Quality of Life Research (2019) 28:163–176
https://doi.org/10.1007/s11136-018-2032-z

Scoring the Child Health Utility 9D instrument: estimation of a Chinese child and adolescent-specific tariff

Gang Chen[1] · Fei Xu[2,3] · Elisabeth Huynh[4] · Zhiyong Wang[2] · Katherine Stevens[5] · Julie Ratcliffe[4]

Accepted: 21 October 2018 / Published online: 29 October 2018

Abstract

Purpose To derive children and adolescents' preferences for health states defined by the Chinese version of Child Health Utility 9D (CHU9D-CHN) instrument in China that can be used to estimate quality-adjusted life years (QALYs) for economic evaluation.

Methods A profile case best–worst scaling (BWS) and a time trade-off (TTO) method were combined to derive a Chinese-specific tariff for the CHU9D-CHN. The BWS survey recruited students from primary and high schools using a multi-stage random sampling method and was administered in a classroom setting, whilst the TTO survey adopted an interviewer-administrated conventional TTO task and was administered to a convenience sample of undergraduate students. A latent class modelling framework was adopted for analysing the BWS data.

Results Two independent surveys were conducted in Nanjing, China, including a valid sample of 902 students (mean age 13 years) from the BWS survey and a valid sample of 38 students (mean age 18 years) from the TTO survey. The poolability of the best and the worst responses was rejected and the optimal result based on the best responses only. The optimal model suggests the existence of two latent classes. The BWS estimates were further re-anchored onto the QALY scale using the TTO generated health state values via a mapping approach.

Conclusion This study provides further insights into the use of the BWS method to generate health state values with young people and highlights the potential different decision rules that young people may employ for determining best vs. worst choices in this context.

Keywords Child Health Utility 9D · Quality-adjusted life years · Economic evaluation · Child · Adolescent · China

Introduction

The Healthy China 2030 plan was approved by China's Central Government in 2016 to ensure the health of the Chinese population and future generations [1]. This initiative recognises that health is essential to economic and social development and places improving health throughout the life course as a key priority in the country's development strategy. Childhood and adolescence represents a critical development stage in the life course, and it is during this period that individuals become increasingly responsible for their own health and health care [2]. Health-related quality of life (HRQoL) is a multi-dimensional construct that measures the impact of health or disease on physical and psychosocial functioning, and it has been widely used as an outcome measure in health policy and planning [3, 4]. Several previous studies have reported upon the application of the Pediatric Quality of Life Inventory (PedsQL)™ 4.0 Generic Core Scales, a validated generic non-preference-based paediatric instrument, in Mainland China to study children and adolescents' HRQoL, among both community respondents [5] and patients [6, 7].

✉ Fei Xu
frankxufei@163.com

[1] Centre for Health Economics, Monash Business School, Monash University, Clayton, Australia

[2] Nanjing Municipal Center for Disease Control and Prevention, 2 Zizhulin, Nanjing 210003, China

[3] School of Public Health, Nanjing Medical University, Nanjing, China

[4] Institute for Choice, University of South Australia, Adelaide, Australia

[5] Health Economics and Decision Science, ScHARR, University of Sheffield, Sheffield, UK

Preference-based HRQoL instruments, which have become increasingly popular in recent years and are particularly relevant for health policy makers, further take into account the relative importance that society places on living in a particular health state and can facilitate the calculation of quality-adjusted life years (QALYs) for conducting cost-utility analysis [3]. In China, the EQ-5D-3L has been routinely embedded in the National Health Services Survey (NHSS) to measure and monitor the HRQoL of adults since 2008 [8]. However, there is currently a lack of children and adolescent-specific preference-based HRQoL instruments in China. Such information is an essential prerequisite for incorporating adolescent preferences into the planning and prioritising of health-related interventions to improve adolescent health.

Among a total of nine available preference-based HRQoL instruments available globally for application with paediatric populations [9], the Child Health Utility 9D (CHU9D) is unique in that it represents the only generic preference-based paediatric HRQoL instrument to date developed exclusively from its inception with young people [10]. The CHU9D has nine dimensions (worried, sad, pain, tired, annoyed, schoolwork, sleep, daily routine and ability to join in activities) with five levels within each dimension. The instrument has undergone psychometric testing and been validated to be used with children and adolescents aged 7–17 years old [11–14]. In 2013, a pilot study with 815 students (aged 9–19 years old) was conducted in Nanjing City, the capital of Jiangsu Province, Mainland China. Results from the pilot study support the feasibility and construct validity of the Chinese version CHU9D (CHU9D-CHN) for measuring and valuing the HRQoL of children and adolescents in China [15]. The translation process for the official CHU9D-CHN questionnaire was designed in line with the recommendations of the ISPOR Task Force for translation and cultural adaptation of patient-reported outcome measures [16], which include dual forward translation, reconciliation, dual back translation including back translation review, followed by cognitive debriefing with participants from the relevant patient group. The translation also underwent proofreading and format check steps prior to finalisation.

Stated preference discrete choice experiments (DCEs) [17], which have their theoretical foundations in random utility theory, have recently become a widely used approach to health state valuation [18–20]. Profile case best–worst scaling (BWS) DCE represents a potentially easier choice task for health state valuation with respondents than traditional DCEs [21]. In BWS, a respondent is required to specify the best and the worst attribute for single health states only presented one at a time. A feasibility study conducted by Ratcliffe et al. [22] compared three different eliciting methods for the valuation of the CHU9D and found that the ordinal method (i.e. the BWS used in the study) for health state valuation was more easily understood and interpreted by children (aged 11–13 years old) than conventional approaches [i.e. standard gamble (SG) and time trade-off (TTO)].

The main objective of this study was to derive children and adolescents' preferences for health states defined by the CHU9D-CHN instrument, thereby enabling the views of young people to be incorporated into the health care priorities decision-making process in China. A secondary objective was to further investigate the potential for scale heterogeneity between best and worst responses from children and adolescents.

Methods

Surveys

Two independent valuation studies were conducted, including a profile case BWS survey which was used to derive children and adolescents' strength of preferences for a series of CHU9D health states, and a TTO survey which was used to facilitate a second-stage re-scaling of the BWS estimates onto the QALY scale. The methods adopted for the Chinese valuation study were similar to the Australian valuation study with two notable exceptions. Firstly, in the Chinese valuation study, students were randomly selected and completed a hard copy of the survey in a classroom setting, as opposed to the online version of the survey completed by an online panel community sample in Australian study, and (2) only one round of best–worst was adopted for the Chinese study, whereas two rounds of best–worst questions (i.e. selecting the best, the worst, followed by the second best and the second worst) were utilised in the Australian study.

BWS survey design

The CHU9D classification system consists of nine dimensions with five levels attached to each dimension. Accordingly, a total of $5^9 = 1{,}953{,}125$ possible health states are defined by the CHU9D descriptive system. A fractional factorial design (with minor adjustments to eliminate a small number of implausible states) was applied to reduce the number of health states to be valued in the BWS survey [23]. A total of fifty health states were selected and blocked into five versions so that each participant answered a manageable ten best–worst questions in practice. Other components in the BWS survey include the CHU9D instrument (which was used to help participants familiarise with the CHU9D classification system) prior to the BWS experiment, several socio-demographic questions and participants' self-reported health status. See Appendix Fig. 2 for an example of the BWS task in Chinese.

TTO survey design

During the one-to-one face-to-face interview, participants were firstly asked to self-complete the CHU9D questionnaire with the objective of introducing participants to the wording and formatting of the CHU9D classification system. Secondly, a series of TTO tasks were conducted using an interviewer-administered mode of administration with a visual prop (TTO board) in the form of a sliding scale to represent life years (an example script see Appendix 4.5 of Brazier et al. [3]). Finally, participants self-recorded their socio-demographic characteristics. See Appendix Fig. 3 for an example TTO task.

Five CHU9D health states (reflected increasing levels of impairment on HRQoL plus the PITS state, i.e. the lowest level for all nine dimensions of the CHU9D) were selected in the conventional TTO task for interviewer administration, following a standard TTO elicitation protocol documented by Brazier et al. [3, 24]. In brief, for a CHU9D health state considered to be better than dead the TTO task involves participants identifying a point of indifference between two alternatives. Alternative one specifies living for 10 years in one of the five CHU9D health states to be valued, whilst alternative two involves full health for time period x where x is less or equal to 10 years. Time x is varied until the respondent is indifferent between the two alternatives. The TTO utility score is calculated as $x/10$.

In instances where a CHU9D health state was considered as worse than death (WTD), a modified version of the TTO task was employed, in which alternative one is immediate death and alternative two is spending a length of time (y) in the CHU9D health state under consideration followed by (10-y) years in full health. Time y is varied until the respondent is indifferent between the two alternatives. The TTO utility score for state WTD is then calculated as $y/10-1$ [25].

Participants

The target sample for the BWS survey was primary and high school students aged between 9 and 17 (grades 4–12). The students were recruited using a multi-stage sampling method. In the first stage, two out of six urban districts were randomly selected. Next, within each chosen urban district one school was chosen from primary, junior and senior high schools, separately. Finally, one class was randomly chosen from each grade within each selected school. This resulted in a total of eighteen classes with two classes from primary, junior and senior high schools (grades 4–12), respectively. All students ($N=923$) within those selected classes were eligible participants. The survey was conducted in October 2013.

Due to ethical concerns relating to the presentation of scenarios involving immediate death within the conventional TTO task with adolescent samples in addition to the cognitive complexities associated with the TTO task [22, 24], the participants completing the TTO survey were a convenience sample of young adults recruited from a pool of first-year undergraduate students from Nanjing Medical University.

Data analysis

The BWS has its theoretical origins in the framework of random utility, which assumes that respondents choose the alternative that maximises their utility. Let U_{iq} represent the utility individual q derives from choosing item i, utility can be written as:

$$U_{\mathrm{iq}} = x'_{\mathrm{iq}}\beta_{\mathrm{iq}} + \varepsilon_{\mathrm{iq}}, \tag{1}$$

where x is a vector of observed attributes (i.e. CHU9D attribute levels), β is a vector of coefficients reflecting the desirability of the attribute levels to be estimated, ε is a random component. Assuming that the random components are distributed extreme value type 1 (EV1), the choice data can be analysed using the conditional logit model.

The profile case BWS task provides two sources of data, i.e. best and worst choices. These two different choice responses may be pooled together for analysis if preferences from the best and worst choices are assumed to be drawn from the same underlying utility (where best represents the highest utility and worst the lowest utility), accounting for the existence of scale heterogeneity across context [26]. An informal investigation of poolability of the best and worst data was firstly conducted by plotting the conditional logit estimates from the best data against the worst data. The Swait and Louviere test was further conducted to formally test for poolability [26].

According to the latent class modelling framework, individual preference heterogeneity is treated as a discrete distribution, i.e. a set of unobservable 'classes'. Class membership is latent in that each respondent belongs to each class up to a modelled probability. Within each class choice probabilities are assumed to be generated by the multinomial logit model. Maximum likelihood estimation is used to fit the best possible model with a predetermined number of classes set by the researcher. The Bayesian information criterion (BIC) can be used to help guide model selection. The final optimal latent class estimates are transformed onto the 0–1 scale by simply subtracting one-ninth of the score of the state '111111111' from each estimate, then dividing by the utility of state '555555555'. To account for the latent class membership, the average values across classes were calculated, by taking the mean of the sets of each class estimates, weighted by probability of class membership. Latent class analysis was conducted by using Latent GOLD® 5.1 software [27].

Since the BWS approach generates raw health state values on an ordinal scale only, it was necessary to adopt a secondary TTO sub-study to generate cardinal values for a series of CHU9D states for the purposes of re-scaling the raw estimates onto the 0–1 death-full health QALY scale [28]. Two re-anchoring approaches were explored here. Approach one re-anchored the BWS estimates using the mean TTO score of the PITS health state (the health state comprising the lowest level on each of the nine attributes of the CHU9D descriptive system) [28]. Approach two involved a mapping analysis, whilst DCE estimates are mapped onto the corresponding mean TTO scores using the ordinary least square (OLS) estimator [29], see Eq. 2:

$$\mathrm{dTTO}_{\mathrm{j}} = \gamma \mathrm{dBWS}_{\mathrm{j}} + \mu_{\mathrm{j}}, \qquad (2)$$

where dTTO equals to 1—TTO score, dBWS equals to 1—BWS estimates, γ is the key parameter to be estimated using OLS, μ is the error term, j indicates health states. A commonly used goodness-of-fit measure in mapping studies, the mean absolute error (MAE), was used to select the optimal re-scaling approach.

Results

Respondents' characteristics

Respondents' characteristics from the BWS and TTO surveys, respectively, are reported in Table 1. A total of 905 (98%) students consented to participate the BWS survey within a classroom setting, among them responses from three participants were excluded as they failed to engage with the task (consistently specifying the same attribute level to be both the best and the worst health state in the survey). A total of 40 undergraduate students took part in the TTO survey, among them two students were excluded as they failed to complete the TTO task.

The analysis samples comprised 902 students (mean age 13 years, ranged 9–17 years) and 38 students (mean age 18 years, ranged 18–19 years) for the BWS and TTO survey, respectively. Among the BWS/TTO survey respondents, 57%/47% were boys, 46%/34% had at least one parent who had received tertiary education, 66%/68% reported their self-assessed health status to be excellent or good in the past 1 month. In relation to survey difficulty, 6%/none respondents felt the BWS/TTO survey was very difficult, whilst 37%/63% respondents felt the BWS/TTO survey was not difficult.[1] A summary of responses to the CHU9D instrument is presented in Table 2. There were some notable consistencies between the most and the least impaired CHU9D dimensions between respondents from two surveys. The top three dimensions that respondents reported the strongest impairments in were 'tired', 'worried' and 'schoolwork/homework' in the BWS survey, as compared to 'worried', 'schoolwork/homework', and 'tired' in the TTO survey. In contrast, the top three dimensions that respondents reported having the least problems with were 'daily routine', 'annoyed' and 'sad' in the BWS survey, as compared to 'annoyed', 'daily routine' and 'sad' in the TTO survey.

Best–worst scaling choice experiment data analysis

The responses to the best and the worst questions were firstly analysed separately using a conditional logit model. The estimated coefficients were then plotted against each other classified according to each of 9 CHU9D dimensions (Fig. 1). If two response sources can be pooled together, it is expected that the coefficients should be linearly proportional across data sources, such as shown in 'tired', and 'schoolwork/homework' dimensions in Fig. 1. However, it can be seen this was not the case for all nine dimensions. The Swait–Louviere test was further conducted to formally test the poolability of two data sources. The log-likelihoods of the conditional logit estimates on the best/worst, and the heteroscedastic conditional logit estimate on the pooled data were $-15558.996/-17603.569$ and -33507.179, respectively. This produced a $\chi^2 = 689.228$, larger than criteria value of 52.192 (with 37 degrees of freedom), thus rejecting the null hypothesis that the parameters of two sources were the same whilst permitting the scale factor to vary.[2] Consistent with the approach adopted for data analysis in the Australian adolescents' sample, the remainder of the data analysis for the BWS survey focused on individual responses for the best data only.[3]

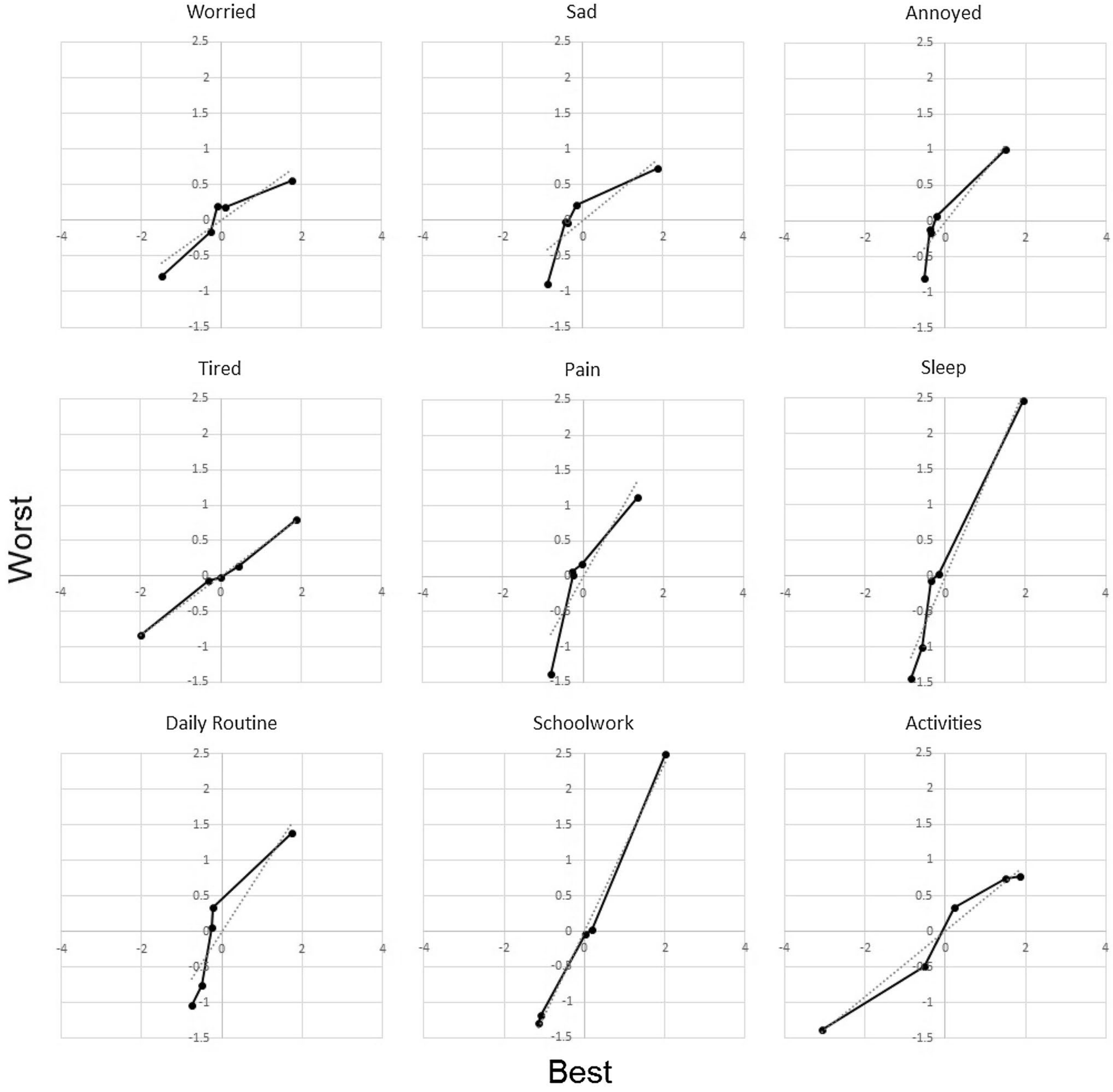
The latent class analysis results indicated that respondent preference patterns could be grouped into two main latent classes[4] (Table 3). Class 1 students (57%) placed the

[1] It should be noted that additional three students failed to complete the BWS choice experiment, and additional two students cannot complete the TTO tasks.

[2] Considering the well-documented issue relating to the sensitivity of the log-likelihood ratio test, we have also tried to relax the constraint of some dimension level parameters (e.g. the 5th level of 'activities' dimension) in estimating the pooled BWS data, i.e. to allow estimated coefficients differ between the best and the worst. However, the conclusion to not pool the best and the worst data remains the same based on the Swait–Louviere test.

[3] We have also estimated the scale-adjusted LC model and the log-scale factor estimated for the worst choices was -0.3 ($p < 0.001$), indicating the worst choices exhibit less consistency (i.e. $\exp(-0.3) = 0.74$) across choice sets than the best choices.

[4] Initially 2–4 classes were considered. The BIC values decreased along with the increased number of classes being specified (i.e. more class is favourable based on the information criteria); however, increasingly more inconsistencies among estimated dimension levels were observed in the 3- or 4-class case. Further considering the % reductions on the likelihood ratio chi-squared statistic L^2 of a 7% reduction from 1-class to 2-classes, versus a 2%/1% from 2- to 3- or 3- to 4-class specifications. The 2-class case was selected as the optimal result, similar to the Australian adolescent study result reported by Ratcliffe et al. [23].



Y-axis: Choice model coefficient values based on worst choices
X-axis: Choice model coefficient values based on best choices
Solid lines: Connection lines between two points
Dashed lines: Linear trend lines (best-fit straight lines)

Fig. 1 Plot of conditional logit coefficients for best versus worst choices

strongest importance on ‘activities’, followed by ‘tired’, and ‘worry’, and the least importance on ‘annoyed’, ‘daily routine’ and ‘sleep’. On the other hand, Class 2 students (43%) placed ‘schoolwork/homework’ as the most important dimension, followed by ‘sleep’ and ‘worry’, and placed the least importance on ‘pain’, ‘sad’ and ‘activities’ dimensions.

Table 1 Characteristics of respondents

Characteristic	BWS survey $N = 902$	TTO survey $N = 38$
Age, years		
Mean (standard deviation)	13.28 (2.59)	18.47 (0.73)
Gender, %		
Boys	56.98	47.37
Girls	43.02	52.63
Grade level (BWS survey only), %		
4 (~ 9 years old)	10.64	
5 (~ 10 years old)	9.53	
6 (~ 11 years old)	9.2	
7 (~ 12 years old)	13.19	
8 (~ 13 years old)	11.42	
9 (~ 14 years old)	10.86	
10 (~ 15 years old)	14.08	
11 (~ 16 years old)	9.87	
12 (~ 17 years old)	11.2	
Parental education, %		
Mother and/or father has tertiary qualification	45.57	34.21
Self-assessed health status (past month), %		
Excellent	28.71	13.16
Good	37.36	55.26
Fair	28.82	26.32
Poor or very poor	5.11	5.26
Difficulty of the survey, %		
Very difficult	6.43	0
Moderately difficult	18.40	2.63
Slightly difficult	38.47	34.21
Not difficult	36.70	63.16

BWS best worst scaling, *TTO* time trade-off

As can be seen in Table 3, for both classes, all 9 CHU9D dimensions were statistically significant in explaining the best preferences of children and adolescents in Mainland China.

In the HRQoL valuation literature, owing to the nature of the ordinal response levels of each dimension, it is expected that the lower levels of functioning should be associated with larger decrements in health state values. This is the case for the vast majority of estimates; however, inconsistencies.[5] were also evident in the latent class estimates (e.g. the estimates on the 4th and 5th levels of 'activities' dimensions in Class 2 were – 0.6058 and – 0.5840).

Following the approach adopted for the Australian study, a decision rule was adopted to restrict or equate the inconsistent estimates if they were not statistically different from each other[6] Further, statistically insignificant levels were constrained to be zero. The final model with collapsed levels is presented in Table 4, with the last column presenting the final transformed score on a 0–1 scale (i.e. the CHU9D health states '111111111' and '555555555' are scored as 1 and 0, respectively).

Time trade-off data analysis and re-scaling onto the QALY scale

The TTO utility scores for the five selected CHU9D health states are presented in Table 5. The worst CHU9D health state '555555555' had a mean utility of – 0.0855, with 16 (42%) of respondents considering it to be worse than being dead. As previously described, two methods were used to re-scale the final BWS scores onto the QALY scale. Based on the goodness-of-fit criterion MAE, the method 2 mapping approach (with an estimated γ in Eq. 2 of 0.9437, $p < 0.001$) achieved a better performance and is recommended (detailed at Table 5).

Calculating CHU9D-CHN utility scores

Combining the BWS estimates and the re-anchor results, the CHU9D-CHN utility scores can be calculated in two steps. Take calculating a health state of "432154321" as an example:

(Step 1, from the last column of Table 4) A summary score can be calculated as:

$0.0569 + 0.0447 + 0.0590 + 0.1131 + 0.0276 + 0.0125 + 0.0485 + 0.0486 + 0.1072 = 0.5181.$

(Step 2, from re-anchor equation) The CHU9D-CHN utility score can be calculated as:

$\text{CHU9D-CHN} = 1 - 0.9437 \times (1 - 0.5181) = 0.5452.$

Discussion

By combining data from a BWS choice experiment and a TTO survey, this study has elicited young people's preferences for key dimensions of HRQoL in Mainland China and has generated the Chinese-specific tariff for the CHU9D-CHN instrument that can be used to calculate health state utility scores. This study represents the first valuation study

[5] The inconsistencies have been widely reported in the valuation studies regardless of which valuation techniques been adopted [23, 42].

[6] The adjustment was conducted within each class and with one inconsistence dimension level at a time. The adjustment slightly changed the proportions of the 2 class membership, from 57.32% versus 42.68–57.98% versus 42.02%. The BIC values were 29993.68 in the main model and 29834.74 in the final model.

Table 2 Responses to the Child Health Utility 9D (CHU9D), %

CHU9D dimensions and levels	BWS survey	TTO survey
	N = 902	*N* = 38
1. Worried (发愁)		
I don't feel worried today	34.9	21.1
I feel a little bit worried today	32.4	44.7
I feel a bit worried today	17.3	23.7
I feel quite worried today	9.8	7.9
I feel very worried today	5.7	2.6
2. Sad (伤心)		
I don't feel sad today	68.4	86.8
I feel a little bit sad today	20.6	7.9
I feel a bit sad today	6.0	5.3
I feel quite sad today	2.7	0.0
I feel very sad today	2.3	0.0
3. Pain (疼痛)		
I don't have any pain today	52.0	68.4
I have a little bit of pain today	29.3	21.1
I have a bit of pain today	11.1	7.9
I have quite a lot of pain today	6.1	2.6
I have a lot of pain today	1.6	0.0
4. Tired (累)		
I don't feel tired today	29.3	29.0
I feel a little bit tired today	36.1	55.3
I feel a bit tired today	15.2	13.2
I feel quite tired today	9.1	2.6
I feel very tired today	10.3	0.0
5. Annoyed (恼怒)		
I don't feel annoyed today	69.7	89.5
I feel a little bit annoyed today	19.6	10.5
I feel a bit annoyed today	4.3	0.0
I feel quite annoyed today	3.1	0.0
I feel very annoyed today	3.2	0.0
6. Schoolwork/homework (课堂作业/家庭作业)		
I have no problems with my schoolwork/homework today	48.1	29.0
I have a few problems with my schoolwork/homework today	37.5	42.1
I have some problems with my schoolwork/homework today	11.0	18.4
I have many problems with my schoolwork/homework today	2.2	7.9
I can't do my schoolwork/homework today	1.2	2.6
7. Sleep (睡觉)		
Last night, I had no problems sleeping	62.3	71.1
Last night, I had a few problems sleeping	22.1	18.4
Last night, I had some problems sleeping	9.4	5.3
Last night, I had many problems sleeping	4.2	5.3
Last night, I couldn't sleep at all	2.0	0.0
8. Daily routine (日常生活)		
I have no problems with my daily routine today	88.4	89.5
I have a few problems with my daily routine today	8.9	7.9
I have some problems with my daily routine today	1.7	0.0
I have many problems with my daily routine today	0.8	2.6
I can't do my daily routine today	0.3	0.0

Table 2 (continued)

CHU9D dimensions and levels	BWS survey $N = 902$	TTO survey $N = 38$
9. Able to join in activities (参加活动的能力)		
I can join in with any activities today	60.1	39.5
I can join in with most activities today	24.4	23.7
I can join in with some activities today	6.7	21.1
I can join in with a few activities today	5.1	10.5
I can join in with no activities today	3.8	5.3

BWS best worst scaling, *TTO* time trade-off

to derive children and adolescent-specific preference weights in China. To date, only two other valuation studies have been conducted in Mainland China both focused on adult populations and employed the traditional TTO valuation technique [30, 31]. Hence, in comparison with other countries, most notably the UK, these types of large scale health state valuation studies remain embryonic in China.

A homogeneous preference assumption has been mainly adopted in the health state valuation literature previously. However, several recent studies conducted separately in both adults and adolescents have found the existence of preference heterogeneity in health state valuation [19, 23, 32]. Comparing the latent classes identified in this study with classes arising from the Australian CHU9D valuation study,[7] some differences are evident in the identified latent classes. The main reasons for these discrepancies are not entirely clear but may be due to cultural differences and/or differences in the education systems operating in these two countries.

Since the CHU9D was first developed and introduced for application in health economics and health technology appraisal processes in 2009, it has gained popularity internationally as a paediatric-specific preference-based HRQoL instrument [33–35]. Two existing scoring algorithms are available for the CHU9D instrument, both of which were developed from English-speaking countries, UK and Australia.[8] It is out of the scope of this particular study to investigate to what extent the above methodological differences explain the variations in the health state values pertaining to the application of each scoring algorithms and health state values. However, it is evident that each of these scoring algorithms cannot be used interchangeably with each other. Take the direct valuation for the CHU9D PITS state as an example, the UK adult value for the PITS state based on the SG method is 0.34, as compared to −0.21 for the Australian adolescent value and −0.09 for the Chinese adolescent value both based on an identical TTO protocol. The potential utility gain from the PITS state to full health is therefore largest where the Australian adolescent scoring algorithm is adopted, followed by the Chinese and UK scoring algorithms.

Profile case BWS has gained popularity in recent years in the health state valuation literature. The completion of a BWS task provides two sources of data (i.e. the best and worst response) for each health state under consideration. However, two CHU9D valuation studies which were conducted in Australia and China with different cultures and were administrated through different methods (online and hardcopy questionnaire within a classroom setting) both found that the best and the worst choices cannot be pooled together for analysis. Differences in choice consistency arise from differences in the size of the random utility components on the latent scale and may arise as a result of differences in participants' certainty of their preferences, and/or different decision rules are adopted in decision-making processes. The poolability was not explicitly studied/reported in other instruments' valuation studies from adult respondents, nor was it reported in a disease control preference study in adolescents using BWS [19, 32, 36]. It is therefore unclear whether this conclusion of non-poolability is unique for adolescent samples in particular or whether it relates specifically

[7] In Australian study, respondents in Class 1 (63%) placed the most important weights on mental health dimensions and the least important weights on activities, and daily routine. In Class 2 (37%), respondents placed equal weights on all dimensions.

[8] Comparison of these established scoring algorithms, with the development of the new Chinese adolescent-specific scoring algorithm documented in this paper indicates some key methodological differences including (i) elicitation methods been used (SG vs. BWS & TTO), (ii) sample size, (iii) whose values were elicited from (adults vs. adolescents), (iv) source of respondents (random street/student sample vs. online panel), (v) the mode of survey delivery

Footnote 8 (continued)

(face-to-face interview vs. online survey vs. hard copy questionnaire), (vi) the utility function to be estimated (whether potential interaction terms were considered and tested), and (vii) econometric techniques used to estimate the utility function (OLS vs. latent class modelling).

Table 3 Latent class estimates—main model with no restrictions ($N = 902$)

Attributes	Class 1	Class 2	Mean
Worry 2	1.4326***	−0.0487	0.8005
Worry 3	0.7788***	0.1390	0.5058
Worry 4	0.7139***	−0.1702	0.3366
Worry 5	−6.7221***	−0.6937***	−4.1494
Sad 2	0.9550***	−0.2456**	0.4426
Sad 3	0.4237*	−0.2600**	0.1320
Sad 4	0.3979	−0.2613**	0.1166
Sad 5	−5.5695***	−0.2289*	−3.2903
Pain 2	0.3532	0.1389	0.2617
Pain 3	0.2872	−0.2502**	0.0579
Pain 4	−0.1435	−0.0451	−0.1015
Pain 5	−3.6918***	−0.3844***	−2.2803
Tired 2	1.7919***	0.0000	1.0272
Tired 3	1.5466***	−0.0893	0.8485
Tired 4	1.2245***	−0.2980***	0.5747
Tired 5	−8.4993***	−0.5363***	−5.1009
Annoy 2	−0.3528	−0.0538	−0.2252
Annoy 3	−0.8863***	−0.1045	−0.5526
Annoy 4	−1.2151***	−0.1608	−0.7652
Annoy 5	−0.7696***	−0.5182***	−0.6623
Schoolwork 2	1.5059***	0.0461	0.8829
Schoolwork 3	0.7789***	0.1475	0.5094
Schoolwork 4	−4.8521***	−0.5668***	−3.0233
Schoolwork 5	−1.2844***	−0.7914***	−1.0740
Sleep 2	0.2646	−0.1098	0.1048
Sleep 3	−0.4728*	−0.0745	−0.3029
Sleep 4	−1.8448***	−0.4572***	−1.2526
Sleep 5	−1.7573***	−0.4751***	−1.2101
Daily 2	0.2576	−0.1550	0.0815
Daily 3	−0.3782	−0.1001	−0.2595
Daily 4	−1.8777***	−0.2706**	−1.1918
Daily 5	−1.5662***	−0.4593***	−1.0938
Activities 2	3.4228***	0.7272***	2.2724
Activities 3	2.3030***	−0.2174*	1.2274
Activities 4	1.0335***	−0.6058***	0.3339
Activities 5	−10.7956***	−0.5840***	−6.4376
Class membership	57.32%	42.68%	
LL	−14748.47		
BIC	29993.68		

Effects coding was used. Reference level: first level of each dimension. Mean class estimates weighted by probability of class membership

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

to this particular CHU9D instrument and its descriptive system. Future studies should investigate in more detail the decision rules that adolescents adopt in making the best vs. the worst choices for health state valuation.

Table 4 Latent class estimates—final model with collapsed levels and monotonicity ($N = 902$)

Attributes	Class 1	Class 2	Rescaled on a 0–1 scale
Worry 1	3.5989***	0.7456***	0.1077
Worry 2	1.2177***	0	0.0630
Worry 3	0.5412**	0	0.0573
Worry 4	0.4900***	0	0.0569
Worry 5	−5.8478***	−0.7456***	−0.0208
Sad 1	3.5980***	0.9757***	0.1154
Sad 2	0.7672***	−0.2449**	0.0510
Sad 3	0	−0.2436***	0.0447
Sad 4	0	−0.2436***	0.0447
Sad 5	−4.3652***	−0.2436***	0.0083
Pain 1	2.9921***	0.5325***	0.0955
Pain 2	0.1490*	0.1490*	0.0590
Pain 3	0	−0.1469**	0.0479
Pain 4	0	−0.1469**	0.0479
Pain 5	−3.1411***	−0.3877***	0.0137
Tired 1	3.7556***	0.8678***	0.1131
Tired 2	1.6200***	0	0.0663
Tired 3	1.3448***	0	0.0640
Tired 4	1.0360***	−0.3147***	0.0509
Tired 5	−7.7564***	−0.5531***	−0.0303
Annoy 1	3.0397***	0.8219***	0.1056
Annoy 2	−0.4146*	0	0.0494
Annoy 3	−0.8347***	−0.1446**	0.0410
Annoy 4	−0.8952***	−0.1446**	0.0405
Annoy 5	−0.8952***	−0.5327***	0.0276
Schoolwork 1	3.6391***	1.1889***	0.1229
Schoolwork 2	1.2784***	0	0.0635
Schoolwork 3	0.5869***	0	0.0577
Schoolwork 4	−2.7522***	−0.5208***	0.0125
Schoolwork 5	−2.7522***	−0.6681***	0.0076
Sleep 1	3.6181***	1.0816***	0.1191
Sleep 2	0	0	0.0528
Sleep 3	−0.5239**	0	0.0485
Sleep 4	−1.4064***	−0.5386***	0.0231
Sleep 5	−1.6878***	−0.5430***	0.0206
Daily 1	3.3744***	0.9689***	0.1133
Daily 2	0	−0.1271*	0.0486
Daily 3	0	−0.1271*	0.0486
Daily 4	−1.6872***	−0.2624**	0.0300
Daily 5	−1.6872***	−0.4523***	0.0236
Activities 1	3.8360***	0.6711***	0.1072
Activities 2	3.2332***	0.7122***	0.1036
Activities 3	2.1139***	−0.2153*	0.0632
Activities 4	0.8619***	−0.5840***	0.0405
Activities 5	−10.0450***	−0.5840***	−0.0503
Class membership	57.98%	42.02%	
LL	−14760.86		
BIC	29834.74		

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table 5 Time trade-off (TTO) utility scores for five selected CHU9D health states and rescaled utility scores

CHU9D health states	TTO mean	TTO SD	BWS mean	Rescaled scores, based on PITS value (method 1)	Rescaled scores, mapping approach (method 2)
434243545	0.4465	0.2591	0.3143	0.2557	0.3529
414355432	0.5568	0.2416	0.4946	0.4514	0.5231
231345314	0.6651	0.2226	0.5176	0.4763	0.5448
423141114	0.7304	0.2180	0.7053	0.6801	0.7219
555555555 (PITS)	−0.0855	0.4200	0	−0.0855	0.0563
MAE (range)	–	–	–	0.1070 (0–0.1908)	0.0796 (0.0085–0.1418)

MAE mean absolute error

Adopting BWS to value health states has been demonstrated to be a feasible approach for health state valuation for children aged 7 years and older. However, the BWS is not without limitations. First, BWS estimates are ordinal and are not generated on the QALY scale. In contrast to a conventional DCE task, in which survival duration can be added as a separated attribute, length of life is not specified in a BWS task [37]. However, owing to the relatively large descriptive system of the CHU9D (i.e. nine dimensions vs. five dimensions of the EQ-5D), using DCEs with survival duration as an additional attribute may not be feasible for children and adolescents. As such a second sub-study utilising a direct cardinal approach to health state valuation (e.g. the TTO or SG) is required to re-scale the BWS estimates on to the QALY scale. Usually those who participate in the direct valuation survey will be a different population and more mature (e.g. young adults in this study). In this regard, the second source of preferences will need to be incorporated into the valuation procedure to generate the final scoring algorithm.[9]

Secondly, only a main-effect model can be estimated using profile case BWS data [19], that is to say, the potential correlations among different HRQoL dimensions (i.e. the multiplicative model) will not be taken into account in empirically estimating the utility function. HRQoL is a multi-dimensional construct, and dimensions are likely to be correlated with each other to some extent (e.g. 'worry' and 'sad' dimensions in CHU9D). The main-effect model may omit potential significant interaction terms among dimension levels that should be taken into account; however, empirical evidence from previous valuation studies for other preference-based HRQoL instruments has reached divergent conclusions [38–40]. The UK CHU9D valuation study based on SG task is the most relevant one to this particular study, and Stevens [40] found that including interaction terms led to an increased chance of inconsistencies within attributes and a decreased number of significant attribute levels that were estimated.

There are several limitations to this study. Firstly, the choice to conduct the valuation survey in Nanjing could be regarded as an outcome of convenience sampling. Nanjing is the capital city of Jiangsu Province and one of the most important cities in the history of China. It is where the initial CHU9D-CHN pilot study was conducted. It should be noted that within the sampling city, a multi-stage random sampling technique was used to select participants for the BWS survey. Secondly, the generalisability of the tariff should be further investigated among ethnic minorities in China. The vast majority of participants in this study are Han Chinese. Although Han ethnic group accounts for 92% of total population in 2010 in Mainland China, it is unclear to what extent their preferences can be representative for the other 55 distinct ethnic groups. Furthermore, since all participants of this study were recruited from urban areas, it is unclear to what extent their preferences would be the same as those who lived in rural areas. A recent empirical study in Western China demonstrates that the CHU9D-CHN (scored based on this Chinese-specific tariff) is a reliable and valid instrument for measuring HRQoL of students aged 8–17 years old in rural areas [41]. Lastly, the sample size for the TTO survey and the health states which were valued in the TTO survey are relatively small.

In conclusion, this is the first study to generate a Chinese-specific tariff for a preference-based HRQoL instrument with a children and adolescent sample. It provides further methodological insights into preference heterogeneity for health state valuation. The scoring algorithm reported in this study can facilitate economic evaluation and health

[9] Two re-anchoring approaches were adopted in this study. Re-anchoring onto the PITS approach arguably uses less information from the second source of preference (i.e. only elicited utility score from the worst health state was used), as compared to the mapping approach (in which a series of elicited utility scores were all used); however, that does not necessary imply the preferential choice of the re-anchoring approach. Better goodness-of-fit (or less prediction errors) from the mapping approach in this study indicates that empirically it is a preferred approach. Whether this conclusion holds in future studies is unclear, but it is evident that re-scaling based on the PITS state only may not be an optimal approach.

technology assessment in China for health services/interventions targeted at young paediatric populations.

Acknowledgements The paper has benefited from participants at the International Society for Quality of Life Research (ISOQOL) 23rd Annual Conference. We are grateful to Chao Li, Qin Ye, Hairong Zhou, Zhenzhen Qin, Shenxiang Qi, Huafeng Yang, Xupeng Chen from Nanjing Municipal CDC for their assistants with data collection. Responsibility for any remaining errors lies solely with the authors.

Funding The study was supported by the Nanjing Municipal Science and Technique Foundation (ZDX12019), China.

Compliance with Ethical Standards

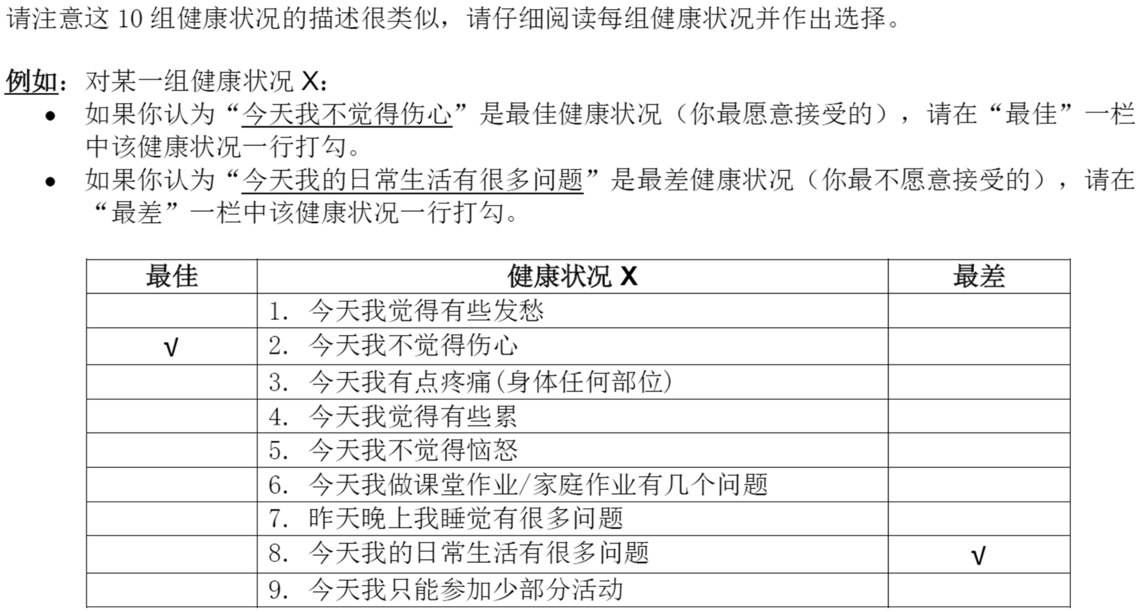
Conflict of interest KS is the developer of the CHU9D and took a small royalty in 2016 for a commercial licence. All other authors declare that they have no conflict of interest.

Ethical Approval The study protocol was reviewed and approved by the academic and ethical committee of Nanjing Municipal Center for Disease Control and Prevention. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed Consent All participants to both surveys provided written consent prior to participation.

Appendix

Figures 2 and 3.



请注意这10组健康状况的描述很类似，请仔细阅读每组健康状况并作出选择。

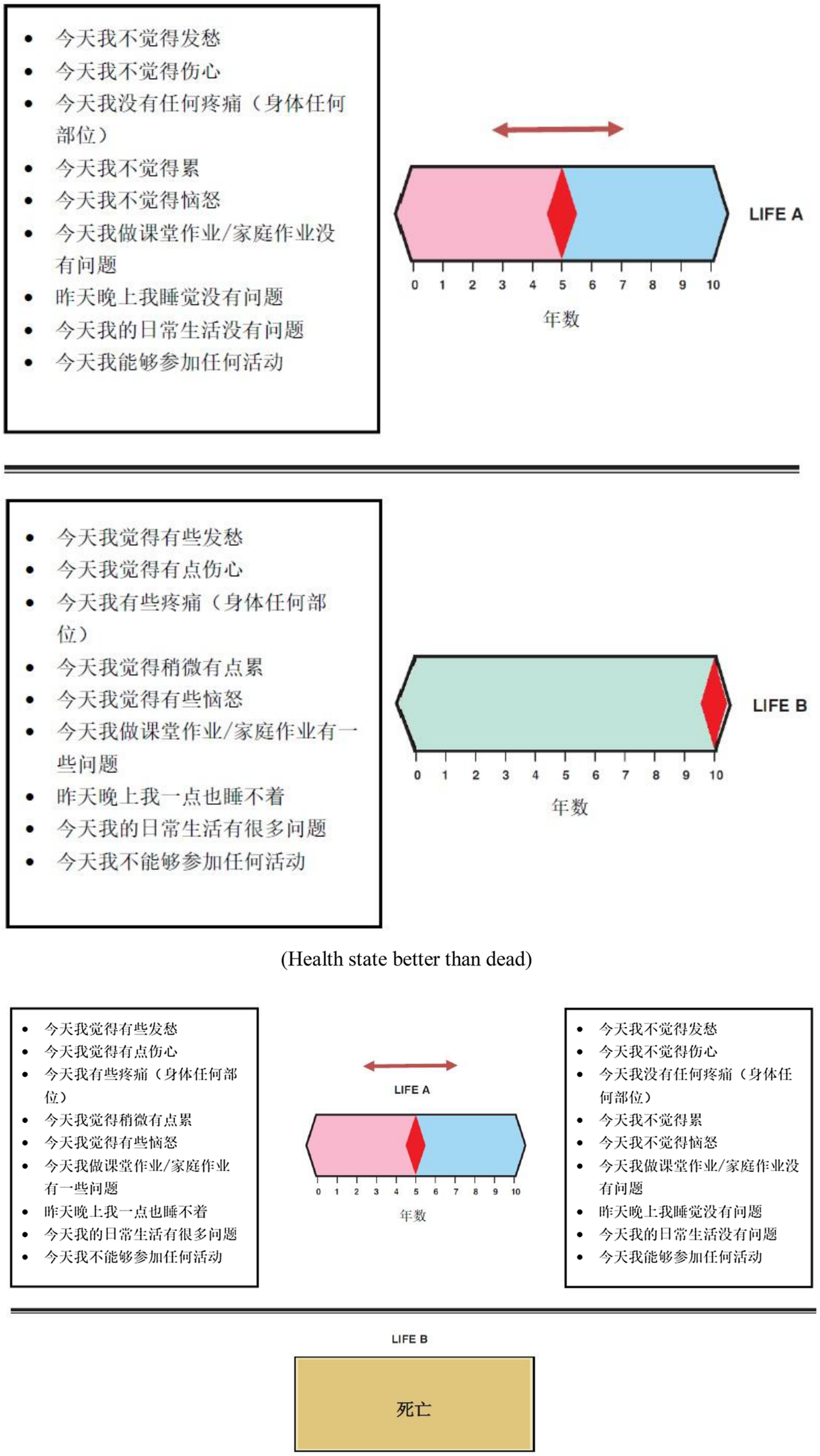
例如：对某一组健康状况X：

- 如果你认为“今天我不觉得伤心”是最佳健康状况（你最愿意接受的），请在“最佳”一栏中该健康状况一行打勾。
- 如果你认为“今天我的日常生活有很多问题”是最差健康状况（你最不愿意接受的），请在“最差”一栏中该健康状况一行打勾。

最佳	健康状况 X	最差
	1. 今天我觉得有些发愁	
√	2. 今天我不觉得伤心	
	3. 今天我有点疼痛(身体任何部位)	
	4. 今天我觉得有些累	
	5. 今天我不觉得恼怒	
	6. 今天我做课堂作业/家庭作业有几个问题	
	7. 昨天晚上我睡觉有很多问题	
	8. 今天我的日常生活有很多问题	√
	9. 今天我只能参加少部分活动	

Fig. 2 Instruction example of the best–worst scaling task in Chinese

Fig. 3 Time trade-off example in Chinese



- 今天我不觉得发愁
- 今天我不觉得伤心
- 今天我没有任何疼痛（身体任何部位）
- 今天我不觉得累
- 今天我不觉得恼怒
- 今天我做课堂作业/家庭作业没有问题
- 昨天晚上我睡觉没有问题
- 今天我的日常生活没有问题
- 今天我能够参加任何活动

- 今天我觉得有些发愁
- 今天我觉得有点伤心
- 今天我有些疼痛（身体任何部位）
- 今天我觉得稍微有点累
- 今天我觉得有些恼怒
- 今天我做课堂作业/家庭作业有一些问题
- 昨天晚上我一点也睡不着
- 今天我的日常生活有很多问题
- 今天我不能够参加任何活动

(Health state better than dead)

- 今天我觉得有些发愁
- 今天我觉得有点伤心
- 今天我有些疼痛（身体任何部位）
- 今天我觉得稍微有点累
- 今天我觉得有些恼怒
- 今天我做课堂作业/家庭作业有一些问题
- 昨天晚上我一点也睡不着
- 今天我的日常生活有很多问题
- 今天我不能够参加任何活动

- 今天我不觉得发愁
- 今天我不觉得伤心
- 今天我没有任何疼痛（身体任何部位）
- 今天我不觉得累
- 今天我不觉得恼怒
- 今天我做课堂作业/家庭作业没有问题
- 昨天晚上我睡觉没有问题
- 今天我的日常生活没有问题
- 今天我能够参加任何活动

(Health state worth than dead)

References

1. The Lancet. (2016). The best science for achieving Healthy China 2030. *Lancet, 388*(10054), 1851.
2. Frisén, A. (2007). Measuring health-related quality of life in adolescence. *Acta Paediatrica, 96*(7), 963–968.
3. Brazier, J., Ratcliffe, J., Salomon, J. A., & Tsuchiya, A. (2017). *Measuring and valuing health benefits for economic evaluation* (2nd ed.). Oxford: Oxford University Press.
4. Schwartzmann, L. (2009). Research and action: Toward good quality of life and equity in health. *Expert Review of Pharmacoeconomics & Outcomes Research, 9*(2), 143–147.
5. Huang, Y., Zhong, X.-N., Li, Q.-Y., Xu, D., Zhang, X.-L., Feng, C., et al. (2015). Health-related quality of life of the rural-China left-behind children or adolescents and influential factors: A cross-sectional study. *Health and Quality of Life Outcomes, 13*(1), 29.
6. Ji, Y., Chen, S., Li, K., Xiao, N., Yang, X., Zheng, S., & Xiao, X. (2011). Measuring health-related quality of life in children with cancer living in mainland China: Feasibility, reliability and validity of the Chinese Mandarin version of PedsQL 4.0 Generic Core Scales and 3.0 Cancer Module. *Health and Quality of Life Outcomes, 9*(1), 103.
7. Yang, X., Xiao, N., & Yan, J. (2011). The PedsQL in pediatric cerebral palsy: Reliability and validity of the Chinese version pediatric quality of life inventory 4.0 generic core scales and 3.0 cerebral palsy module. *Quality of Life Research, 20*(2), 243–252.
8. Zhou, Z., Fang, Y., Zhou, Z., Li, D., Wang, D., Li, Y., et al. (2017). Assessing income-related health inequality and horizontal inequity in China. *Social Indicators Research, 132*(1), 241–256.
9. Chen, G., & Ratcliffe, J. (2015). A review of the development and application of generic multi-attribute utility instruments for paediatric populations. *Pharmacoeconomics, 33*(10), 1013–1028.
10. Stevens, K. (2009). Developing a descriptive system for a new preference-based measure of health related quality of life for children. *Quality of Life Research, 18*(8), 1105–1113.
11. Stevens, K., & Ratcliffe, J. (2012). Measuring and valuing health benefits for economic evaluation in adolescence: An assessment of the practicality and validity of the Child Health Utility 9D in the Australian adolescent population. *Value in Health, 15*(8), 1092–1099.
12. Ratcliffe, J., Stevens, K., Flynn, T., Brazier, J., & Sawyer, M. (2012). An assessment of the construct validity of the CHU9D in the Australian adolescent general population. *Quality of Life Research, 21*(4), 717–725.
13. Chen, G., Flynn, T., Stevens, K., Brazier, J., Huynh, E., Sawyer, M., et al. (2015). Assessing the health-related quality of life of Australian adolescents: An empirical comparison of the Child Health Utility 9D and EQ-5D-Y instruments. *Value in Health, 18*(4), 432–438.
14. Petersen, K. D., Chen, G., Mpundu-Kaambwa, C., Stevens, K., Brazier, J., & Ratcliffe, J. (2018). Measuring health related quality of life in adolescent populations: An empirical comparison of the CHU9D and the PedsQL™ 4.0 Short Form 15. *The Patient, 11*(1), 29–37.
15. Xu, F., Chen, G., Stevens, K., Zhou, H., Qi, S., Wang, Z., et al. (2014). Measuring and valuing health-related quality of life among children and adolescents in Mainland China—A pilot study. *PLoS ONE, 9*(2), e89222.
16. Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., et al. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health, 8*(2), 94–104.
17. Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling, 3*(3), 57–72.
18. McCabe, C., Brazier, J., Gilks, P., Tsuchiya, A., Roberts, J., O'Hagan, A., & Stevens, K. (2006). Using rank data to estimate health state utility models. *Journal of Health Economics, 25*(3), 418–431.
19. Flynn, T. N., Huynh, E., Peters, T. J., Al-Janabi, H., Clemens, S., Moody, A., & Coast, J. (2015). Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. *Health Economics, 24*(3), 258–269.
20. Coast, J., Huynh, E., Kinghorn, P., & Flynn, T. (2016). Complex valuation: Applying ideas from the complex intervention framework to valuation of a new measure for end-of-life care. *Pharmacoeconomics, 34*(5), 499–508.
21. Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: Theory, methods and applications*. Cambridge: Cambridge University Press.
22. Ratcliffe, J., Couzner, L., Flynn, T., Sawyer, M., Stevens, K., Brazier, J., & Burgess, L. (2011). Valuing Child Health Utility 9D health states with a young adolescent sample: A feasibility study to compare best-worst scaling discrete-choice experiment, standard gamble and time trade-off methods. *Applied Health Economics and Health Policy, 9*(1), 15–27.
23. Ratcliffe, J., Huynh, E., Chen, G., Stevens, K., Swait, J., Brazier, J., et al. (2016). Valuing the Child Health Utility 9D: Using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. *Social Science & Medicine, 157*, 48–59.
24. Ratcliffe, J., Chen, G., Stevens, K., Bradley, S., Couzner, L., Brazier, J., et al. (2015). Valuing Child Health Utility 9D health states with young adults: Insights from a time trade off study. *Applied Health Economics and Health Policy, 13*(5), 485–492.
25. Dolan, P., Gudex, C., Kind, P., & Williams, A. (1996). The time trade-off method: Results from a general population study. *Health Economics, 5*(2), 141–152.
26. Swait, J., & Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research, 30*(3), 305–314.
27. Vermunt, J. K., & Magidson, J. (2016). *Upgrade manual for latent GOLD 5.1*. Belmont Massachusetts: Statistical Innovations Inc.
28. Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2010). Using discrete choice experiments to understand preferences for quality of life. Variance-scale heterogeneity matters. *Social Science & Medicine, 70*(12), 1957–1965.
29. Rowen, D., Brazier, J., & Van Hout, B. (2015). A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Medical Decision Making, 35*(3), 328–340.
30. Liu, G. G., Wu, H., Li, M., Gao, C., & Luo, N. (2014). Chinese time trade-off values for EQ-5D health states. *Value in Health, 17*(5), 597–604.
31. Luo, N., Liu, G., Li, M., Guan, H., Jin, X., & Rand-Hendriksen, K. (2017). Estimating an EQ-5D-5L value set for China. *Value in Health, 20*(4), 662–669.
32. Krucien, N., Watson, V., & Ryan, M. (2017). Is best-worst scaling suitable for health state valuation? A comparison with discrete choice experiments. *Health Economics, 26*(12), e1–e16.
33. Canaway, A. G., & Frew, E. J. (2013). Measuring preference-based quality of life in children aged 6–7 years: A comparison of the performance of the CHU-9D and EQ-5D-Y—The WAVES Pilot Study. *Quality of Life Research, 22*(1), 173–183.
34. Chen, G., Ratcliffe, J., Olds, T., Magarey, A., Jones, M., & Leslie, E. (2014). BMI, health behaviors, and quality of life in children and adolescents: A school-based study. *Pediatrics, 133*(4), e868–e874.

35. Boyer, N. R. S., Miller, S., Connolly, P., & McIntosh, E. (2016). Paving the way for the use of the SDQ in economic evaluations of school-based population health interventions: An empirical analysis of the external validity of SDQ mapping algorithms to the CHU9D in an educational setting. *Quality of Life Research, 25*(4), 913–923.
36. Ungar, W. J., Hadioonzadeh, A., Najafzadeh, M., Tsao, N. W., Dell, S., & Lynd, L. D. (2015). Parents and adolescents preferences for asthma control: A best-worst scaling choice experiment using an orthogonal main effects design. *BMC Pulmonary Medicine, 15*, 146.
37. Flynn, T. N. (2010). Using conjoint analysis and choice experiments to estimate QALY values. *Pharmacoeconomics, 28*(9), 711–722.
38. Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care, 35*(11), 1095–1108.
39. Viney, R., Norman, R., King, M. T., Cronin, P., Street, D. J., Knox, S., & Ratcliffe, J. (2011). Time trade-off derived EQ-5D weights for Australia. *Value in Health, 14*(6), 928–936.
40. Stevens, K. (2012). Valuation of the Child Health Utility 9D index. *Pharmacoeconomics, 30*(8), 729–747.
41. Yang, P., Chen, G., Wang, P., Zhang, K., Deng, F., Yang, H., & Zhuang, G. (2018). Psychometric evaluation of the Chinese version of the Child Health Utility 9D (CHU9D-CHN): A school-based study in China. *Quality of Life Research, 27*(7), 1921–1931.
42. Brazier, J., & Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care, 42*(9), 851–859.