



This is a repository copy of *Valuation of SF-6Dv2 health states in China using time trade-off and discrete-choice experiment with a duration dimension*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/171683/>

Version: Published Version

Article:

Wu, J., Xie, S., He, X. et al. (9 more authors) (2021) Valuation of SF-6Dv2 health states in China using time trade-off and discrete-choice experiment with a duration dimension. *PharmacoEconomics*, 39 (5). pp. 521-535. ISSN 1170-7690

<https://doi.org/10.1007/s40273-020-00997-1>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



Valuation of SF-6Dv2 Health States in China Using Time Trade-off and Discrete-Choice Experiment with a Duration Dimension

Jing Wu^{1,2} · Shitong Xie^{1,2} · Xiaoning He^{1,2} · Gang Chen³ · Gengliang Bai⁴ · Da Feng⁵ · Ming Hu⁶ · Jie Jiang⁷ · Xiaohui Wang⁸ · Hongyan Wu⁹ · Qunhong Wu^{10,11} · John E. Brazier¹²

Accepted: 24 December 2020
© The Author(s) 2021

Abstract

Objectives Our objective was to generate a value set for the SF-6Dv2 using time trade-off (TTO) and a discrete-choice experiment with a duration dimension (DCE_{TTO}) in China.

Methods A large representative sample of the Chinese general population was recruited from eight provinces/municipalities in China, stratified by age, sex, education level, and proportion of urban/rural residence. Respondents completed eight TTO tasks and ten DCE_{TTO} tasks during face-to-face interviews. Ordinary least squares (OLS), random-effects, fixed-effects, and Tobit models were used for TTO data, and conditional logit and mixed logit models were used for DCE_{TTO}. The monotonicity of model coefficients and the consistency of the predicted values according to intraclass correlation coefficient (ICC), mean absolute difference (MAD), and mean squared difference (MSD) were compared between the two approaches.

Results In total, 3320 respondents (50.3% male; range 18–90 years) were recruited. The random-effects model and the conditional logit model were preferred for the TTO and DCE_{TTO}, respectively. The TTO values ranged from – 0.277 to 1, with 927 (4.94%) states considered as worse than dead (WTD). The corresponding range for DCE_{TTO} was – 0.535 to 1, with a higher WTD of 8.50%. DCE_{TTO} presented minor nonmonotonicity with the coefficients in two dimensions. Values from the two approaches were highly consistent (ICC 0.9804, MAD 0.0588, MSD 0.0055), albeit those with DCE_{TTO} were slightly lower than those with TTO. The value set generated by TTO was preferred given the better monotonicity and the statistical significance of coefficients.

Conclusions The Chinese value set for the SF-6Dv2 was established based on the TTO approach, but the DCE_{TTO} also performed well. Minor issues of nonmonotonicity did present for DCE_{TTO}.

Key Points for Decision Makers

The Chinese value set for the SF-6Dv2 was established using a time trade-off (TTO) approach, which will facilitate the calculation of quality-adjusted life-years.

A direct comparison between the TTO and discrete-choice experiment with a duration dimension (DCE_{TTO}) approaches indicated a good performance for both; however, minor issues of non-monotonicity existed in DCE_{TTO} estimates.

A systematic difference was found between value sets developed using the TTO and DCE_{TTO} approaches.

✉ Jing Wu
jingwu@tju.edu.cn

Extended author information available on the last page of the article

1 Introduction

Economic evaluations of healthcare interventions are becoming integral to the reimbursement decision-making process in many countries, including China [1, 2]. Cost-utility analysis is a form of economic evaluation that quantifies health outcomes on a standardized metric, typically the quality-adjusted life-year (QALY), a single value produced by multiplying a quality adjustment weight (or health utility) by life duration [2–4]. The health utility, which lies on a 0–1 death–full health QALY scale, is calculated by a value set for a range of possible health states described by the health state classification system of generic preference-based measures. Examples of the most used measures include the EuroQol 5-Dimensions (EQ-5D) and the Short-Form Six-Dimension (SF-6D) [2, 5, 6], both of which are recommended for use in Chinese guidelines for pharmacoeconomic evaluations [7].

The SF-6D is derived from the Short-Form 36 (SF-36) health survey, which is one of the most widely-used health-related quality-of-life measures worldwide, including in China [2, 6, 8, 9]. The original health state classification system of the SF-6D comprises six dimensions with four to six levels in each, including physical functioning (PF), role limitation (RL), social functioning (SF), pain (PN), mental health (MH), and vitality (VT). Recently, a second version of the SF-6D, SF-6Dv2, was developed, which revisited the items selected from the SF-36 and modified the ambiguity between dimension levels and inconsistency of wording in the original version [8, 10]. The SF-6Dv2 has the same six dimensions as the SF-6Dv1, with five to six levels in each dimension, yielding up to 18,750 health states [2, 8, 10]. More details on the development of the SF-6Dv2 and comparisons with the SF-6Dv1 can be found elsewhere [8, 10]. The Simplified Chinese version of the SF-6Dv2 was developed after translation and cross-cultural adaptation, and preliminary psychometric testing was also conducted among the Chinese general population [11]. A country-specific value set for the SF-6Dv2 is currently available in the UK [12].

Health state utility values are commonly elicited using time trade-off (TTO) and standard gamble (SG) approaches [2, 13, 14]. Although TTO is generally regarded as simpler than SG, it is still considered too cognitively demanding for certain populations because of its iterative process, which may further result in response inconsistencies and subsequent data exclusions [2, 14–16]. A choice-based approach, the discrete-choice experiment (DCE), which some studies have argued may be simpler than the iterative process of TTO tasks, has recently gained popularity [16–19]. DCE tasks present two or more alternative health states, and respondents indicate their preference for one state over the other. However, a key problem in using DCEs has been how to anchor the values estimated by logit models, i.e., latent

utilities, onto the QALY scale [20–23]. The DCE with a duration dimension (DCE_{TTO}) approach, in which an additional dimension of life duration is presented with the health state, provides a valid alternative requiring no separate task or data manipulation for anchoring [19, 24–30].

Until now, no Chinese value set for the SF-6Dv2 has been available for the calculation of QALYs. A pilot study in 2018, based on a representative sample of the general population in Tianjin, China, was conducted to compare the acceptability, consistency, and accuracy of the TTO, DCE, and DCE_{TTO} approaches in utility elicitation by using the SF6Dv2 [31]. DCE and DCE_{TTO} were found to be feasible in the establishment of value sets, but they were not considered easier to understand or answer than TTO, which is consistent with a previous study [19]. In the pilot study, DCE_{TTO} had the highest completion rates and shortest completion time but showed a slight non-monotonicity on model coefficients [31], which has also been reported in other studies [12, 24, 28–30]. Therefore, this study aimed to generate a Chinese value set for SF-6Dv2 and to compare TTO and DCE_{TTO} approaches in a large representative sample of the general population in China.

2 Methods

Face-to-face interviews were conducted among a large representative sample of the general population of China to collect TTO and DCE_{TTO} responses, which were then modeled to estimate utility values for all health states in the SF-6Dv2.

2.1 Elicitation Tasks Design

Both TTO and DCE_{TTO} elicitation tasks were employed in this study. The composite TTO approach, which was developed by the EuroQol group [32, 33], was used in the TTO task (hereafter TTO) (Fig. 1a in the electronic supplementary material [ESM]), where “better than dead” and “worse than dead” (WTD) states were valued by conventional TTO and lead-time TTO, respectively. A detailed description of the composite TTO approach can be found elsewhere [31–33]. The health states “being in a wheelchair” and “being in a health state worse than dead” were used as warm-up questions to make sure respondents understood the concept of TTO before proceeding to the formal tasks. In the DCE_{TTO} task (Fig. 1b in the ESM), respondents were presented with a pair of health states described by the SF-6Dv2, with a further dimension representing the number of years living in that health state followed by death. Four levels of life-years were chosen: 1, 4, 7, and 10 years [12]. The longest duration was set to 10 years to be commensurate with the standard timeframe of the TTO task. Two stepwise warm-up questions were used in the DCE_{TTO} tasks. The

first warm-up question consisted of a pair of health states described by three dimensions, the first two of which were randomly chosen from SF-6Dv2 dimensions, and the third dimension represented the life duration. In the second warm-up question, two extra dimensions were further randomly chosen from the remaining SF-6Dv2 dimensions and added to describe the health states (i.e., five dimensions in total).

2.2 Health State Selection

The SF-6Dv2 defines a total of 18,750 health states, with more than 175 million potential pairwise combinations generated in the full factorial design. The number of possible combinations is even larger if the life duration dimension is added. A trade-off between the number of health states directly valued and the cognitive burden on respondents was considered following previous studies [31, 34, 35]. For TTO tasks, 295 health states described by the SF-6Dv2 were selected, including the six mildest imperfect health states (211111, 121111, 112111, 111211, 111121, 111112), the worst state (555655), and 288 other states generated based on near orthogonal arrays using SAS[®] Studio. The mildest health states were deliberately included because it allowed direct observations to distinguish the mildest impairments from full health. The 288 states were first distributed over 48 blocks, the state 555655 (included in all 48 blocks), and the six mildest states (each randomly included in eight blocks) were then added in the blocks. Each respondent was randomly assigned a block (i.e., eight TTO tasks) for valuation.

For DCE_{TTO} tasks, 300 pairs of health states (split into 30 blocks) were generated using the balanced overlap method. Both main effects and two-way interactions between the levels of each dimension and life-years were considered in the experimental design. Statistical efficiency was maximized with regard to the D-efficiency using Lighthouse Studio 9.6.0 (Sawtooth Software, Inc.) [36–38]. Each respondent was randomly assigned a block (i.e., ten DCE_{TTO} tasks) for valuation; the task order and the left–right position of health states within each task were all randomized.

2.3 Respondents

For each pair of DCE_{TTO} tasks, 100 observations are expected to result in robust model estimation [12]. Accordingly, the total target sample size was set at 3000. Respondents were recruited from eight cities, including Wuhan (central), Tianjin (north), Nanjing (east), Guangzhou (south), Lanzhou (northwest), Harbin (northeast), Chengdu, and Guiyang (southwest), as well as their surrounding rural areas, to achieve sufficient geographical spread and varied economic development levels in China (Fig. 2a in the ESM) [39, 40].

A stratified sampling method was used, in which four quotas were set for age, sex, education level, and proportion

of urban/rural residence, to ensure these distributions of the sample resembled those of the Chinese general population [39, 40]. In each of the eight cities chosen in this study, seven to ten districts (for urban areas) and villages (for rural areas) were selected, and 40–60 respondents were then recruited in each district/village. Recruitment was conducted in publicly accessible places (parks, shops, streets, or university campuses) and private places (participants' residence). Respondents were also required to meet the following inclusion criteria: (1) aged ≥ 18 years; (2) had Chinese nationality; (3) lived in mainland China during the past 5 years; (4) were literate and had no disease that limited cognitive function, such as dementia; and (5) gave informed consent.

2.4 Data Collection

Data were collected through two-to-one face-to-face computer-assisted personal interviews. The structure of the interview was as follows. First, respondents answered quota and inclusion criteria questions to confirm they were eligible for the interview. Second, respondents recorded their health state on the SF-6Dv2. Third, respondents completed the TTO and DCE_{TTO} tasks in the randomized order. Last, respondents provided a series of socio-demographic characteristics. Sound recordings of all interviews were collected with the respondents' permission.

In each of the eight selected cities, interviews were conducted by a local team from a local university. Each team was led by a local lead investigator and supervised by the principal investigators. A total of 146 interviewers with a bachelor's degree or higher were involved in this study. The interviewers attended a 2-day training to ensure equivalent task understanding, procedures, and interactions with respondents. Before the beginning of data collection, each interviewer was asked to complete three pilot interviews under the supervision of both the local lead investigator and the principal investigators of this study.

The study protocol was approved by the institutional review board of the School of Pharmaceutical Science and Technology, Tianjin University (no. 20180615). Informed consent was obtained from all respondents included in the study.

2.5 Quality Control

The quality of the collected data was monitored daily by the principal investigators. Interviews were directly excluded if (1) the interview was not completed; (2) respondents were not patient enough to follow the interviewers' guidance; or (3) interviewers failed to ask the questions or operate the questionnaire system according to the study protocol. Potentially problematic data were also identified, including respondents who gave the same values for all tasks;

gave the worst state (555655) a higher value (at least 0.5) than the other states in the TTO exercise [41–44]; always selected the same options, such as “AAAAA”; or selected “ABABAB” in the DCE_{TTO} [19, 43, 44]. Furthermore, we randomly selected 30% of the interview sound recordings for further daily double checking by the principal investigators to ensure the data quality.

2.6 Data Analysis

TTO data were analyzed according to main-effect specification using ordinary least squares (OLS) and Tobit models [2, 33]. The basic equation for the OLS model is as shown in Equation 1.

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

where y_i represents the disutility value; α represents the intercept; x_{dl} represents 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); β_{dl} represents the estimated disutility on dimension d at level l ; and ϵ represents the error term. Considering each respondent completed multiple TTO tasks, in addition to the OLS estimator with cluster-robust standard errors, the fixed- and random-effects models were also considered to account for the panel structure in the data.

The Tobit model has a potentially favorable characteristic because observed values were left-censored by the TTO methodology at -1 , whereas latent preferences of respondents might include valuations lower than -1 for health states WTD (Fig. 4 in the ESM). As shown in Eq. (2), the Tobit model assumes that a latent variable y_i^* underlies the observed y_i TTO disutility value and uses a likelihood function to adjust the parameter estimates for the probability of the y_i^* value beyond the censored value (i.e., lower than -1). Detailed information for the Tobit model is described elsewhere [41, 42, 45].

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > -1 \\ -1 & \text{if } y_i^* \leq -1 \end{cases} \quad (2)$$

The DCE_{TTO} data were analyzed under the random utility framework using both a conditional logit model (which assumes a homogenous preference from the respondents) and a mixed logit model (which allows for potential preference heterogeneity among respondents), following the model specification proposed by Bansback et al. [19] (Eq. 3).

$$U_{ij} = \beta t_{ij} + \lambda x_{ij} t_{ij} + \epsilon_{ij}, \quad (3)$$

where U_{ij} represents the binary choice of respondent i for DCE_{TTO} task j ; t_{ij} represents the life duration, which is modeled as a linear, continuous variable; β represents the

coefficient for the life duration; $x_{ij} t_{ij}$ represents the interactions between dimension levels and life duration; λ represents the coefficients for the interactions; and ϵ_{ij} represents the error term, which is assumed to be independently and identically distributed with Gumbel distribution. The mixed-logit model considers preference heterogeneity by estimating both the mean (which represents the average preferences of respondents) and the standard deviation (SD). In this study, a SF-6Dv2 dimension was considered random (with normal distribution) as long as the SD of at least one response level was statistically significant. The DCE_{TTO} value for each health state can be anchored on the QALY scale as shown in Eq. (4) [19, 26, 27, 29, 30].

$$U_{\text{anchored}} = 1 + \frac{\lambda}{\beta} x_{ij}. \quad (4)$$

2.7 Model Evaluation

The preferred models for both TTO and DCE_{TTO} approaches were selected based on (1) the monotonicity of logical ordering of the model coefficients, meaning that theoretically, the coefficients of more severe levels should have lower values than the coefficients of milder levels within each dimension; (2) the goodness of fit of the model using Akaike information criterion (AIC) and Bayesian information criterion (BIC); and (3) the parsimony of the model, meaning that the most parsimonious model would be selected if two or more models exhibited similar prediction performances. Furthermore, for TTO data, the prediction accuracy could be assessed by comparing predicted and observed mean values for health states valued in the study, using the intraclass correlation coefficient (ICC), the mean absolute difference (MAD), and the mean squared difference (MSD). Lower MAD and MSD and higher ICC values indicated better accuracy. Several interaction terms were also tested based on the preferred model for both TTO and DCE_{TTO}, which can be found in Tables 2 and 3 in the ESM. The final model, which would be used to calculate the health utility values and inform policy, requires the monotonicity of model coefficients [30, 46, 47]. The adjacent inconsistent levels in the preferred models were combined in this study to produce a fully consistent model.

2.8 Value Set Comparison

Based on the preferred model specification, after handling the potential issue of monotonicity, the comparison of the characteristics of health utility value sets generated by TTO and DCE_{TTO} was evaluated by the descriptive features, including the range of the utility value, the utility distribution of all 18,750 health states in SF-6Dv2, and the number

of health states WTD. The consistency between two value sets was also evaluated using ICC, MAD, and MSD values. The degree of agreement between utility values of TTO and DCE_{TTO} was assessed using a Bland–Altman plot. The cross-validation method was further used to demonstrate and compare the robustness of model estimation for both approaches. Specifically, data for one of the eight cities were excluded and the data for the remaining seven cities used for model estimation. This process was repeated eight times, in turn excluding data for each of the eight cities. Then, the MAD between coefficients of these fitted models and coefficients of the whole sample model was compared.

All statistical analyses were conducted using STATA 15.1. To compare the distribution of characteristics between subgroups, the t test was used for continuous variables and the χ^2 or Fisher's exact test was used for categorical variables. Differences in distribution of characteristics and model coefficients were considered statistically significant if $p < 0.05$.

3 Results

3.1 Respondents

A total of 3575 respondents were interviewed from June to September in 2019 (Fig. 3 in the ESM), of which 255 interviews were excluded because the respondents did not complete the whole interview ($N = 174$) or the interviews did not pass the quality control process ($N = 81$). Finally, a total of 3320 respondents were included, with geographic distribution as shown in Fig. 2b in the ESM. As illustrated in Table 1, the mean \pm SD age of respondents was 44.6 ± 16.1 years (range 18–90); 50.3% were males, and 40.4% lived in rural areas. The characteristics of respondents were close to those of the Chinese general population. The distributions for four quota characteristics were comparable across respondents in eight cities, and various distributions were observed for other characteristics, reflecting the geographical spread and different economic development levels in China (Table 1 in the ESM).

The mean \pm SD time spent in the interview was 39.4 ± 17.0 min, and the duration for TTO tasks was significantly longer than for DCE_{TTO} tasks (16.2 vs. 12.9 min; $p < 0.001$). Health problems were most frequently reported in the VT dimension (76.7%) and least frequently in PF (35.9%) (Fig. 1).

3.2 Data Characteristics

Mean observed TTO values ranged from -0.243 for state 555655 to 0.885 for state 111112 and ranged from 0.862 to 0.885 for the six mildest imperfect health states. Of 26,560

responses, 5011 (18.9%) were considered WTD. The distribution of observed TTO values for 295 states is presented in Fig. 4 in the ESM. For DCE_{TTO} data, as the difference in overall severity between the two states increased, respondents were more likely to choose the state with the lower severity; as expected, several inconsistencies were found because of the additional life duration dimension (Fig. 5 in the ESM).

Nevertheless, potentially problematic answer patterns were observed, including three respondents who gave the same values for all tasks, 51 who gave the worst state a higher value (at least 0.5) than the other states in TTO data, and respondents who always selected the same options (e.g., 20 responded “AAAAA” and 19 responded “ABABAB”) in the DCE_{TTO} . These respondents were few, with no noticeable differences in demographic characteristics, and some answers may be due to random errors. Therefore, these respondents were not excluded from this study.

3.3 Model Estimation

The estimated coefficients of the models on TTO data are presented in Table 2. The random-effects model performed better as measured by the criteria mentioned and was selected for the final data analysis for TTO data. Although the mixed logit model performed better in AIC and BIC, the conditional logit model was chosen for DCE_{TTO} data given that there were fewer non-monotonic coefficients and that the preference heterogeneity was not substantial (only four dimension levels had statistically significant SDs in the mixed logit model) (Table 3). In these two preferred models, all of the coefficients for TTO were ordered as expected. Level 2 in MH and VT dimensions for DCE_{TTO} showed slight non-monotonicity, while the coefficients were not statistically significant. The goodness of fit was slightly improved after combining the inconsistent levels. Most of the coefficients in both TTO and DCE_{TTO} models were significantly different from 0 ($p < 0.001$). All of the interaction terms were excluded in the final models because they resulted in non-monotonicity, varying degrees of impairment of the model estimations, or the parsimony of the model (Tables 2 and 3 in the ESM).

Following the previous study, the linear adjustment to the predicted values of TTO was made using the formula $U_{Adjusted} = U_{Predicted} / (1 - \text{intercept})$ (Table 4) [48]. This additional step was to remove the effect of the non-zero intercept in TTO, which leads to a predicted value of less than 1 for full health (111111).

3.4 Value Set Comparison

As illustrated in Table 4, values in the two approaches were highly consistent (ICC 0.9804, MAD 0.0588, MSD 0.0055).

Table 1 Respondent characteristics

| Characteristics | Chinese general population ^a (%) | Total sample (N = 3320) N (%) | Difference (%) |
|---|---|-------------------------------|----------------|
| Sex^b | | | |
| Male | 51.2 | 1670 (50.3) | - 0.9 |
| Female | 48.8 | 1650 (49.7) | +0.9 |
| Age (mean ± SD) | NA | 44.6 ± 16.1 | - |
| Age group (years)^b | | | |
| 18–29 | 21.5 | 708 (21.3) | - 0.2 |
| 30–39 | 18.7 | 613 (18.5) | - 0.2 |
| 40–49 | 21.1 | 670 (20.2) | - 0.9 |
| 50–59 | 17.1 | 614 (18.5) | + 1.4 |
| ≥ 60 | 21.6 | 715 (21.5) | - 0.1 |
| Education^b | | | |
| Primary or lower | 26.2 | 820 (24.7) | - 1.5 |
| Junior high school | 40.3 | 1288 (38.8) | - 1.5 |
| Senior high school | 17.2 | 601 (18.1) | + 0.9 |
| College or higher | 16.3 | 611 (18.4) | + 2.1 |
| Region^b | | | |
| Urban | 59.6 | 1980 (59.6) | - 0.05 |
| Rural | 40.4 | 1340 (40.4) | + 0.05 |
| Marital status | | | |
| Unmarried | NA | 709 (21.4) | - |
| Married | NA | 2434 (73.3) | - |
| Divorced | NA | 73 (2.2) | - |
| Widowed | NA | 104 (3.1) | - |
| Health insurance | | | |
| Urban employee | NA | 1576 (47.5) | - |
| Urban and rural resident | NA | 1476 (44.5) | - |
| Commercial | NA | 449 (13.5) | - |
| Other | NA | 74 (2.2) | - |
| No | NA | 188 (5.7) | - |
| Employment status | | | |
| Employed | NA | 2043 (61.5) | - |
| Retired | NA | 604 (18.2) | - |
| Student | NA | 229 (6.9) | - |
| Unemployed | NA | 444 (13.4) | - |
| Monthly income (RMB) | | | |
| < 2000 | NA | 858 (25.8) | - |
| 2000–5000 | NA | 1831 (55.2) | - |
| 5000–10,000 | NA | 481 (14.5) | - |
| > 10,000 | NA | 150 (4.5) | - |
| Number of chronic conditions^c | | | |
| 0 | NA | 2063 (62.1) | - |
| 1 | NA | 831 (25.0) | - |
| 2 | NA | 265 (8.0) | - |
| 3 | NA | 93 (2.8) | - |
| ≥ 4 | NA | 68 (2.0) | - |

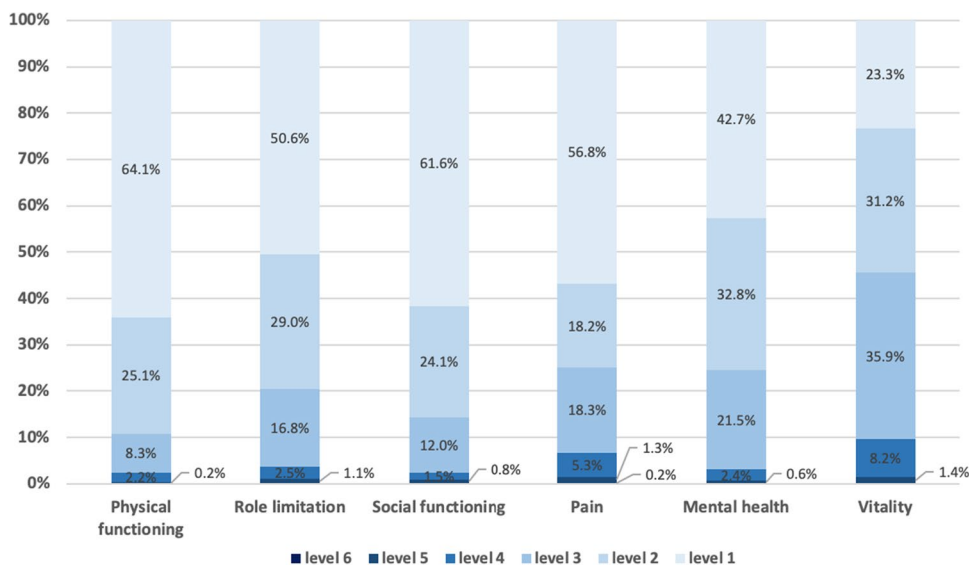
NA data not included in the publicly available data source, *RMB* renminbi, *SD* standard deviation

^aStatistics data for the Chinese general population were extracted from the Sixth National Census of China [39] and the China Statistical Yearbook [40]

^bQuota sampling was used in this study; sex, age, education status, and region were predefined on the basis of their distribution in the Chinese general population

^cChronic 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 psychiatric problems, memory-related disease, arthritis or rheumatism, asthma, or other respondent-reported chronic conditions

Fig. 1 Distribution across levels of the SF-6Dv2 dimensions. All dimensions have five levels, except for the pain dimension, which has six levels; higher values represent more severe states



The orders of overall decrement of the dimensions were the same for both approaches, as follows: PN, PF, MH, VT, SF, and RL. The Bland–Altman plot (Fig. 6 in the ESM) also showed that the mean difference of 0.02 was close to zero, the 95% limits of agreement between TTO and DCE_{TTO} ranged from − 0.11 to 0.16, and 95.7% of points lay within limits. Although agreement was generally good, we also observed that TTO values tended to be lower than DCE_{TTO} values for milder health states and higher than DCE_{TTO} values for worse health states. The comparison of the tendency of coefficients between TTO and DCE_{TTO} is presented in Fig. 2a, which shows that the coefficients of TTO decreased more smoothly than DCE_{TTO}. The estimated utility values for the 18,750 health states for SF-6Dv2 of both approaches, with the benchmark of observed TTO values, are shown in Fig. 2b. A similar trend can be found with the Bland–Altman plot (Fig. 6 in ESM). In total, 927 (4.94%) health states were estimated to be WTD in TTO, which were less than 1593 (8.50%) in DCE_{TTO}. The utility values of the worst state 555655 were − 0.277 for TTO and − 0.535 for DCE_{TTO}. The cross-validation results showed that excluding the data from one of the eight cities had only trivial effects on the coefficients for both TTO (less than 0.003) and DCE_{TTO} (less than 0.002) (Tables 4 and 5 in the ESM).

The value set generated by the random-effects model of TTO data after adjusting the intercept (Table 4) was preferred over that generated by the conditional logit model of DCE_{TTO} data, based on its performance concerning the monotonicity and statistical significance of the coefficients. In applying this preferred model as the value set of SF-6Dv2 in China, a health state utility value was obtained by subtracting coefficients for each dimension level of the health state from 1. For example, for the health state 232154, the

utility value would be $1 - (0.038 + 0.059 + 0.047 + 0 + 0.134 + 0.108) = 0.614$.

4 Discussion

This study collected TTO and DCE_{TTO} responses via face-to-face interviews with 3320 respondents who were representative of the general population of China in terms of age, sex, education, and proportion of urban/rural population. All of these response data were modeled to estimate utility values for all health states in the SF-6Dv2. This study presents the first empirical evidence of the systematic difference between these two approaches that directly compared value sets of the SF-6Dv2 generated by TTO and DCE_{TTO} approaches. Value sets for the EQ-5D-3L and the EQ-5D-5L have already been developed for China [48–50], and this study reports a Chinese-specific value set for the SF-6Dv2 that can be used for economic evaluations. Furthermore, as the first to generate a value set for the SF-6Dv2 in Asia, this study facilitates cross-country comparisons, which could provide further information on the health preference differences between eastern and western populations.

Both TTO and DCE_{TTO} approaches were feasible for eliciting health state utility values, and the orders of overall decrement of the dimensions were the same for both approaches. There were some (statistically insignificant) inconsistent coefficients in the DCE_{TTO} model and, following previous literature, the adjacent inconsistent levels were combined when developing value sets. It should be noted that this issue is not unique to this study and has been found in several previous valuation studies using DCE or DCE_{TTO} [12, 24, 28–31, 43]. Non-monotonicity of the coefficients can be caused by many factors, including respondents’

Table 2 Estimated coefficients (standard errors) of the fitted models on time trade-off data

| | M1: OLS model | | | M2: RE model | | | M3: FE model | | | M4: Tobit model | | | M5: RE Tobit model | | |
|-----------------------------|----------------------------------|-------|---------|--------------------------------|-------|---------|--------------|--------------|-------------------|-----------------|-------|---------|--------------------|--------------|-------------------|
| | Coef. | SE | p value | Coef. | SE | p value | Coef. | SE | p value | Coef. | SE | p value | Coef. | SE | p value |
| Intercept | 0.139 | 0.007 | < 0.001 | 0.130 | 0.005 | < 0.001 | 0.128 | 0.007 | < 0.001 | 0.113 | 0.010 | < 0.001 | 0.105 | 0.009 | < 0.001 |
| Physical functioning | | | | | | | | | | | | | | | |
| PF2 | 0.033 | 0.008 | < 0.001 | 0.033 | 0.007 | < 0.001 | 0.033 | 0.008 | < 0.001 | 0.038 | 0.010 | < 0.001 | 0.037 | 0.008 | < 0.001 |
| PF3 | 0.065 | 0.010 | < 0.001 | 0.069 | 0.008 | < 0.001 | 0.070 | 0.008 | < 0.001 | 0.069 | 0.010 | < 0.001 | 0.074 | 0.008 | < 0.001 |
| PF4 | 0.109 | 0.010 | < 0.001 | 0.122 | 0.008 | < 0.001 | 0.125 | 0.008 | < 0.001 | 0.114 | 0.010 | < 0.001 | 0.128 | 0.008 | < 0.001 |
| PF5 | 0.342 | 0.011 | < 0.001 | 0.344 | 0.010 | < 0.001 | 0.344 | 0.008 | < 0.001 | 0.347 | 0.010 | < 0.001 | 0.348 | 0.008 | < 0.001 |
| Role limitation | | | | | | | | | | | | | | | |
| RL2 | 0.041 | 0.009 | < 0.001 | 0.044 | 0.007 | < 0.001 | 0.044 | 0.008 | < 0.001 | 0.045 | 0.010 | < 0.001 | 0.047 | 0.008 | < 0.001 |
| RL3 | 0.053 | 0.011 | < 0.001 | 0.052 | 0.009 | < 0.001 | 0.051 | 0.009 | < 0.001 | 0.058 | 0.011 | < 0.001 | 0.056 | 0.009 | < 0.001 |
| RL4 | 0.087 | 0.009 | < 0.001 | 0.083 | 0.008 | < 0.001 | 0.083 | 0.008 | < 0.001 | 0.092 | 0.010 | < 0.001 | 0.087 | 0.008 | < 0.001 |
| RL5 | 0.089 | 0.010 | < 0.001 | 0.084 | 0.008 | < 0.001 | 0.083 | 0.008 | < 0.001 | 0.093 | 0.010 | < 0.001 | 0.088 | 0.008 | < 0.001 |
| Social functioning | | | | | | | | | | | | | | | |
| SF2 | 0.040 | 0.009 | < 0.001 | 0.041 | 0.008 | < 0.001 | 0.041 | 0.008 | < 0.001 | 0.044 | 0.010 | < 0.001 | 0.045 | 0.008 | < 0.001 |
| SF3 | 0.053 | 0.010 | < 0.001 | 0.052 | 0.008 | < 0.001 | 0.052 | 0.008 | < 0.001 | 0.057 | 0.010 | < 0.001 | 0.056 | 0.008 | < 0.001 |
| SF4 | 0.079 | 0.010 | < 0.001 | 0.081 | 0.008 | < 0.001 | 0.082 | 0.008 | < 0.001 | 0.084 | 0.010 | < 0.001 | 0.086 | 0.008 | < 0.001 |
| SF5 | 0.090 | 0.010 | < 0.001 | 0.094 | 0.008 | < 0.001 | 0.095 | 0.008 | < 0.001 | 0.094 | 0.010 | < 0.001 | 0.098 | 0.008 | < 0.001 |
| Pain | | | | | | | | | | | | | | | |
| PN2 | 0.041 | 0.009 | < 0.001 | 0.041 | 0.008 | < 0.001 | 0.041 | 0.008 | < 0.001 | 0.048 | 0.010 | < 0.001 | 0.047 | 0.008 | < 0.001 |
| PN3 | 0.067 | 0.011 | < 0.001 | 0.072 | 0.009 | < 0.001 | 0.073 | 0.009 | < 0.001 | 0.073 | 0.012 | < 0.001 | 0.078 | 0.009 | < 0.001 |
| PN4 | 0.127 | 0.011 | < 0.001 | 0.134 | 0.009 | < 0.001 | 0.136 | 0.009 | < 0.001 | 0.135 | 0.011 | < 0.001 | 0.141 | 0.009 | < 0.001 |
| PN5 | 0.330 | 0.012 | < 0.001 | 0.338 | 0.010 | < 0.001 | 0.339 | 0.009 | < 0.001 | 0.338 | 0.011 | < 0.001 | 0.346 | 0.009 | < 0.001 |
| PN6 | 0.369 | 0.012 | < 0.001 | 0.372 | 0.010 | < 0.001 | 0.372 | 0.009 | < 0.001 | 0.376 | 0.011 | < 0.001 | 0.378 | 0.009 | < 0.001 |
| Mental health | | | | | | | | | | | | | | | |
| MH2 | 0.020 | 0.009 | 0.035 | 0.028 | 0.008 | < 0.001 | 0.030 | 0.008 | < 0.001 | 0.023 | 0.010 | 0.022 | 0.032 | 0.008 | < 0.001 |
| MH3 | 0.052 | 0.011 | < 0.001 | 0.043 | 0.009 | < 0.001 | 0.041 | 0.008 | < 0.001 | 0.058 | 0.010 | < 0.001 | 0.048 | 0.008 | < 0.001 |
| MH4 | 0.119 | 0.011 | < 0.001 | 0.115 | 0.008 | < 0.001 | 0.115 | 0.008 | < 0.001 | 0.125 | 0.010 | < 0.001 | 0.124 | 0.008 | < 0.001 |
| MH5 | 0.120 | 0.010 | < 0.001 | 0.116 | 0.008 | < 0.001 | 0.114 | 0.008 | < 0.001 | 0.127 | 0.010 | < 0.001 | 0.123 | 0.008 | < 0.001 |
| Vitality | | | | | | | | | | | | | | | |
| VT2 | 0.017 | 0.009 | 0.049 | 0.025 | 0.007 | < 0.001 | 0.027 | 0.008 | 0.001 | 0.019 | 0.010 | 0.054 | 0.027 | 0.008 | 0.001 |
| VT3 | 0.053 | 0.011 | < 0.001 | 0.053 | 0.008 | < 0.001 | 0.053 | 0.008 | < 0.001 | 0.056 | 0.010 | < 0.001 | 0.056 | 0.008 | < 0.001 |
| VT4 | 0.090 | 0.011 | < 0.001 | 0.094 | 0.008 | < 0.001 | 0.095 | 0.009 | < 0.001 | 0.094 | 0.011 | < 0.001 | 0.097 | 0.009 | < 0.001 |
| VT5 | 0.093 | 0.010 | < 0.001 | 0.101 | 0.008 | < 0.001 | 0.103 | 0.008 | < 0.001 | 0.096 | 0.010 | < 0.001 | 0.104 | 0.008 | < 0.001 |
| Breusch Pagan LM test | < 0.001 (RE model was preferred) | | | | | | | | | - | | | - | | |
| Hausman test | | | | 0.409 (RE model was preferred) | | | | | | - | | | - | | |
| Log likelihood | - 18813.96 | | | - 14711.84 | | | - 9739.92 | | | - 19169.09 | | | - 15101.28 | | |
| AIC | 37679.92 | | | 29479.69 | | | 19531.84 | | | 38392.18 | | | 30258.57 | | |

Table 2 (continued)

| | M1: OLS model | | | M2: RE model | | | M3: FE model | | | M4: Tobit model | | | M5: RE Tobit model | | |
|-----|---------------|----|---------|--------------|----|---------|--------------|----|---------|-----------------|----|---------|--------------------|----|---------|
| | Coef. | SE | p value | Coef. | SE | p value | Coef. | SE | p value | Coef. | SE | p value | Coef. | SE | p value |
| BIC | 37892.79 | | | 29708.93 | | | 19744.70 | | | 38613.24 | | | 30487.81 | | |
| MAD | 0.0610 | | | 0.0616 | | | 0.0618 | | | 0.0622 | | | 0.0626 | | |
| MSD | 0.0062 | | | 0.0063 | | | 0.0064 | | | 0.0065 | | | 0.0066 | | |
| ICC | 0.9772 | | | 0.9307 | | | 0.9301 | | | 0.9306 | | | 0.9289 | | |

Bold formatting represents non-monotonicity

AIC Akaike information criterion, BIC Bayesian information criterion, Coef. coefficient, FE fixed effects, ICC intraclass correlation coefficient, M1–5 model 1–5, MAD mean absolute difference, MH mental health, MSD mean squared difference, OLS ordinary least squares, PF physical functioning, PN pain, RE random effects, RL role limitation, SE standard error, SF social functioning, VT vitality

characteristics, instruments used to describe health states, health states selected for valuation, and the model chosen to estimate the data. It has also been reported in studies conducted in different countries, among respondents with different characteristics and cultural backgrounds [12, 24, 28, 31, 43]; using different instruments [12, 24, 29, 30]; estimating data based on different models [12, 29, 30]; or even using different health states [12, 31]. Further research exploring the issue of inconsistent coefficients in DCE approaches is encouraged.

Although the value set generated by TTO data was favored over that generated by DCE_{TTO} data given the monotonicity and statistical significance of the coefficients, DCE_{TTO} did generate sensible results. The utility values generated by DCE_{TTO} were generally lower than those provided by TTO, which is consistent with previous studies [19, 26]. When compared with the UK value set for the SF-6Dv2, the range of values was similar, despite the different health states and experimental designs used for the DCE_{TTO} approach [12]. Specifically, the range of values was from 1 (111111) to – 0.535 (555655) for China and from 1 (111111) to – 0.574 (555655) for the UK, with the UK value set producing a slightly lower value [12]. The number of health states WTD was 1593 (8.50%) for China and 2850 (15.2%) for the UK [12]. The PN dimension had the largest decrement, and RL had the smallest, for both the China and the UK value sets; nevertheless, the order of the other dimensions was not identical. Further studies are warranted to compare the TTO and DCE_{TTO} value sets to provide more evidence when using DCE_{TTO} as a promising alternative to TTO, based on previous discussions [19, 31].

The value set generated by the TTO data was preferred in this study, even though statistically significant non-zero intercepts were observed. This was mainly because respondents gave low values for the very mild health state. This finding also existed in the Chinese EQ-5D-5L valuation study, which had an intercept of 0.121 [48]. Therefore, this issue may be related to the health preferences of the Chinese population, which tend to give very mild states a lower value. A significant intercept would favor and could result in overinvestment in treatments for very mild health problems. Therefore, following the Chinese EQ-5D-5L valuation study [48], a linear adjustment to all model coefficients was applied in this study in terms of using the SF-6Dv2 value set to better inform healthcare decision making.

This study also found that the decrement of the PN and the PF dimensions were the largest, indicating that the Chinese general population gave more weight to these two dimensions than other dimensions in SF-6Dv2. This is consistent with SF-6Dv1 value sets for Hong Kong China and Japan, which had the largest decrement for the same two dimensions [51, 52]. Similarly, in both Chinese EQ-5D-3L (2014) and EQ-5D-5L value sets, the decrement for pain/

Table 3 Estimated coefficients (standard errors) of the fitted models on DCE_{TTO} data

| | Conditional logit model (after combination) | | | Mixed logit model | | | | | | Conditional logit model (after combination) | | | Conditional logit model Anchored utility | |
|-----------------------------|---|--------------|--------------|-------------------|--------------|----------------|-------|-------|---------|---|-------|---------|--|----------------------|
| | Coef. | SE | p value | Coef. | SE | p value | SD | SE | p value | Coef. | SE | p value | Coef. | 95% CI |
| Year | 0.365 | 0.009 | < 0.001 | 0.465 | 0.013 | <0.001 | 0.247 | 0.006 | < 0.001 | 0.370 | 0.008 | < 0.001 | – | – |
| Physical functioning × year | | | | | | | | | | | | | | |
| PF2 | – 0.009 | 0.005 | 0.052 | – 0.008 | 0.006 | 0.168 | 0.068 | 0.012 | 0.060 | – 0.010 | 0.005 | 0.043 | – 0.027 | – 0.0011 to – 0.0520 |
| PF3 | – 0.018 | 0.005 | < 0.001 | – 0.022 | 0.006 | < 0.001 | 0.046 | 0.013 | 0.075 | – 0.019 | 0.004 | < 0.001 | – 0.050 | – 0.0267 to – 0.0736 |
| PF4 | – 0.054 | 0.005 | < 0.001 | – 0.069 | 0.005 | < 0.001 | 0.005 | 0.020 | 0.791 | – 0.054 | 0.005 | < 0.001 | – 0.145 | – 0.1206 to – 0.1688 |
| PF5 | – 0.150 | 0.005 | < 0.001 | – 0.217 | 0.007 | < 0.001 | 0.171 | 0.009 | < 0.001 | – 0.150 | 0.005 | < 0.001 | – 0.404 | – 0.3774 to – 0.4308 |
| Role limitation × year | | | | | | | | | | | | | | |
| RL2 | – 0.011 | 0.004 | 0.010 | 0.001 | 0.006 | 0.965 | 0.006 | 0.020 | 0.760 | – 0.011 | 0.004 | 0.010 | – 0.029 | – 0.0073 to – 0.0516 |
| RL3 | – 0.011 | 0.005 | 0.017 | – 0.023 | 0.006 | < 0.001 | 0.066 | 0.011 | 0.078 | – 0.011 | 0.005 | 0.018 | – 0.030 | – 0.0056 to – 0.0552 |
| RL4 | – 0.030 | 0.004 | < 0.001 | – 0.024 | 0.006 | < 0.001 | 0.043 | 0.016 | 0.006 | – 0.030 | 0.004 | < 0.001 | – 0.081 | – 0.0589 to – 0.1034 |
| RL5 | – 0.041 | 0.005 | < 0.001 | – 0.009 | 0.006 | 0.137 | 0.005 | 0.058 | 0.926 | – 0.041 | 0.005 | < 0.001 | – 0.112 | – 0.0880 to – 0.1354 |
| Social functioning × year | | | | | | | | | | | | | | |
| SF2 | – 0.013 | 0.005 | 0.004 | – 0.010 | 0.006 | 0.071 | – | – | – | – 0.013 | 0.005 | 0.004 | – 0.035 | – 0.0114 to – 0.0586 |
| SF3 | – 0.014 | 0.004 | 0.001 | – 0.027 | 0.006 | < 0.001 | – | – | – | – 0.014 | 0.004 | 0.001 | – 0.038 | – 0.0151 to – 0.0613 |
| SF4 | – 0.039 | 0.004 | < 0.001 | – 0.051 | 0.006 | < 0.001 | – | – | – | – 0.039 | 0.004 | < 0.001 | – 0.104 | – 0.0819 to – 0.1266 |
| SF5 | – 0.042 | 0.004 | < 0.001 | – 0.048 | 0.006 | < 0.001 | – | – | – | – 0.042 | 0.004 | < 0.001 | – 0.114 | – 0.0923 to – 0.1348 |
| Pain × year | | | | | | | | | | | | | | |
| PN2 | – 0.027 | 0.005 | < 0.001 | – 0.032 | 0.006 | < 0.001 | 0.037 | 0.014 | 0.068 | – 0.027 | 0.005 | < 0.001 | – 0.072 | – 0.0446 to – 0.0996 |
| PN3 | – 0.029 | 0.005 | < 0.001 | – 0.013 | 0.007 | 0.054 | 0.058 | 0.014 | 0.213 | – 0.029 | 0.005 | < 0.001 | – 0.079 | – 0.0536 to – 0.1054 |
| PN4 | – 0.057 | 0.005 | < 0.001 | – 0.061 | 0.007 | < 0.001 | 0.032 | 0.014 | 0.020 | – 0.057 | 0.005 | < 0.001 | – 0.155 | – 0.1295 to – 0.1802 |
| PN5 | – 0.173 | 0.006 | < 0.001 | – 0.216 | 0.007 | < 0.001 | 0.078 | 0.012 | 0.093 | – 0.173 | 0.006 | < 0.001 | – 0.466 | – 0.4382 to – 0.4948 |
| PN6 | – 0.200 | 0.005 | < 0.001 | – 0.263 | 0.007 | < 0.001 | 0.127 | 0.011 | < 0.001 | – 0.200 | 0.005 | < 0.001 | – 0.541 | – 0.5126 to – 0.5688 |
| Mental health × year | | | | | | | | | | | | | | |
| MH2 | 0.002 | 0.004 | 0.686 | 0.004 | 0.006 | 0.543 | – | – | – | 0.000 | – | – | 0.000 | – |
| MH3 | – 0.003 | 0.004 | 0.568 | – 0.026 | 0.006 | < 0.001 | – | – | – | – 0.004 | 0.004 | 0.341 | – 0.010 | 0.0104 to – 0.0303 |
| MH4 | – 0.053 | 0.005 | < 0.001 | – 0.075 | 0.006 | < 0.001 | – | – | – | – 0.054 | 0.004 | < 0.001 | – 0.146 | – 0.1266 to – 0.1656 |
| MH5 | – 0.072 | 0.005 | < 0.001 | – 0.099 | 0.006 | < 0.001 | – | – | – | – 0.073 | 0.004 | < 0.001 | – 0.197 | – 0.1750 to – 0.2193 |
| Vitality × year | | | | | | | | | | | | | | |
| VT2 | 0.007 | 0.005 | 0.145 | 0.019 | 0.006 | 0.001 | – | – | – | 0.000 | – | – | 0.000 | – |
| VT3 | – 0.027 | 0.005 | < 0.001 | – 0.011 | 0.006 | 0.058 | – | – | – | – 0.031 | 0.004 | < 0.001 | – 0.083 | – 0.0611 to – 0.1040 |
| VT4 | – 0.029 | 0.005 | < 0.001 | – 0.044 | 0.006 | < 0.001 | – | – | – | – 0.033 | 0.004 | < 0.001 | – 0.089 | – 0.0702 to – 0.1085 |
| VT5 | – 0.058 | 0.005 | < 0.001 | – 0.070 | 0.006 | < 0.001 | – | – | – | – 0.062 | 0.004 | < 0.001 | – 0.167 | – 0.1475 to – 0.1871 |

Table 3 (continued)

| | Conditional logit model (after combination) | | | Mixed logit model | | | Conditional logit model (after Anchored utility) | | | |
|----------------|---|----|---------|-------------------|----|---------|--|----|---------|--------|
| | Coef. | SE | p value | Coef. | SE | p value | Coef. | SE | p value | 95% CI |
| Log likelihood | - 35,563.72 | | | - 29,583.62 | | | - 35,430.03 | | | |
| AIC | 71,179.45 | | | 59,228.25 | | | 70,914.06 | | | |
| BIC | 71,395.13 | | | 59,427.93 | | | 71,119.74 | | | |

For the mixed logit model, a dimension was specified as having random coefficients (with a normal distribution) as long as the SD of one response level was statistically significant ($p < 0.05$). 500 Halton draws were used for the simulation. Among all six dimensions, the SDs of SF, MH, and VT were consistently insignificant. As such, these three dimensions were set as fixed coefficients. Although the mixed logit model estimates had a better model fit according to AIC and BIC information criteria, the conditional logit model was selected as the final model given that (1) among all estimations, less non-monotonic coefficients were found; (2) only four dimension levels had statistically significant SDs (as such, the preference heterogeneity was not substantial). Coefficients presented in bold represent the non-monotonicity

AIC Akaike information criterion, BIC Bayesian information criterion, CI confidence interval, Coef. coefficient, DCE_{TTO} discrete-choice experiments with a duration dimension, MH mental health, PF physical functioning, PN pain, RE random effects, RL role limitation, SD standard deviation, SE standard error, SF social functioning, VT vitality

Table 4 Comparison between health utility value sets generated by time trade-off and discrete-choice experiments with a duration dimension

| | TTO (adjusted) Coef. | DCE _{TTO} Coef. |
|--------------------------------|-------------------------|-----------------------------|
| Physical functioning | | |
| PF1 | 0.000 | 0.000 |
| PF2 | - 0.038 | - 0.027 |
| PF3 | - 0.080 | - 0.050 |
| PF4 | - 0.140 | - 0.145 |
| PF5 | - 0.395 | - 0.404 |
| Role limitation | | |
| RL1 | 0.000 | 0.000 |
| RL2 | - 0.050 | - 0.029 |
| RL3 | - 0.059 | - 0.030 |
| RL4 | - 0.096 | - 0.081 |
| RL5 | - 0.097 | - 0.112 |
| Social functioning | | |
| SF1 | 0.000 | 0.000 |
| SF2 | - 0.047 | - 0.035 |
| SF3 | - 0.060 | - 0.038 |
| SF4 | - 0.093 | - 0.104 |
| SF5 | - 0.108 | - 0.114 |
| Pain | | |
| PN1 | 0.000 | 0.000 |
| PN2 | - 0.047 | - 0.072 |
| PN3 | - 0.083 | - 0.079 |
| PN4 | - 0.154 | - 0.155 |
| PN5 | - 0.388 | - 0.466 |
| PN6 | - 0.427 | - 0.541 |
| Mental health | | |
| MH1 | 0.000 | 0.000 |
| MH2 | - 0.033 | 0.000 |
| MH3 | - 0.050 | - 0.010 |
| MH4 | - 0.132 | - 0.146 |
| MH5 | - 0.134 | - 0.197 |
| Vitality | | |
| VT1 | 0.000 | 0.000 |
| VT2 | - 0.029 | 0.000 |
| VT3 | - 0.060 | - 0.083 |
| VT4 | - 0.108 | - 0.089 |
| VT5 | - 0.116 | - 0.167 |
| No. (%) of worse than death | 927 (4.94%) | 1593 (8.50%) |
| The worst state value (55E655) | - 0.277 | - 0.535 |
| MAD | 0.0588 | |
| MSD | 0.0055 | |
| ICC | 0.9804 | |

The value set generated by TTO was based on model 2 (random-effects model) shown in Table 2, and the linear adjustment to remove the effect of the non-zero intercept was made using the formula $U_{Adjusted} = U_{Predicted}/(1 - intercept)$. The value set generated by DCE_{TTO} was based on the anchored coefficients of the conditional logit model (Table 3) with the combination of inconsistent coefficients

Coef. coefficient, DCE_{TTO} discrete-choice experiments with a duration dimension, ICC intraclass correlation coefficient, MAD mean absolute difference, MH mental health, MSD mean squared difference, PF physical functioning, PN pain, RL role limitation, SF social functioning, TTO time trade-off, VT vitality

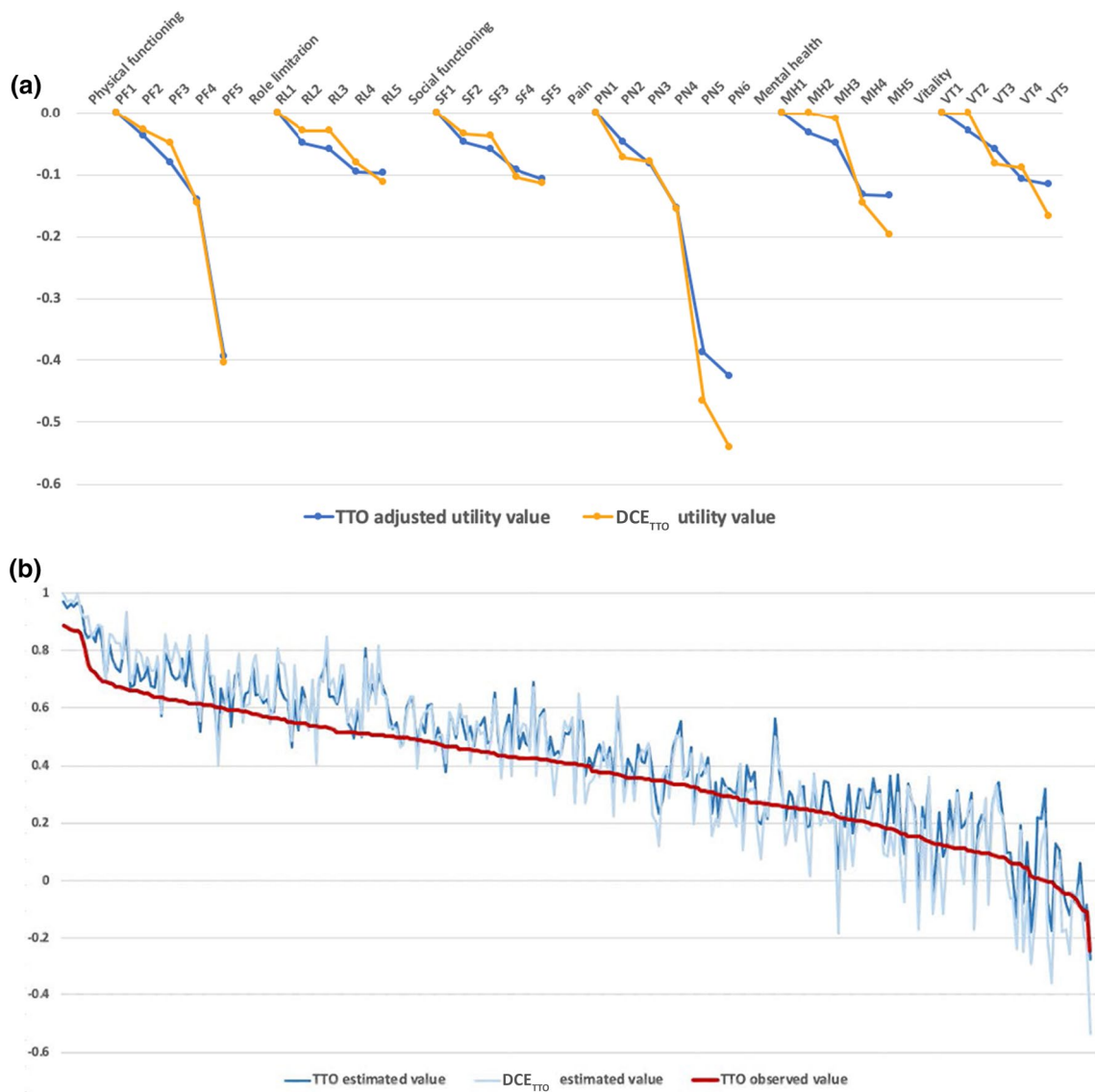


Fig. 2 Comparison of the value sets generated by TTO and DCE_{TTO} . **a** Comparison of the tendency of coefficients between TTO and DCE_{TTO} . **b** Estimated utility values for the 18,750 health states for the SF-6Dv2 based on the TTO and DCE_{TTO} data, with the benchmark

of observed TTO values (ordered by the TTO observed values). TTO time trade-off, DCE_{TTO} discrete-choice experiment with a duration dimension

discomfort and mobility dimensions was the largest [48, 50] similar to the EQ-5D-5L value sets for South Korea, Malaysia, and Thailand [53–55]. However, in the Chinese EQ-5D-3L (2018) value set [49], the decrement for the self-care dimension was the largest, and pain/discomfort was the smallest, which differed from these studies. This inconsistency may be partly because of the different TTO task design used to generate the Chinese EQ-5D-3L (2018) value set [49]. In western countries, such as the USA, the UK, Germany, and the Netherlands, the decrements in the PN and MH dimensions in the SF-6Dv1 [56, 57] and in the pain/discomfort and anxiety/depression dimensions in the EQ-5D-5L [41, 45, 58], were the largest. Although different

measures were used in these studies, the similarities in the health state classification system provided good comparability. Populations of both eastern and western countries may give similarly large preferences for PN. In contrast, populations of eastern countries may give more weight to PF, while those of western countries may have more preference for MH. The similarities and distinctions in the ranking of the dimensions reflect cultural and socioeconomic factors, which are essential to shaping the preferences of populations. Further investigation is needed to explore and compare the impact on the results of economic evaluations by using the newly established SF-6Dv2 value set in this study and the existing Chinese EQ-5D value sets as mentioned.

A particular strength of this study is its sample size, which was larger than most of the other valuation studies [59–62] and helped to reduce the standard errors of model coefficients (no larger than 0.01 in this study). Besides, given the large proportion of rural residents in China, an important factor that may affect health preferences [49, 63, 64], a specific quota of the urban and rural proportion of the Chinese general population was employed for the first time in this study. This improved the representativeness of the study sample and provided a more reliable health utility value set to reflect the health preferences of the Chinese population.

Several limitations of this study need to be noted. First, 146 interviewers involved in this study had the same extensive training but came from different backgrounds and used different communication skills when conducting the interviews. Although the cross-validation results showed that excluding the data from one of the eight cities had only trivial effects on the model estimation, there may be some unobservable effects [65]. Second, to achieve the maximum statistical efficiency of modeling, implausible health states in SF-6Dv2 were not excluded in the experimental design for both TTO and DCE_{TTO}. Asking respondents to consider implausible health states was likely to have had an impact on the quality of their responses and may have affected the model estimation results. There was also a lack of agreement among respondents on which states were implausible [66]. Furthermore, the order of the eight tasks in each TTO block was not completely random. The mildest state and the worst state were always the seventh and eighth states, respectively, because of the technical limitations during production of the survey. This could have had some minor impacts on the estimates.

5 Conclusions

The Chinese value set for the SF-6Dv2 was established based on the TTO approach, and both TTO and DCE_{TTO} approaches performed well when eliciting health state utility values in China. Minor issues of nonmonotonicity did present for DCE_{TTO}.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40273-020-00997-1>.

Acknowledgements The authors thank Dr. Zhihao Yang, Dr. Keren Zhang, Dr. Hong Zhu, Dr. Yunfang Jiang, Jie Tian, Mengqian Zhang, Jiahui Zhang, Pinan Chen, Nan Fang, Lili Chen, Meng Lv, Jia Wang, Li Zhou, Zhuoru Liang and Yunyu Li, for the excellent quality control work during this study. We also thank all of the interviewers and respondents for taking part in this study.

Declarations

Funding This study was funded by the National Natural Science Foundation of China (Grant no. 71673197 and no. 71804122) and the Science and Technology Program of Guangzhou, China (Grant no. 201704020198).

Conflict of interest JW and XH reported receiving grants from the National Natural Science Foundation of China during the conduct of the study. JJ reported receiving grants from the Science and Technology Program of Guangzhou, China during the conduct of the study. JB is one of the developers and holds the patent for SF-6D with royalties paid to the University of Sheffield. SX, GC, GB, DF, MH, XW, HW, and QW have no conflicts of interest that are directly relevant to the content of this article.

Ethics approval This study was approved by the Institutional Review Board of School of Pharmaceutical Science and Technology, Tianjin University (No. 20180615) and was conducted in accordance with the Declaration of Helsinki.

Consent to participate Informed consent was obtained from all individual participants included in the study. Participants were informed about their freedom of refusal. Anonymity and confidentiality were maintained throughout the research process.

Consent for publication Informed consent for publication was obtained from all individual participants. Participants were informed about their freedom of refusal for data publication. Anonymity and confidentiality were maintained in this publication.

Availability of data and material The predicted values of the 18,750 health states, together with standard errors and 95% confidence intervals, are available from the corresponding author on reasonable request.

Code availability Not applicable.

Author Contributions Concept and design: JW, SX, XH, GC, and JB. Acquisition of data: SX, GB, DF, MH, JJ, XW, HW, and QW. Analysis and interpretation of data: JW, SX, XH, GC, and JB. Drafting of the manuscript: JW, SX, and GC. Statistical analysis: SX and GC. Obtaining funding: JW, XH, and JJ. Supervision: JW, JB. All authors commented on previous versions of the manuscript and approved the final manuscript.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

- Zhao R, Zhao K. Health technology assessment takes the road of development with Chinese characteristics. *China Health*. 2019;10:76–8.
- Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press; 2017.
- Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. Cost-effectiveness in health and medicine. 2nd ed. New York: Oxford University Press; 2016.
- Rascati K. Essentials of pharmacoeconomics. Philadelphia: Commonwealth of Pennsylvania, Lippincott Williams & Wilkins; 2013.
- The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208. [https://doi.org/10.1016/0168-8510\(90\)90421-9](https://doi.org/10.1016/0168-8510(90)90421-9).
- Brazier J, Usherwood T, Harper R, et al. Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol*. 1998;51:1115–28.
- Liu GG, Hu S, Wu JH, Wu J, Dong C, Li H. China guidelines for pharmacoeconomic evaluations (2020). Beijing, China: China Market Press; 2020.
- Poder TG, Fauteux V, He J, et al. Consistency between three different ways of administering the short form 6 dimension version 2. *Value Health*. 2019;22:837–42.
- Liu X, Li S, Chen G. Development of the short form health survey and introduction of short form 6-dimension (SF-6D). *Chin Health Econ*. 2019;38(02):8–11.
- Brazier J, 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;58:557–65.
- Wu J, Xie S, He X, et al. The simplified Chinese version of SF-6Dv2: translation, cross-cultural adaptation and preliminary psychometric testing. *Qual Life Res*. 2020;29:1385–91.
- Mulhern BJ, Bansback N, Norman R, et al. Valuing the SF-6Dv2 classification system in the United Kingdom using a discrete-choice experiment with duration. *Med Care*. 2020;58:566–73.
- Martin AJ, Glasziou PP, Simes RJ, et al. A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *Int J Technol Assess Health Care*. 2000;16:137–47.
- Morimoto T, Fukui T. Utilities measured by rating scale, time trade-off, and standard gamble: review and reference for health care professionals. *J Epidemiol*. 2002;12:160–78.
- Craig BM, Busschbach JJ, Salomon JA. Keep it simple: ranking health states yields values similar to cardinal measurement approaches. *J Clin Epidemiol*. 2009;62(3):296–305.
- Craig BM, Busschbach JJ. The episodic random utility model unifies time trade-off and discrete choice approaches in health state valuation. *Popul Health Metrics*. 2009;7:3.
- Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics*. 2008;26:661–77.
- Stolk EA, Oppe M, Scalone L, et al. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. *Value Health*. 2010;13:1005–13.
- Bansback N, Brazier J, Tsuchiya A, et al. Using a discrete choice experiment to estimate health state utility values. *J Health Econ*. 2012;31:306–18.
- Salomon JA. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. *Popul Health Metrics*. 2003;1:12.
- McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. *J Health Econ*. 2006;25:418–31.
- Ratcliffe J, Brazier J, Tsuchiya A, et al. Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. *Health Econ*. 2009;18:1261–76.
- 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. *Eur J Health Econ*. 2012;13:575–87.
- Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. *Appl Health Econ Health Policy*. 2013;11:287–98.
- Mulhern B, Bansback N, Brazier J, et al. Preparatory study for the revaluation of the EQ-5D tariff: methodology report. *Health Technol Assess (Winchester, England)*. 2014;18:1–191 (vii-xxvi).
- Norman R, Viney R, Brazier J, et al. Valuing SF-6D health states using a discrete choice experiment. *Med Decis Making*. 2014;34:773–86.
- Viney R, Norman R, Brazier J, et al. An Australian discrete choice experiment to value EQ-5D health states. *Health Econ*. 2014;23:729–42.
- Mulhern B, Bansback N, Hole AR, et al. Using discrete choice experiments with duration to model EQ-5D-5L health state preferences: testing experimental design strategies. *Med Decis Mak*. 2017;37(3):285–97.
- King MT, Viney R, Simon Pickard A, et al. Australian Utility Weights for the EORTC QLU-C10D, a multi-attribute utility instrument derived from the cancer-specific quality of life questionnaire, EORTC QLQ-C30. *Pharmacoeconomics*. 2018;36:225–38.
- Rowen D, Mulhern B, Stevens K, et al. Estimating a Dutch value set for the pediatric preference-based CHU9D using a discrete choice experiment with duration. *Value Health*. 2018;21:1234–42.
- 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 SF6Dv2 in China. *Value Health*. 2020;23:1391–9.
- Janssen BM, Oppe M, Versteegh MM, et al. Introducing the composite time trade-off: a test of feasibility and face validity. *Eur J Health Econ*. 2013;14(Suppl 1):S5-13.
- Oppe M, Rand-Hendriksen K, Shah K, et al. EuroQol protocols for time trade-off valuation of health outcomes. *Pharmacoeconomics*. 2016;34:993–1004.
- Oppe M, Devlin NJ, van Hout B, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health*. 2014;17:445–53.
- Burgess L, Street DJ, Wasi N. Comparing designs for choice experiments: a case study. *J Stat Theory Pract*. 2011;5:25–46.
- Chrzan K, Orme B. An overview and comparison of design strategies for choice-based conjoint analysis. Sawtooth software research paper series. 2000; 98382.
- Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health*. 2013;16:3–13.
- 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.
- National Bureau of Statistics of China. China Sixth National Census. 2010. <http://stats.tj.gov.cn/nianjian/2017nj/zk/indexch.htm>. Accessed 25 Sept 2020.
- National Bureau of Statistics of China. China Statistical Yearbook. 2018. <http://www.stats.gov.cn/tjsj/ndsj/2018/indexch.htm>. Accessed 25 Sept 2020.
- Ludwig K, von der Schulenburg JMG, Greiner W. German value set for the EQ-5D-5L. *Pharmacoeconomics*. 2018;36:663–74.

42. Pickard AS, Law EH, Jiang R, et al. United States valuation of EQ-5D-5L health states using an international protocol. *Value Health*. 2019;22:931–41.
43. Purba FD, Hunfeld JAM, Iskandarsyah A, et al. The Indonesian EQ-5D-5L value set. *Pharmacoeconomics*. 2017;35:1153–65.
44. Ramos-Goni JM, Oppe M, Slaap B, et al. Quality control process for EQ-5D-5L valuation studies. *Value Health*. 2017;20:466–73.
45. Versteegh MM, Vermeulen KM, Evers SMAA, et al. Dutch Tariff for the Five-Level Version of EQ-5D. *Value Health*. 2016;19:343–52.
46. Brazier J, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42:851–9.
47. Mukuria C, Rowen D, Brazier J, et al. Deriving a preference-based measure for myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health*. 2015;18:846–55.
48. Luo N, Liu G, Li M, et al. Estimating an EQ-5D-5L value set for China. *Value Health*. 2017;20:662–9.
49. Zhuo L, Xu L, Ye J, et al. Time trade-off value set for EQ-5D-3L based on a nationally representative Chinese Population Survey. *Value Health*. 2018;21(11):1330–7.
50. Liu GG, Wu H, Li M, et al. Chinese time trade-off values for EQ-5D health states. *Value Health*. 2014;17:597–604.
51. Lam CL, Brazier J, McGhee SM. Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population. *Value Health*. 2008;11:295–303.
52. Brazier J, Fukuhara S, Roberts J, et al. Estimating a preference-based index from the Japanese SF-36. *J Clin Epidemiol*. 2009;62:1323–31.
53. Kim SH, Ahn J, Ock M, et al. The EQ-5D-5L valuation study in Korea. *Qual Life Res*. 2016;25:1845–52.
54. Shafie AA, Vasan Thakumar A, Lim CJ, et al. EQ-5D-5L valuation for the Malaysian population. *Pharmacoeconomics*. 2019;37:715–25.
55. Pattanaphesaj J, Thavorncharoensap M, Ramos-Goni JM, et al. The EQ-5D-5L Valuation study in Thailand. *Expert Rev Pharmacoecon Outcomes Res*. 2018;18:551–8.
56. Craig BM, Pickard AS, Stolk E, et al. US valuation of the SF-6D. *Med Decis Mak*. 2013;33:793–803.
57. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21:271–92.
58. Devlin NJ, Shah KK, Feng Y, et al. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Econ*. 2018;27:7–22.
59. Wang P, Liu GG, Jo MW, et al. Valuation of EQ-5D-5L health states: a comparison of seven Asian populations. *Expert Rev Pharmacoecon Outcomes Res*. 2019;19:445–51.
60. 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.
61. Zhou T, Guan H, Yao J, et al. The quality of life in Chinese population with chronic non-communicable diseases according to EQ-5D-3L: a systematic review. *Qual Life Res*. 2018;27(11):2799–814.
62. Xie F, Gaebel K, Perampaladas K, et al. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. *Med Decis Mak*. 2014;34:8–20.
63. Zhou Z, Zhou Z, Gao J, et al. Urban-rural difference in the associations between living arrangements and the health-related quality of life (HRQOL) of the elderly in China-evidence from Shaanxi province. *PLoS One*. 2018;13(9):e0204118.
64. Chen Y, Sun G, Guo X, et al. Factors affecting the quality of life among Chinese rural general residents: a cross-sectional study. *Public Health*. 2017;146:140–7.
65. Yang Z, van Busschbach J, Timman R, et al. Logical inconsistencies in time trade-off valuation of EQ-5D-5L health states: whose fault is it? *PLoS One*. 2017;12(9):e0184883.
66. Yang Z, Feng Z, Busschbach J, et al. How prevalent are implausible EQ-5D-5L health states and how do they affect valuation? A study combining quantitative and qualitative evidence. *Value Health*. 2019;22(7):829–36.

Authors and Affiliations

Jing Wu^{1,2}  · Shitong Xie^{1,2} · Xiaoning He^{1,2} · Gang Chen³ · Gengliang Bai⁴ · Da Feng⁵ · Ming Hu⁶ · Jie Jiang⁷ · Xiaohui Wang⁸ · Hongyan Wu⁹ · Qunhong Wu^{10,11} · John E. Brazier¹²

¹ School of Pharmaceutical Science and Technology, Tianjin University, Room 209, 24th building, 92th Weijin Road, Nankai District, Tianjin 300072, China

² Center for Social Science Survey and Data, Tianjin University, Tianjin, China

³ Monash Business School, Centre for Health Economics, Monash University, Melbourne, VIC, Australia

⁴ School of Health Economics and Management, Nanjing University of Chinese Medicine, Nanjing, China

⁵ School of Pharmacy, Huazhong University of Science and Technology, Wuhan, China

⁶ West China School of Pharmacy, Sichuan University, Chengdu, China

⁷ College of Pharmacy, Jinan University, Guangzhou, China

⁸ School of Public Health, Lanzhou University, Lanzhou, China

⁹ School of Medicine and Health Management, Guizhou Medical University, Guiyang, China

¹⁰ Department of Health Policy, Health Management College, Harbin Medical University, Harbin, China

¹¹ Department of Social Medicine, School of Public Health, Harbin Medical University, Harbin, China

¹² School of Health and Related Research, University of Sheffield, Sheffield, UK