The DCE data were analyzed under the random utility framework using the conditional logit model (8). The utility function consisted of 25 dummy variables similar to what has been shown in Eq. 1. The error term was assumed to be independently and identically distributed with Gumbel distribution. The adjacent inconsistent levels were combined for the DCE model to produce a fully monotonic model, considering the goodness-of-fit of model estimation based on the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).

The DCE latent utilities were then anchored onto the full health-dead scale using their own TTO data and other's TTO data. Three anchoring methods were used (7, 14): 1) using the mean observed TTO value for the worst state; 2) using the model estimated TTO value for the worst state; and 3) mapping DCE values onto TTO values.

For the first two methods, the value of the worst state in the DCE model is anchored at the TTO value of the worst state. Specifically, the coefficients on a latent utility scale estimated in the DCE model were normalized onto the full health-dead scale using an adjusted weight containing the TTO value and DCE latent value for the worst state (7, 14). The mean observed TTO value of the worst state was used in the first anchoring method, and the estimated TTO value of the worst state generated by the TTO model was used in the second anchoring method. This is achieved using the Eq. 2:

$$\beta_{DCE-rescaled} = \beta_{DCE-original} \cdot {}^{W_{TTO}}/_{W_{DCE}} \tag{2}$$

where $\beta_{DCE-original}$ is the original coefficient for level $l$ of dimension $d$; $\beta_{DCE-rescaled}$ is the rescaled coefficient for level $l$ of dimension $d$; $W_{TTO}$ is the mean observed TTO value for the worst state (Method 1) or the estimated TTO value for the worst state generated using the TTO model (Method 2); and $W_{DCE}$ is the DCE value for the worst state estimated in the DCE model.

The mapping method was then used in this study to convert DCE latent values onto TTO values (14, 25). Specifically, the DCE model generated values on a latent utility scale for all 18,750 states. 115 of these states were also directly valued using TTO by the respondents during the interviews. Both DCE latent values and TTO values of these 115 states were used to develop the mapping function. All DCE latent values of the 18,750 states in the SF-6Dv2 were then converted onto TTO values using this mapping algorithm. The mapping function from DCE to TTO was specified as (Eq. 3):

$$TTO_j = f(DCE_j) \tag{3}$$

Where $TTO_j$ represents the mean TTO value of health state $j$, $DCE_j$ represents the modeled latent utility value for health state $j$. The mapping was undertaken using ordinary least squares (OLS) model assuming a linear relationship.

The prediction accuracy of each anchoring method was first evaluated by the agreement between the anchored values and observed TTO values using the intraclass correlation coefficient (ICC). The ICC was computed with the two-way mixed-effects model based on absolute agreement. Mean absolute difference (MAD), root mean squared difference (RMSD), and the number of states with prediction errors > 0.05 and > 0.1 compared with the mean observed TTO values were also employed. Higher ICC and lower MAD and RMSD values indicated better prediction accuracy. For each anchoring method, the prediction accuracy was compared between the anchoring results using own and other's TTO data. Comparisons of the anchored utilities for all 18,750 health states between the two samples were also conducted.

All statistical analyses were conducted using STATA 15.1 (StataCorp LLC, College Station, TX, USA). A two-tailed p-value < 0.05 was considered statistically significant.

## RESULTS

### *Characteristics of the study sample*

A total of 503 respondents (53.7% male; age range 18-86 years) were completed the tasks with 252 randomly assigned to the own TTO sample and 251 to the other's TTO sample. As shown in Table 1, The mean (SD) age was 45.2 [16.6] years for the own TTO sample vs. 45.6 [16.8] for the other's TTO sample (p = 0.830). The proportion of male respondents was 52.0% for the own TTO sample vs. 55.4% for the other's TTO sample (p = 0.445). Other demographic characteristics were also comparable between the two samples, except for employment status (p = 0.023). The mean (SD) self-reported SF-6Dv2 utility values were 0.793 (0.154) for the own TTO sample and 0.794 (0.176) for the other's TTO sample (p = 0.998).

### *Observed TTO values*

The mean (SD) TTO values were 0.406 (0.599) for the own TTO sample and 0.375 (0.634) for the other's TTO sample (p = 0.015). The difference in the mean TTO values between the two samples was statistically significant for all 115 states elicited from the TTO tasks. A comparison of observed TTO values for all 115 states between the two samples in terms of mean, SD, and median is reported in Supplementary Table 1. For the six mildest states, higher mean TTO values of the three of them (state 112111, 111121, 111211) were observed in the own TTO sample, with a range of difference from 0.002 to 0.014. The other three states showed a higher mean TTO value in the other's sample, while with a narrower range of the difference (0.001 to 0.007). Larger differences could be observed among moderate states (Supplementary Table 1), with a range from 0.002 (state 434454) to 0.900 (state 122631). The mean (SD) observed TTO value for the worst state was -0.259 (0.591) for the own TTO sample and -0.236 (0.616) for the other's TTO sample ($p < 0.001$). Bland–Altman plot (Figure 2) showed a comparable distribution above and below zero across the severity spectrum between the two samples. Five (4.3%) of 115 states lay out the 95% limits of agreement.

***Anchored values***

The estimated TTO values for the worst state were therefore -0.223 for the own TTO sample and -0.237 for the other's TTO sample (Table 2). The anchored coefficients of the anchoring methods (1) and (2) using the Eq. 2 and the modeled coefficients of the method (3) are shown in Supplementary Table 2.

As shown in Table 3, the ICCs of anchoring results using own TTO data were consistently higher (0.837 vs. 0.771, 0.835 vs. 0.771, and 0.873 vs. 0.804) than those using other's TTO data across all anchoring methods. Similar trend was also observed on MAD between the own and other's TTO samples (0.174 vs. 0.203, 0.181 vs. 0.203, and 0.127 vs. 0.146) and RMSD (0.228 vs. 0.270, 0.237 vs. 0.268, and 0.164 vs. 0.192) than in the other's TTO sample for the three methods, respectively (Table 3). The number of states with prediction error > 0.05 (21-25 vs. 21-26) and > 0.1 (61-72 vs. 64-75) also performed better in anchoring results using the own TTO data than using the other's TTO data (Table 3).

Comparisons of the anchored utility distributions between the two samples for the three anchoring methods are presented in Figure 3. For the anchoring method using the observed TTO value, the utility values anchored using the other's TTO data were higher than those using the own TTO data due to the higher observed TTO value for the worst state in the other's TTO sample. For the anchoring method using the estimated TTO value and mapping method, the utility values anchored using the other's TTO data were lower than those using the own TTO data. A larger difference between the two samples was observed for the mapping method than for methods using observed and estimated TTO values.

## DISCUSSION

Our study found that there were differences in DCE anchored utility values using own versus other's TTO data. Prediction accuracy was better with the use of own TTO across all three anchoring methods. The differences in the anchored utility values were larger when using the mapping approach compared with the anchoring approaches using the worst health state value.

The TTO responses were different between the two samples despite the randomization. The difference in moderate states was greater than in those mild and very severe states. This explains that the differences in anchored values were smaller when using the utility for worst health state than the mapping approach with 115 health states. Moreover, anchoring DCE values using the TTO value of a single health state was noticeably less precise than the mapping method.

In published EQ-5D-Y-3L valuation studies, different anchoring methods have been used. The method of anchoring using the TTO value of the worst state was used in Slovenia (28). Both methods of anchoring using the TTO value of the worst state and using the hybrid model were employed in Spain (27), in which a lower MSD for the hybrid model that included all health states was observed. The mapping method was used in Japan, in which the RMSD between the observed and estimated TTO values ranged from 0.040 (state 21111) to 0.284 (state 33333) (26), versus of 0.164 (across 115 states) in our study.

The international valuation protocol for the EQ-5D-Y-3L recommends the use of TTO values elicited from a separate, smaller sample for anchoring purposes. This approach does have a practical advantage but at the cost of prediction accuracy. The question is how much decrease in accuracy are we willing to accept in exchange for better feasibility of data collection?

No specific anchoring method is recommended in the international protocol (17), but this is an important and relevant factor. The prediction accuracy was better with the use of the mapping approach on other's TTO data than that of using single health state value based on own TTO data. This evidence supports the use of the mapping approach for the purpose of anchoring.

This study has several limitations. First, considering the relatively small number of health states evaluated given the large descriptive system of the SF-6Dv2, and the limited sample size in this study, there could be an impact on the statistical efficiency of the model estimation. Second, all 18 blocks of TTO tasks were assigned to both samples. However, 8 of the blocks were evaluated slightly more frequently in the own TTO sample while the remaining 10 blocks were evaluated more frequently in the other's TTO sample. This might have an impact on the comparison of utility values between the two samples. Third, this study was part of the SF-6Dv2 valuation study in which the adult's own perspective was used. This is different from the EQ-5D-Y-3L valuation tasks, in which the adults are asked to value health states by taking the perspective of a 10-year-old child.

## CONCLUSIONS

Anchored DCE values were different when using own TTO data compared with using other's TTO data. Prediction accuracy was better with the use of own TTO across all three anchoring methods. The mapping method has better prediction accuracy than the anchoring methods using single health state value.

## References:


1. Brazier J, Ratcliffe J, Saloman J, et al. *Measuring and valuing health benefits for economic evaluation*. OXFORD university press, 2017.

2. Neumann PJ, Sanders GD, Russell LB, et al. *Cost-effectiveness in health and medicine*. Oxford University Press, 2016.

3. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *Journal of health economics*. 2002; 21: 271-92.

4. Oppe M, Rand-Hendriksen K, Shah K, et al. EuroQol Protocols for Time Trade-Off Valuation of Health Outcomes. *PharmacoEconomics*. 2016; 34: 993-1004.

5. Dolan P. Modeling valuations for EuroQol health states. *Medical care*. 1997; 35: 1095-108.

6. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health economics*. 2002; 11: 447-56.

7. Brazier J, Rowen D, Yang Y, et al. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. *The European journal of health economics*. 2012; 13: 575-87.

8. 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 in Health*. 2020.

9. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health economics*. 2012; 21: 145-72.

10. Clark MD, Determann D, Petrou S, et al. Discrete choice experiments in health economics: a review of the literature. PharmacoEconomics. 2014; 32: 883-902.

11. Mulhern B, Norman R, Street DJ, et al. One Method, Many Methodological Choices: A Structured Review of Discrete-Choice Experiments for Health State Valuation. *PharmacoEconomics*. 2019; 37: 29-43.

12. Stolk EA, Oppe M, Scalone L, et al. Discrete Choice Modeling for the Quantification of Health States: The Case of the EQ-5D. *Value in Health*. 2010; 13: 1005-13.

13. Bansback N, Brazier J, Tsuchiya A, et al. Using a discrete choice experiment to estimate health state utility values. *Journal of health economics*. 2012; 31: 306-18.

14. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Medical decision making*. 2015; 35: 328-40.

15. Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. *Applied health economics and health policy*. 2013; 11: 287-98.

16. Kreimeier S, Greiner W. EQ-5D-Y as a Health-Related Quality of Life Instrument for Children and Adolescents: The Instrument's Characteristics, Development, Current Use, and Challenges of Developing Its Value Set. *Value in Health*. 2019; 22: 31-37.

17. Ramos-Goñi JM, Oppe M, Stolk E, et al. International Valuation Protocol for the EQ-5D-Y-3L. *PharmacoEconomics*. 2020; 38: 653-63.

18. 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. *Medical care*. 2020; 58: 557-65.

19. Wu J, Xie S, He X, et al. The Simplified Chinese version of SF-6Dv2: translation, cross-cultural adaptation and preliminary psychometric testing. *Quality of life research*. 2020;29:1385–1391.

20. Janssen BM, Oppe M, Versteegh MM, et al. Introducing the composite time trade-off: a test of feasibility and face validity. *The European journal of health economics*. 2013; 14 Suppl 1: S5-13.

21. 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: 521-35.

22. Chrzan K, Orme B. An overview and comparison of design strategies for choice-based conjoint analysis. *Sawtooth software research paper series*. 2000; 98382.

23. Marshall DA, Deal K, Bombard Y, et al. How do women trade-off benefits and risks in chemotherapy treatment decisions based on gene expression profiling for early-stage breast cancer? A discrete choice experiment. *BMJ open*. 2016; 6: e010981.

24. Norman R, Viney R, Brazier J, et al. Valuing SF-6D Health States Using a Discrete Choice Experiment. *Medical decision making*. 2014; 34: 773-86.

25. Chen G, Khan MA, Iezzi A, et al. Mapping between 6 Multiattribute Utility Instruments. *Medical decision making*. 2016; 36: 160-75.

26. Shiroiwa T, Ikeda S, Noto S, et al. Valuation Survey of EQ-5D-Y Based on the International Common Protocol: Development of a Value Set in Japan. *Medical Decision Making*. 2021: 0272989X2110018.

27. Ramos-Goñi JM, Oppe M, Estévez-Carrillo A, et al. Accounting for Unobservable Preference Heterogeneity and Evaluating Alternative Anchoring Approaches to Estimate Country-Specific EQ-5D-Y Value Sets: A Case Study Using Spanish Preference Data. *Value in Health*. 2021.

28. Rupel VP, Ogorevc M, Greiner W, et al. EQ-5D-Y Value Set for Slovenia. *PharmacoEconomics*. 2021; 39: 463-71.

## Table 1 Characteristics of respondents

| Characteristics | Own TTO sample (N=252) N (%) | Other's TTO sample (N=251) N (%) | P-value [a] |
|---|---|---|---|
| **Male** | 131 (52.0%) | 139 (55.4%) | 0.445 |
| **Age (mean [SD])** | 45.2 (16.6) | 45.6 (16.8) | 0.830 |
| **Age group (y)** | | | 0.934 |
| 18-29 | 50 (19.8%) | 53 (21.2%) | |
| 30-39 | 52 (20.6%) | 48 (19.1%) | |
| 40-49 | 47 (18.7%) | 41 (16.3%) | |
| 50-59 | 46 (18.3%) | 48 (19.1%) | |
| ≥ 60 | 57 (22.6%) | 61 (24.3%) | |
| **Education** | | | 0.929 |
| Primary or lower | 46 (18.3%) | 47 (18.7%) | |
| Junior high school | 82 (32.5%) | 87 (34.7%) | |
| Senior high school | 58 (23.0%) | 57 (22.7%) | |
| College or higher | 66 (26.2%) | 60 (23.9%) | |
| **Household registration** | | | 0.653 |
| Urban | 170 (67.5%) | 174 (69.3%) | |
| Rural | 82 (32.5%) | 77 (30.7%) | |
| **Marital status** | | | 0.658 |
| Single | 55 (21.8%) | 56 (22.3%) | |
| Married | 176 (69.8%) | 176 (70.1%) | |
| Divorced | 6 (2.4%) | 9 (3.6%) | |
| Widowed | 15 (6.0%) | 10 (4.0%) | |
| **Employment status** | | | **0.023** |
| Employed | 162 (64.4%) | 135 (53.7%) | |
| Retired | 52 (20.6%) | 73 (29.1%) | |
| Student | 19 (7.5%) | 30 (12.0%) | |
| Unemployed | 19 (7.5%) | 13 (5.2%) | |
| **Monthly income (RMB)** | | | 0.117 |
| < ¥2000 | 43 (17.1%) | 63 (25.1%) | |
| ¥2000-5000 | 151 (59.9%) | 142 (56.6%) | |
| ¥5000-10000 | 42 (16.7%) | 36 (14.3%) | |
| > ¥10000 | 16 (6.3%) | 10 (4.0%) | |
| **Self-reported SF-6Dv2 utility** | 0.793 (0.154) | 0.794 (0.176) | 0.998 |
| **Number of chronic conditions [b]** | | | 0.331 |
| 0 | 154 (61.1%) | 140 (55.8%) | |
| 1 | 56 (22.2%) | 68 (27.1%) | |
| 2 | 22 (8.7%) | 22 (8.8%) | |
| 3 or more | 20 (8.0%) | 21 (8.3%) | |

[a] The comparison of characteristics distributions between the two samples by t-test or $chi^2$ test as appropriate.

[b] The chronic conditions include: Hypertension, dyslipidemia, diabetes or high blood sugar, cancer or malignant tumor, chronic lung disease, liver disease, heart disease, stroke, kidney disease, stomach or other digestive disease, emotional or psychatric problems, memory-related disease, arthritis or rheumatism, asthma, or other respondent-reported chronic conditions.

## Table 2 Model coefficients of DCE and TTO data

| | Own TTO sample (N=252) | | | | Other's TTO sample (N=251) | |
|---|---|---|---|---|---|---|
| | DCE data Conditional logit model [a] | | TTO data Fixed effects model | | TTO data Fixed effects model | |
| | Coef. | SE | Coef. | SE | Coef. | SE |
| PF2 | -0.188 | 0.105 | -0.037 | 0.033 | -0.032 | 0.032 |
| PF3 | -0.260* | 0.102 | -0.032 | 0.034 | -0.047 | 0.035 |
| PF4 | -0.427*** | 0.108 | -0.122*** | 0.030 | -0.151*** | 0.032 |
| PF5 | -1.806*** | 0.133 | -0.406*** | 0.031 | -0.418*** | 0.031 |
| RL2 | -0.031 | 0.104 | -0.034 | 0.030 | -0.048 | 0.031 |
| RL3 | -0.143 | 0.105 | -0.058 | 0.033 | -0.054 | 0.032 |
| RL4 | -0.186 | 0.103 | -0.065* | 0.033 | -0.065 | 0.034 |
| RL5 | -0.518*** | 0.115 | -0.094** | 0.033 | -0.076* | 0.033 |
| SF2 | 0.000 | -- | -0.094** | 0.029 | -0.119*** | 0.030 |
| SF3 | 0.000 | -- | -0.122*** | 0.032 | -0.104** | 0.031 |
| SF4 | -0.453*** | 0.091 | -0.110*** | 0.031 | -0.143*** | 0.031 |
| SF5 | -0.535*** | 0.089 | -0.113*** | 0.031 | -0.121*** | 0.032 |
| PN2 | 0.000 | -- | -0.079* | 0.032 | -0.087** | 0.033 |
| PN3 | -0.079 | 0.080 | -0.080* | 0.033 | -0.099** | 0.034 |
| PN4 | -0.079 | 0.080 | -0.053 | 0.034 | -0.109** | 0.033 |
| PN5 | -1.328*** | 0.113 | -0.333*** | 0.034 | -0.331*** | 0.036 |
| PN6 | -1.700*** | 0.127 | -0.327*** | 0.034 | -0.380*** | 0.033 |
| MH2 | -0.044 | 0.111 | -0.053 | 0.029 | -0.022 | 0.029 |
| MH3 | -0.217 | 0.112 | -0.111** | 0.033 | -0.126*** | 0.035 |
| MH4 | -0.676*** | 0.099 | -0.124*** | 0.032 | -0.118*** | 0.031 |
| MH5 | -0.676*** | 0.099 | -0.147*** | 0.032 | -0.126*** | 0.032 |
| VT2 | 0.000 | -- | -0.067* | 0.030 | -0.075* | 0.032 |
| VT3 | 0.000 | -- | -0.068* | 0.033 | -0.056 | 0.033 |
| VT4 | -0.350*** | 0.080 | -0.111*** | 0.031 | -0.116*** | 0.032 |
| VT5 | -0.553*** | 0.083 | -0.136*** | 0.033 | -0.116*** | 0.032 |
| Value of the worst state | -5.788 | | -1.223 | | -1.237 | |
| Log-Likelihood | -2371.623 | | -762.051 | | -809.484 | |
| AIC | 4794.458 | | 1567.722 | | 1671.595 | |
| BIC | 4962.925 | | 1679.900 | | 1778.088 | |

[a] In DCE model, levels 1 to 3 of SF dimension, levels 1 to 2 of PN dimension, levels 4 to 5 of MH dimension, and levels 1 to 3 of VT dimension were combined to produce a fully monotonic model.

**Abbr:** AIC, Akaike information criterion; BIC, Bayesian information criterion. PF, physical functioning; RL, role limitation; SF, social functioning; PN, pain; MH, mental health; VT, Vitality.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

**Table 3 Comparison of the prediction accuracy between anchoring results using own versus other's TTO data**

| | **(1) Anchoring with observed TTO value of the worst state** | | **(2) Anchoring with estimated TTO value of the worst state** | | **(3) Mapping DCE onto TTO** | |
|---|---|---|---|---|---|---|
| | Own TTO (N=252) | Other's TTO (N=251) | Own TTO (N=252) | Other's TTO (N=251) | Own TTO (N=252) | Other's TTO (N=251) |
| Range | (1, -0.259) | (1, -0.236) | (1, -0.223) | (1, -0.237) | (0.824, -0.380) | (0.812, -0.434) |
| ICC | 0.837 | 0.771 | 0.835 | 0.771 | 0.873 | 0.804 |
| No. of predictions >0.05 from observed TTO | 24 (20.9%) | 26 (22.6%) | 21 (18.3%) | 23 (20.0%) | 25 (21.7%) | 21 (18.3%) |
| No. predictions >0.1 from observed TTO | 69 (60.0%) | 75 (65.2%) | 72 (62.6%) | 75 (65.2%) | 61 (53.0%) | 64 (55.7%) |
| MAD | 0.174 | 0.203 | 0.181 | 0.203 | 0.127 | 0.146 |
| RMSD | 0.228 | 0.270 | 0.237 | 0.268 | 0.164 | 0.192 |

**Abbr:** DCE, discrete choice experiment; TTO, time trade-off; MAD, mean absolute difference; RMSD, root mean squared difference.

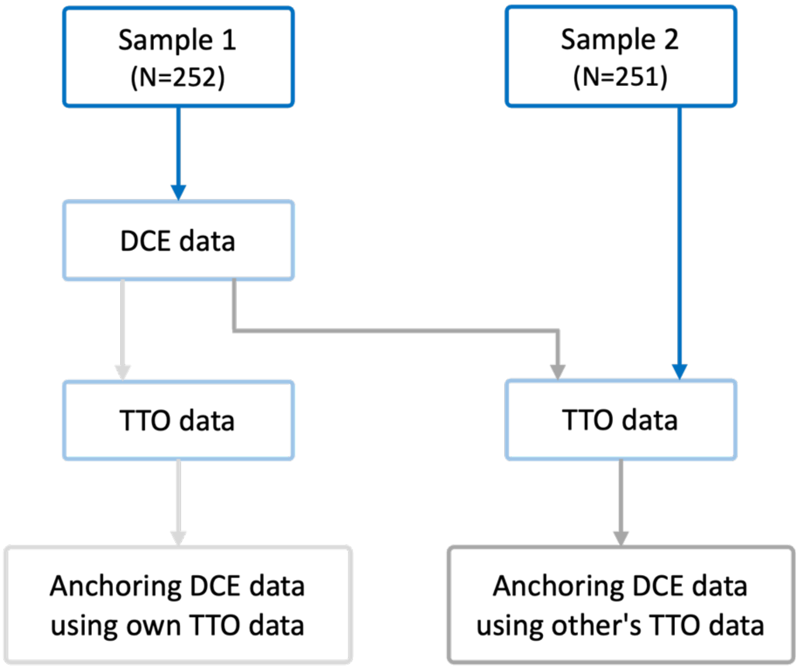


**Figure 1 Flow chart of the data included in this study**

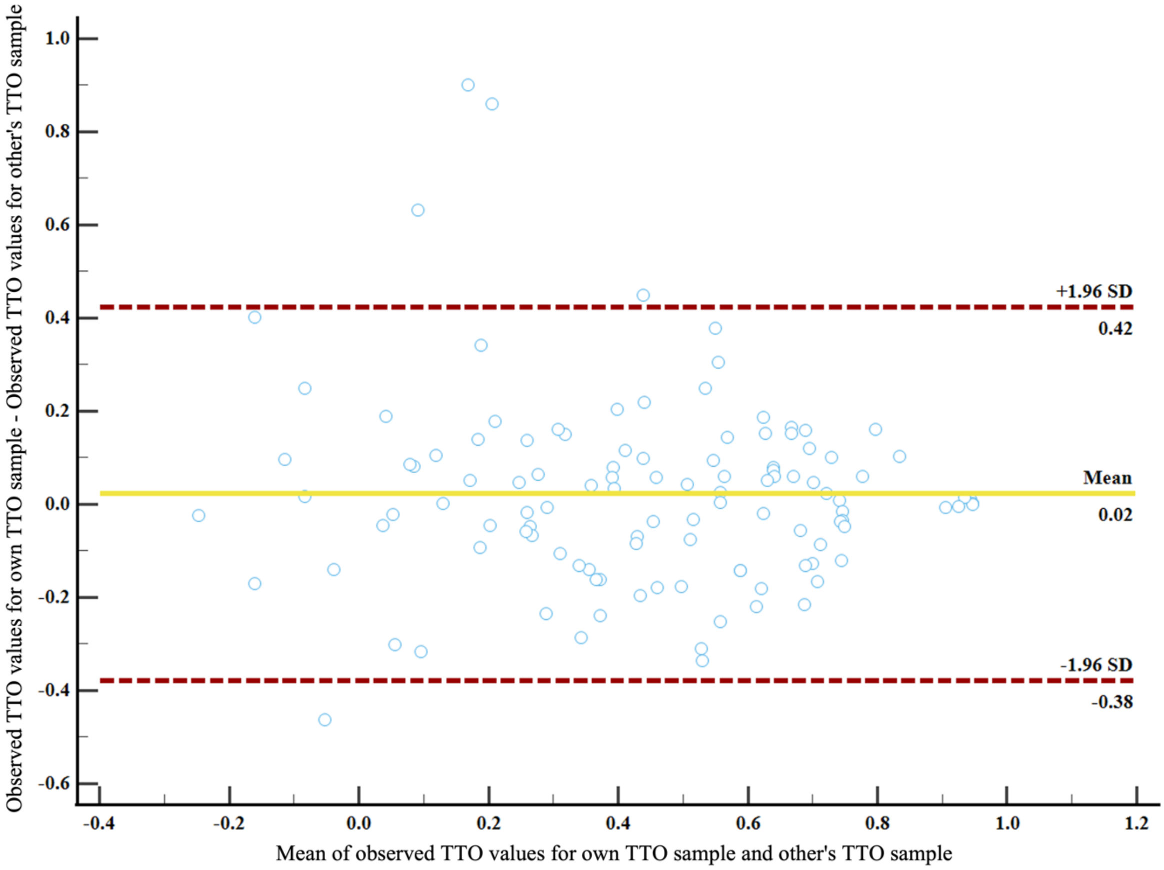


**Figure 2 Bland-Altman plots of observed TTO values between the two samples**

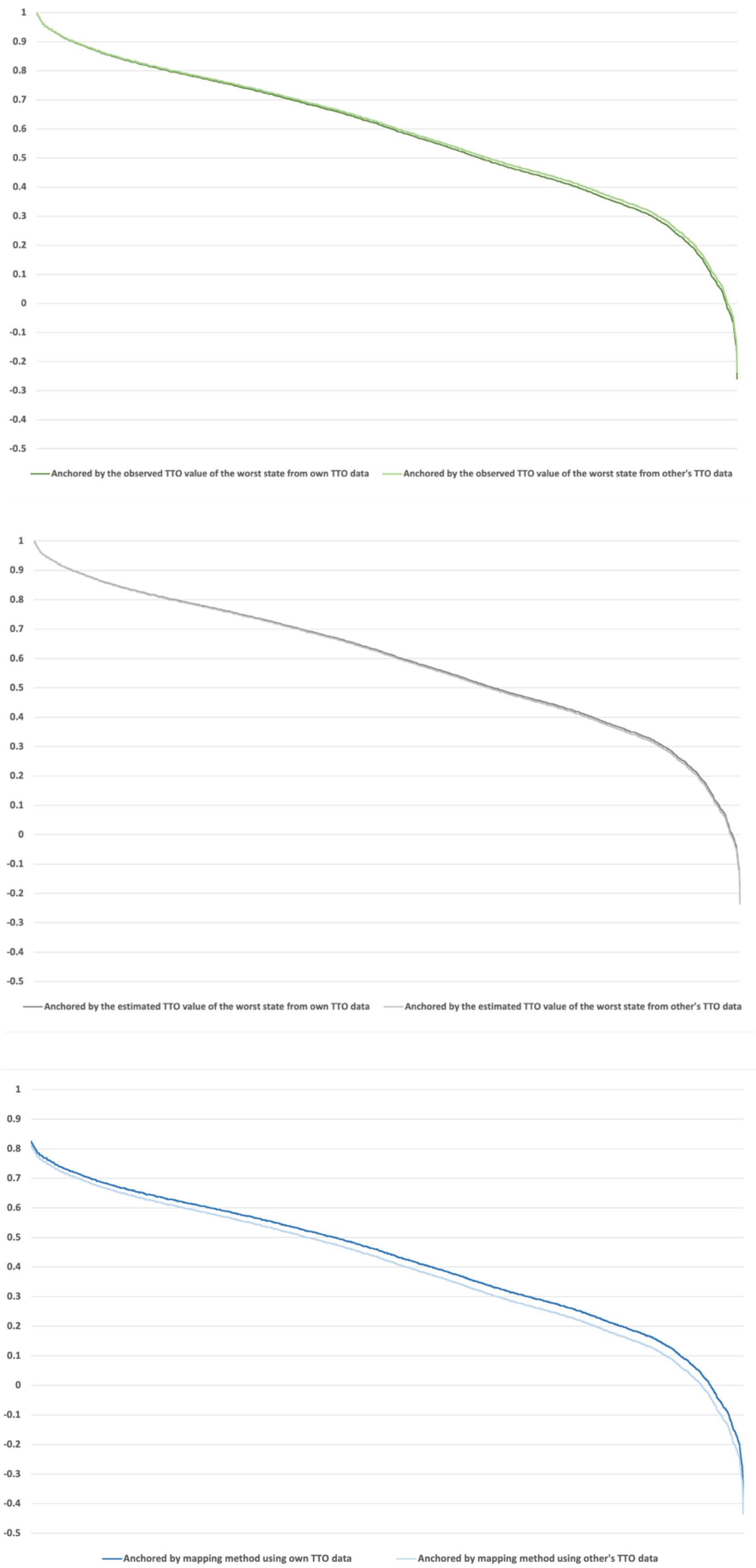


**Figure 3 Distributions of anchored utility values between the two samples**

**Supplementary Table 1 TTO responses of 115 health states between the two samples**

| state | Own TTO sample (N=252) | | | | Other's TTO sample (N=251) | | | | Mean difference |
|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | SD | Median | N | Mean | SD | Median | |
| 111112 | 52 | 0.946 | 0.028 | 0.95 | 44 | 0.947 | 0.020 | 0.95 | 0.001 |
| 111121 | 47 | 0.948 | 0.021 | 0.95 | 34 | 0.946 | 0.019 | 0.95 | 0.002 |
| 111211 | 51 | 0.942 | 0.025 | 0.95 | 36 | 0.928 | 0.075 | 0.95 | 0.014 |
| 112111 | 37 | 0.949 | 0.008 | 0.95 | 35 | 0.937 | 0.053 | 0.95 | 0.012 |
| 121111 | 37 | 0.923 | 0.089 | 0.95 | 51 | 0.927 | 0.045 | 0.95 | -0.004 |
| 211111 | 28 | 0.902 | 0.081 | 0.95 | 51 | 0.909 | 0.082 | 0.95 | -0.007 |
| 111313 | 14 | 0.886 | 0.113 | 0.95 | 18 | 0.783 | 0.235 | 0.9 | 0.102 |
| 112352 | 14 | 0.671 | 0.264 | 0.7 | 18 | 0.611 | 0.242 | 0.65 | 0.060 |
| 113331 | 16 | 0.725 | 0.473 | 0.925 | 9 | 0.761 | 0.193 | 0.85 | -0.036 |
| 114244 | 12 | 0.733 | 0.142 | 0.775 | 11 | 0.709 | 0.304 | 0.85 | 0.024 |
| 115225 | 7 | 0.729 | 0.175 | 0.7 | 8 | 0.763 | 0.201 | 0.825 | -0.034 |
| 121252 | 17 | 0.579 | 0.605 | 0.8 | 9 | 0.794 | 0.107 | 0.8 | -0.215 |
| 122631 | 14 | 0.618 | 0.206 | 0.575 | 14 | -0.282 | 0.795 | -0.65 | 0.900 |
| 123644 | 17 | 0.179 | 0.614 | 0.3 | 12 | 0.225 | 0.614 | 0.55 | -0.046 |
| 124625 | 23 | 0.470 | 0.420 | 0.55 | 15 | 0.353 | 0.437 | 0.4 | 0.116 |
| 125213 | 17 | 0.624 | 0.484 | 0.8 | 9 | 0.756 | 0.228 | 0.85 | -0.132 |
| 131325 | 14 | 0.768 | 0.268 | 0.875 | 18 | 0.608 | 0.299 | 0.65 | 0.160 |
| 133552 | 12 | -0.063 | 0.720 | 0.25 | 17 | 0.253 | 0.644 | 0.5 | -0.315 |
| 134531 | 17 | 0.371 | 0.560 | 0.5 | 9 | 0.550 | 0.247 | 0.55 | -0.179 |
| 135544 | 7 | 0.286 | 0.727 | 0.55 | 8 | 0.425 | 0.522 | 0.6 | -0.139 |
| 141131 | 19 | 0.779 | 0.252 | 0.85 | 9 | 0.678 | 0.632 | 0.9 | 0.101 |
| 142144 | 10 | 0.725 | 0.296 | 0.8 | 16 | 0.678 | 0.459 | 0.8 | 0.047 |
| 143125 | 12 | 0.504 | 0.498 | 0.65 | 11 | 0.723 | 0.284 | 0.8 | -0.219 |
| 144613 | 12 | 0.171 | 0.727 | 0.375 | 17 | 0.406 | 0.547 | 0.55 | -0.235 |
| 145152 | 19 | 0.571 | 0.342 | 0.65 | 9 | 0.544 | 0.627 | 0.8 | 0.027 |
| 152425 | 12 | 0.200 | 0.619 | 0.425 | 17 | 0.485 | 0.587 | 0.65 | -0.285 |
| 153413 | 14 | 0.739 | 0.205 | 0.775 | 14 | 0.361 | 0.671 | 0.525 | 0.379 |
| 154452 | 17 | 0.594 | 0.253 | 0.55 | 12 | 0.533 | 0.522 | 0.675 | 0.061 |
| 155431 | 14 | 0.704 | 0.154 | 0.675 | 18 | 0.550 | 0.189 | 0.5 | 0.154 |
| 211441 | 7 | 0.614 | 0.184 | 0.5 | 21 | 0.633 | 0.402 | 0.75 | -0.019 |
| 212424 | 16 | 0.669 | 0.527 | 0.9 | 9 | 0.756 | 0.133 | 0.75 | -0.087 |
| 213415 | 12 | 0.363 | 0.679 | 0.575 | 17 | 0.697 | 0.244 | 0.75 | -0.335 |
| 221524 | 16 | 0.272 | 0.642 | 0.525 | 17 | 0.224 | 0.817 | 0.6 | 0.048 |
| 222515 | 9 | 0.411 | 0.580 | 0.6 | 19 | 0.376 | 0.514 | 0.5 | 0.035 |
| 223553 | 17 | 0.241 | 0.646 | 0.35 | 12 | 0.288 | 0.581 | 0.45 | -0.046 |
| 224532 | 16 | 0.288 | 0.611 | 0.45 | 9 | 0.294 | 0.558 | 0.4 | -0.007 |
| 225341 | 12 | 0.500 | 0.535 | 0.625 | 11 | 0.532 | 0.595 | 0.75 | -0.032 |
| 231132 | 7 | 0.879 | 0.064 | 0.9 | 21 | 0.717 | 0.281 | 0.85 | 0.162 |
| 232141 | 16 | 0.684 | 0.214 | 0.6 | 11 | 0.805 | 0.198 | 0.9 | -0.120 |

| state | Own TTO sample (N=252) | | | | Other's TTO sample (N=251) | | | | Mean difference |
|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | SD | Median | N | Mean | SD | Median | |
| 233624 | 7 | 0.121 | 0.718 | 0.35 | 21 | 0.036 | 0.733 | 0.35 | 0.086 |
| 234115 | 12 | 0.517 | 0.584 | 0.775 | 17 | 0.659 | 0.479 | 0.8 | -0.142 |
| 235153 | 15 | 0.677 | 0.242 | 0.75 | 18 | 0.603 | 0.270 | 0.625 | 0.074 |
| 241315 | 9 | 0.700 | 0.200 | 0.7 | 19 | 0.639 | 0.329 | 0.7 | 0.061 |
| 242353 | 17 | 0.432 | 0.573 | 0.6 | 9 | 0.683 | 0.290 | 0.75 | -0.251 |
| 243232 | 7 | 0.636 | 0.295 | 0.65 | 8 | 0.763 | 0.158 | 0.75 | -0.127 |
| 244241 | 10 | 0.755 | 0.186 | 0.775 | 16 | 0.634 | 0.453 | 0.7 | 0.121 |
| 245324 | 10 | 0.625 | 0.506 | 0.85 | 16 | 0.791 | 0.161 | 0.85 | -0.166 |
| 251653 | 7 | 0.379 | 0.614 | 0.55 | 8 | 0.338 | 0.439 | 0.525 | 0.041 |
| 253641 | 10 | 0.420 | 0.576 | 0.575 | 16 | 0.363 | 0.595 | 0.5 | 0.058 |
| 254224 | 17 | 0.374 | 0.677 | 0.55 | 9 | 0.683 | 0.255 | 0.75 | -0.310 |
| 255215 | 7 | 0.336 | 0.614 | 0.55 | 8 | 0.531 | 0.502 | 0.675 | -0.196 |
| 313242 | 12 | 0.292 | 0.672 | 0.525 | 17 | 0.453 | 0.673 | 0.75 | -0.161 |
| 314221 | 9 | 0.744 | 0.216 | 0.8 | 19 | 0.737 | 0.145 | 0.75 | 0.008 |
| 322642 | 19 | 0.274 | 0.698 | 0.55 | 9 | 0.406 | 0.622 | 0.6 | -0.132 |
| 323121 | 7 | 0.807 | 0.110 | 0.8 | 21 | 0.748 | 0.191 | 0.8 | 0.060 |
| 324114 | 17 | 0.738 | 0.212 | 0.75 | 12 | 0.754 | 0.144 | 0.775 | -0.016 |
| 325155 | 9 | 0.394 | 0.804 | 0.65 | 19 | 0.463 | 0.411 | 0.55 | -0.069 |
| 331314 | 15 | 0.680 | 0.177 | 0.6 | 18 | 0.600 | 0.171 | 0.6 | 0.080 |
| 332255 | 17 | 0.435 | 0.594 | 0.65 | 9 | 0.472 | 0.626 | 0.75 | -0.037 |
| 333233 | 16 | 0.653 | 0.272 | 0.625 | 11 | 0.709 | 0.271 | 0.8 | -0.056 |
| 334342 | 23 | 0.641 | 0.204 | 0.65 | 15 | 0.497 | 0.461 | 0.55 | 0.145 |
| 335321 | 16 | 0.488 | 0.590 | 0.65 | 17 | 0.388 | 0.705 | 0.65 | 0.099 |
| 341442 | 9 | 0.594 | 0.625 | 0.85 | 19 | 0.500 | 0.404 | 0.55 | 0.094 |
| 343414 | 7 | 0.529 | 0.191 | 0.5 | 21 | 0.486 | 0.321 | 0.6 | 0.043 |
| 344455 | 14 | 0.636 | 0.218 | 0.575 | 14 | -0.225 | 0.715 | 0.075 | 0.861 |
| 345433 | 16 | 0.656 | 0.284 | 0.725 | 10 | 0.605 | 0.274 | 0.7 | 0.051 |
| 351521 | 16 | 0.409 | 0.574 | 0.55 | 11 | 0.586 | 0.436 | 0.7 | -0.177 |
| 352514 | 14 | 0.407 | 0.639 | 0.6 | 14 | -0.225 | 0.799 | -0.275 | 0.632 |
| 353555 | 23 | 0.393 | 0.371 | 0.4 | 15 | 0.243 | 0.585 | 0.4 | 0.150 |
| 354333 | 15 | 0.140 | 0.614 | 0.4 | 18 | 0.233 | 0.539 | 0.4 | -0.093 |
| 355542 | 17 | 0.300 | 0.420 | 0.4 | 12 | 0.121 | 0.614 | 0.4 | 0.179 |
| 411522 | 17 | 0.488 | 0.481 | 0.5 | 12 | 0.429 | 0.503 | 0.55 | 0.059 |
| 412511 | 12 | 0.308 | 0.643 | 0.575 | 17 | 0.244 | 0.606 | 0.5 | 0.064 |
| 413354 | 19 | 0.474 | 0.435 | 0.5 | 9 | 0.550 | 0.626 | 0.85 | -0.076 |
| 414535 | 12 | -0.096 | 0.675 | 0.15 | 17 | 0.206 | 0.616 | 0.35 | -0.302 |
| 415543 | 16 | 0.234 | 0.554 | 0.45 | 9 | 0.300 | 0.506 | 0.4 | -0.066 |
| 421211 | 16 | 0.725 | 0.491 | 0.925 | 9 | 0.772 | 0.236 | 0.9 | -0.047 |
| 423335 | 23 | 0.500 | 0.211 | 0.5 | 15 | 0.297 | 0.500 | 0.35 | 0.203 |
| 424343 | 16 | 0.659 | 0.259 | 0.675 | 10 | 0.410 | 0.652 | 0.625 | 0.249 |

| state | Own TTO sample (N=252) | | | | Other's TTO sample (N=251) | | | | Mean difference |
|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | SD | Median | N | Mean | SD | Median | |
| 425322 | 16 | 0.750 | 0.181 | 0.8 | 10 | 0.585 | 0.451 | 0.725 | 0.165 |
| 431443 | 10 | 0.560 | 0.502 | 0.675 | 16 | 0.556 | 0.503 | 0.775 | 0.004 |
| 432422 | 23 | 0.717 | 0.161 | 0.7 | 15 | 0.530 | 0.438 | 0.6 | 0.187 |
| 433411 | 12 | 0.517 | 0.584 | 0.775 | 17 | 0.659 | 0.215 | 0.55 | -0.142 |
| 434454 | 16 | 0.131 | 0.695 | 0.375 | 17 | 0.129 | 0.782 | 0.5 | 0.002 |
| 441654 | 19 | 0.126 | 0.742 | 0.55 | 9 | 0.044 | 0.857 | 0.3 | 0.082 |
| 442235 | 23 | 0.550 | 0.238 | 0.55 | 15 | 0.330 | 0.532 | 0.4 | 0.220 |
| 443243 | 16 | 0.531 | 0.514 | 0.675 | 9 | 0.711 | 0.187 | 0.75 | -0.180 |
| 444622 | 16 | 0.041 | 0.574 | 0.125 | 17 | -0.209 | 0.850 | -0.7 | 0.249 |
| 445611 | 10 | 0.285 | 0.642 | 0.5 | 16 | 0.447 | 0.484 | 0.525 | -0.162 |
| 451635 | 16 | 0.253 | 0.590 | 0.4 | 11 | 0.491 | 0.378 | 0.5 | -0.238 |
| 452143 | 7 | 0.707 | 0.174 | 0.75 | 21 | 0.402 | 0.551 | 0.65 | 0.305 |
| 453122 | 14 | 0.664 | 0.300 | 0.7 | 14 | 0.214 | 0.831 | 0.525 | 0.450 |
| 454111 | 16 | 0.744 | 0.202 | 0.8 | 10 | 0.590 | 0.436 | 0.575 | 0.154 |
| 455254 | 15 | 0.137 | 0.635 | 0.3 | 18 | -0.053 | 0.770 | 0.35 | 0.189 |
| 511134 | 16 | 0.388 | 0.441 | 0.525 | 11 | 0.227 | 0.755 | 0.5 | 0.160 |
| 513123 | 16 | -0.075 | 0.710 | 0.025 | 17 | -0.091 | 0.896 | 0.2 | 0.016 |
| 514112 | 14 | 0.432 | 0.490 | 0.55 | 18 | 0.353 | 0.452 | 0.35 | 0.079 |
| 521445 | 12 | 0.329 | 0.460 | 0.35 | 11 | 0.191 | 0.746 | 0.4 | 0.138 |
| 522223 | 15 | 0.253 | 0.532 | 0.55 | 18 | 0.114 | 0.719 | 0.55 | 0.139 |
| 524451 | 7 | 0.386 | 0.279 | 0.5 | 8 | 0.469 | 0.415 | 0.6 | -0.083 |
| 525434 | 16 | 0.359 | 0.430 | 0.4 | 11 | 0.018 | 0.733 | 0.4 | 0.341 |
| 531251 | 9 | -0.283 | 0.644 | -0.2 | 19 | 0.179 | 0.530 | 0.5 | -0.462 |
| 532234 | 16 | 0.197 | 0.606 | 0.325 | 10 | 0.145 | 0.746 | 0.5 | 0.052 |
| 533645 | 12 | -0.246 | 0.545 | -0.075 | 17 | -0.076 | 0.678 | 0.25 | -0.169 |
| 534623 | 15 | 0.040 | 0.557 | 0.25 | 18 | -0.361 | 0.550 | -0.475 | 0.401 |
| 535612 | 12 | 0.042 | 0.662 | 0.3 | 11 | 0.064 | 0.765 | -0.05 | -0.022 |
| 541523 | 14 | 0.229 | 0.538 | 0.375 | 18 | 0.286 | 0.262 | 0.225 | -0.058 |
| 542312 | 19 | 0.171 | 0.661 | 0.4 | 9 | 0.067 | 0.817 | 0.4 | 0.104 |
| 543551 | 16 | 0.016 | 0.633 | 0.225 | 10 | 0.060 | 0.762 | 0.375 | -0.044 |
| 545545 | 14 | -0.300 | 0.691 | -0.55 | 14 | -0.571 | 0.590 | -1 | 0.271 |
| 551212 | 12 | 0.250 | 0.785 | 0.65 | 17 | 0.268 | 0.666 | 0.6 | -0.018 |
| 552351 | 16 | -0.066 | 0.666 | 0.025 | 17 | -0.162 | 0.849 | -0.5 | 0.096 |
| 553334 | 12 | 0.258 | 0.621 | 0.5 | 11 | 0.364 | 0.699 | 0.55 | -0.105 |
| 554345 | 12 | -0.108 | 0.730 | 0.225 | 17 | 0.032 | 0.706 | 0.35 | -0.141 |
| 555655 | 252 | -0.259 | 0.591 | -0.175 | 251 | -0.236 | 0.616 | -0.1 | -0.023 |

**Note:** The mean TTO value was statistically significantly different ($p < 0.001$) between two subsamples across all 115 states using the Mann-Whitney test.

**Supplementary Table 2 Model coefficients of the three anchoring methods**

| | Own TTO sample (N=252) | | | | Other's TTO sample (N=251) | |
|---|---|---|---|---|---|---|
| | DCE data | | Anchoring using own TTO data | | Anchoring using other's TTO data | |
| | Coef. | SE | Method (1) | Method (2) | Method (1) | Method (2) |
| PF2 | -0.188 | 0.105 | -0.041 | -0.040 | -0.040 | -0.040 |
| PF3 | -0.260* | 0.102 | -0.057 | -0.055 | -0.056 | -0.056 |
| PF4 | -0.427*** | 0.108 | -0.093 | -0.090 | -0.091 | -0.091 |
| PF5 | -1.806*** | 0.133 | -0.393 | -0.382 | -0.386 | -0.386 |
| RL2 | -0.031 | 0.104 | -0.007 | -0.006 | -0.007 | -0.007 |
| RL3 | -0.143 | 0.105 | -0.031 | -0.030 | -0.030 | -0.030 |
| RL4 | -0.186 | 0.103 | -0.040 | -0.039 | -0.040 | -0.040 |
| RL5 | -0.518*** | 0.115 | -0.113 | -0.109 | -0.111 | -0.111 |
| SF2 | 0.000 | -- | 0.000 | 0.000 | 0.000 | 0.000 |
| SF3 | 0.000 | -- | 0.000 | 0.000 | 0.000 | 0.000 |
| SF4 | -0.453*** | 0.091 | -0.099 | -0.096 | -0.097 | -0.097 |
| SF5 | -0.535*** | 0.089 | -0.116 | -0.113 | -0.114 | -0.114 |
| PN2 | 0.000 | -- | 0.000 | 0.000 | 0.000 | 0.000 |
| PN3 | -0.079 | 0.080 | -0.017 | -0.017 | -0.017 | -0.017 |
| PN4 | -0.079 | 0.080 | -0.017 | -0.017 | -0.017 | -0.017 |
| PN5 | -1.328*** | 0.113 | -0.289 | -0.281 | -0.284 | -0.284 |
| PN6 | -1.700*** | 0.127 | -0.370 | -0.359 | -0.363 | -0.363 |
| MH2 | -0.044 | 0.111 | -0.010 | -0.009 | -0.009 | -0.009 |
| MH3 | -0.217 | 0.112 | -0.047 | -0.046 | -0.046 | -0.046 |
| MH4 | -0.676*** | 0.099 | -0.147 | -0.143 | -0.144 | -0.145 |
| MH5 | -0.676*** | 0.099 | -0.147 | -0.143 | -0.144 | -0.145 |
| VT2 | 0.000 | -- | 0.000 | 0.000 | 0.000 | 0.000 |
| VT3 | 0.000 | -- | 0.000 | 0.000 | 0.000 | 0.000 |
| VT4 | -0.350*** | 0.080 | -0.076 | -0.074 | -0.075 | -0.075 |
| VT5 | -0.553*** | 0.083 | -0.120 | -0.117 | -0.118 | -0.118 |
| | | | Method (3) | | Method (3) | |
| | | | Coef. | SE | Coef. | SE |
| DCE latent utility value | | | 0.208*** | 0.013 | 0.215*** | 0.016 |
| Constant | | | 0.824*** | 0.029 | 0.812*** | 0.034 |

**Note:** In DCE model, levels 1 to 3 of SF dimension, levels 1 to 2 of PN dimension, levels 4 to 5 of MH dimension, and levels 1 to 3 of VT dimension were combined to produce a fully monotonic model.

**Abbr:** PF, physical functioning; RL, role limitation; SF, social functioning; PN, pain; MH, mental health; VT, vitality.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.