

Assessing the Use of a Feedback Module to Model EQ-5D-5L Health States Values in Hong Kong

Eliza L. Y. Wong¹ · Juan Manuel Ramos-Goñi² · Annie W. L. Cheung¹ · Amy Y. K. Wong¹ · Oliver Rivero-Arias³ 

Published online: 10 October 2017
© The Author(s) 2017. This article is an open access publication

Abstract

Background An international valuation protocol exists for obtaining societal values for each of the 3125 health states of the five-level EuroQol-five dimensions (EQ-5D-5L) questionnaire. A feedback module (FM) that can be related to theoretical models used in behavioral economics was recently included in this protocol.

Objectives Our objective was to assess the impact of using an FM to estimate an EQ-5D-5L value set in Hong Kong. **Methods** EQ-5D-5L health states were elicited using a composite time trade-off (C-TTO) and a discrete-choice (DC) experiment. Use of the FM according to participant characteristics and the impact of the FM on the number of inconsistent C-TTO responses were assessed. We employed a main-effects hybrid model that combined data from both elicitation techniques.

Results In total, 1014 individuals completed the survey. The sample was representative of the general Chinese Hong Kong population in terms of sex, educational

attainment, marital status, and most age groups but not for employment status. The use of the FM reduced the number of C-TTO inconsistencies. Participant characteristics differed significantly between those who used and did not use the FM. The model without a constant resulted in logical consistent coefficients and was therefore selected as the model to produce the value set. The predicted EQ-5D-5L values ranged from -0.864 to 1 .

Conclusions The use of an FM to allow participants to exclude C-TTO responses reduced the number of inconsistent responses and improved the quality of the data when estimating an EQ-5D-5L value set in Hong Kong.

Eliza L. Y. Wong and Oliver Rivero-Arias are joint lead authors.

Electronic supplementary material The online version of this article (doi:[10.1007/s40271-017-0278-0](https://doi.org/10.1007/s40271-017-0278-0)) contains supplementary material, which is available to authorized users.

✉ Oliver Rivero-Arias
oliver.rivero@npeu.ox.ac.uk

¹ The JC School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

² The EuroQol Research Foundation, Rotterdam, The Netherlands

³ National Perinatal Epidemiology Unit (NPEU), Nuffield Department of Population Health, University of Oxford, Old Road Campus, Headington, Oxford OX3 7LF, UK

Key Points for Decision Makers

The first series of studies estimating value sets for the five-level EuroQol-five dimensions (EQ-5D-5L) questionnaire in selected countries identified areas for improvement. A feedback module is now included after the composite time trade-off (C-TTO) tasks in a new wave of country-specific EQ-5D-5L value set studies.

This study shows that using a feedback module to allow participants to identify and exclude incorrect C-TTO responses reduces the number of inconsistent responses and improves both the quality of the data and the estimation of an EQ-5D-5L value set.

This study also reports an EQ-5D-5L value set for Hong Kong that can be used to measure the impact of interventions on health-related quality of life in economic evaluations for resource allocation in this jurisdiction.

1 Introduction

The EuroQol-five dimensions (EQ-5D) questionnaire is a generic preference-based patient-reported outcome measure (PROM) used to capture health-related quality of life [1]. It is the most widely used PROM instrument when measuring health benefits in economic evaluations of health technologies worldwide [2]. The original EQ-5D (also known as the EQ-5D-3L) questionnaire included three levels in each dimension (no problems, some problems, and extreme problems) and defined 243 (3^5) health states. The EQ-5D-3L questionnaire has been used for almost two decades and has been shown to be a reliable and valid PROM in areas such as diabetes and cancer [3, 4]. However, evidence about the measurement properties of the EQ-5D-3L in other diseases such as mental health is mixed [5, 6]. Researchers have also indicated that the three-level version of the EQ-5D may not capture minor changes between two different health states in patients with milder conditions [7]. In response to these concerns, a five-level version of the EQ-5D was developed [7]. The five-level version (known as the EQ-5D-5L) included the original five dimensions but increased the levels to no problems, slight problems, moderate problems, severe problems, extreme problems/unable to, defining 3125 (5^5) health states. New valuation exercises were therefore needed to create value sets for this five-level version. Based on experience from previous EQ-5D-3L valuation exercises, the EuroQol Group developed an international valuation protocol. The objectives of the protocol were to improve the comparability of value sets between countries and the identification of cross-cultural differences of health states between jurisdictions [8]. The original version (known as version 1) of the protocol [8] was developed using a large multi-country pilot project assessing the performance of different elicitation methods and different modes of administration of health state valuations [9]. The main novelties of that protocol compared with previous EQ-5D-3L valuation exercises were the use of the composite time trade-off (C-TTO) [10, 11] and a discrete-choice experiment (DCE) as elicitation techniques [12, 13]. Valuation exercises using version 1 identified data quality issues in the C-TTO responses [14]; to correct for this, version 2 of the valuation protocol incorporated a feedback module (FM) [15].

In this FM, participants were presented with the rank ordering of the health states based on their C-TTO valuations and asked to exclude health state valuations they considered did not have the appropriate location in the ranking. The inclusion of the FM in this new version was based on the hypothesis that it would reduce the number of

C-TTO inconsistencies but was not linked to any theoretical model of economic or human behavior [15]. C-TTO exercises are complex, and the sequential nature of the task is subject to challenges that affect comparability [16–18]. It has been shown that the majority of participants completing TTO tasks need interviewer help [19] and, as most individuals have never previously completed a C-TTO exercise, the level of understanding is expected to vary between individuals. Interviewers play an important role in explaining the task to participants, but it is expected that the cognitive ability of the person also contributes to this understanding. A recent qualitative study reported that individuals with inadequate health literacy were more likely to provide inconsistent C-TTO valuations [20]. This can be associated with theoretical models that link cognitive abilities and economic behavior [21]. A line of thinking in experimental economics investigates whether individual heterogeneity in decision making can be explained by the cognitive ability of the subjects [22–24]. A recent special issue in the *Journal of Behavioral and Experimental Economics* included the most recent developments in the area [21]. Research conducted to date has found evidence that individuals with higher cognitive abilities are more likely to undertake optimal strategic actions in games where behavior is not affected by social preferences (e.g., trust) [25–28]. As part of the C-TTO task, the FM is a tool that facilitates participants to select their final responses to the exercise. It is possible that subjects with higher cognitive skills are more likely than those with lower cognitive skills to better understand the task and agree with their final responses without engaging with the FM. However, it is also possible that subjects with higher cognitive skills perceive the FM as a tool to refine their final C-TTO responses, identifying inconsistent responses and engaging more with the tool than individuals with lower cognitive skills.

Version 2 of the protocol was recently employed in a valuation exercise in Hong Kong. We report the results of that valuation exercise and evaluate the impact of using an FM when modelling an EQ-5D-5L value set for Hong Kong. We hypothesize that whether a participant engages with the FM depends on their cognitive ability and background characteristics.

2 Methods

Ethics approval to conduct this study was obtained from the Joint Chinese University of Hong Kong—New Territories East Cluster Clinical Research Ethics Committee (CUHK-NTEC; Approval no: CRE-2013.464).

2.1 The EQ-5D-5L Hong Kong Valuation Exercise

The EuroQol valuation protocol version 2 was followed in this valuation study and included five sections: (1) a general welcome to the participant; (2) introduction to the context of the research and completion of self-reported health using the EQ-5D-5L descriptive system and visual analog scale (VAS), and background information; (3) ten C-TTO tasks and an FM; (4) seven DCE tasks; and (5) a general thank you and goodbye. The implementation of the interview protocol was facilitated using the EuroQol Valuation Technology version 2 (EQ-VT 2.0) software [8].

2.1.1 Sample and Data Collection

The EuroQol valuation protocol recommended the collection of 10,000 C-TTO responses, and—given that each participant valued ten health states—suggested a minimum of 1000 respondents [29]. Sample size calculations were based on obtaining mean C-TTO values with some level of precision; details of these power calculations have been reported elsewhere [8]. Data were collected between June 2014 and October 2015, and the 2016 Hong Kong census was used to evaluate the representativeness of our sample. A representative sample in terms of sex, age, and highest educational attainment was recruited from 18 geographical districts throughout Hong Kong. The survey included Cantonese-speaking Hong Kong residents aged ≥ 18 years. Face-to-face interviews were conducted in community centers in each of the geographical regions included in the study. The study was advertised in the community centers using leaflets and posters that included information about the research being conducted and how to participate in the survey. Participants received an incentive of supermarket coupons to the value of \$HK50 (\$US6.5, £5). Interviews were conducted by a team of six interviewers well-trained in the use of EQ-VT software. All the interviewer's training material, originally prepared in English by the EuroQol Group for previous valuation exercises, was professionally translated to Hong Kong Chinese.

2.1.2 Methods to Elicit Preferences

We used a C-TTO and a DCE to obtain preferences in this study. These techniques have already been applied in several EQ-5D-5L valuation exercises across different countries [29–34].

The C-TTO involved using the conventional TTO for the health states better than dead and lead-time TTO for states worse than dead [35, 36]. The TTO tasks included 86 EQ-5D-5L health states identified through simulation and divided in ten blocks with similar severity levels. All the

blocks were constructed to include one very mild state (one dimension at level 2, the remaining dimensions at level 1), the state 55555, and a set of intermediate health states. Participants were randomized to one of the ten blocks, and the order in which the states were presented was also randomized.

The use of DCEs is a complementary approach to elicit individual preferences [12, 13, 37]. It has been suggested that a DCE is able to measure health preferences using a test that is cognitively easier to understand [38]. A recent multinational study showed the feasibility of using DCEs in the context of EQ-5D-5L valuations [39]. However, the values obtained from DC models are not on the same scale 0 (death)–1 (full health) as its TTO counterpart and need to be re-scaled for any subsequent quality-adjusted life-year (QALY) calculations. In a DCE, respondents are presented with multiple (usually two) health states and asked to indicate which they prefer. Our DCE included 196 pairs of EQ-5D-5L states divided into 28 blocks with similar severity representation using an efficient Bayesian design [40], and each block included seven DC pairs. Participants were randomly allocated to one of the blocks, and the order of the pairs and the position of the health states (i.e., left or right) were also randomized.

2.1.3 Quality Control

EQ-VT 2.0 included a data quality control process to monitor data collection using the EQ-VT QC tool [14]. The quality control process included different measures to monitor how C-TTO and DCE information was collected. These measures were presented in a weekly quality control report that was generated to identify any suspicious or questionable performances by interviewers. Additional information about this quality control process is available upon request.

2.1.4 Feedback Module (FM)

Version 2 of the protocol also included an FM for C-TTO tasks. This FM consisted of presenting each respondent with a ranking of the ten health states ordered by their own C-TTO responses and provided the opportunity to indicate and exclude any state that—in their opinion—did not have the correct ranking position (Fig. 1). No option for re-valuing the selected states by the participant was allowed. We argue here that the use of an FM as part of the C-TTO task can be supported using theoretical models that link cognitive abilities and economic behavior [21]. In particular, we hypothesize that the engagement of the subject with the FM depends on their cognitive skills and background characteristics. To test such an hypothesis, an explicit measure of cognitive skills and a number of

participant characteristics were needed. The protocol used in valuation exercises for the EQ-5D-5L instrument did not include an explicit measure of cognitive skill and included the following background characteristics: age, sex, whether the participant had experience with serious illness, educational level, marital status, and whether the participant was diagnosed with a chronic condition. Educational level was used in this study as a proxy for cognitive skills, as research has shown that cognitive skill is a good predictor of academic achievement [41, 42]. Therefore, in this study, we assumed that participants with higher educational attainment had greater cognitive skills.

We first explored whether the background characteristics of participants selecting and excluding states during the FM differed from those not engaging (not selecting and excluding states). Then we identified the number of inconsistent health states before and after using the tool and the number of states excluded during the FM. This provided information about whether using the FM helped reduce inconsistent C-TTO responses. We defined a logical dominance relationship between two health states as follows: state A dominates state B when state A is better than

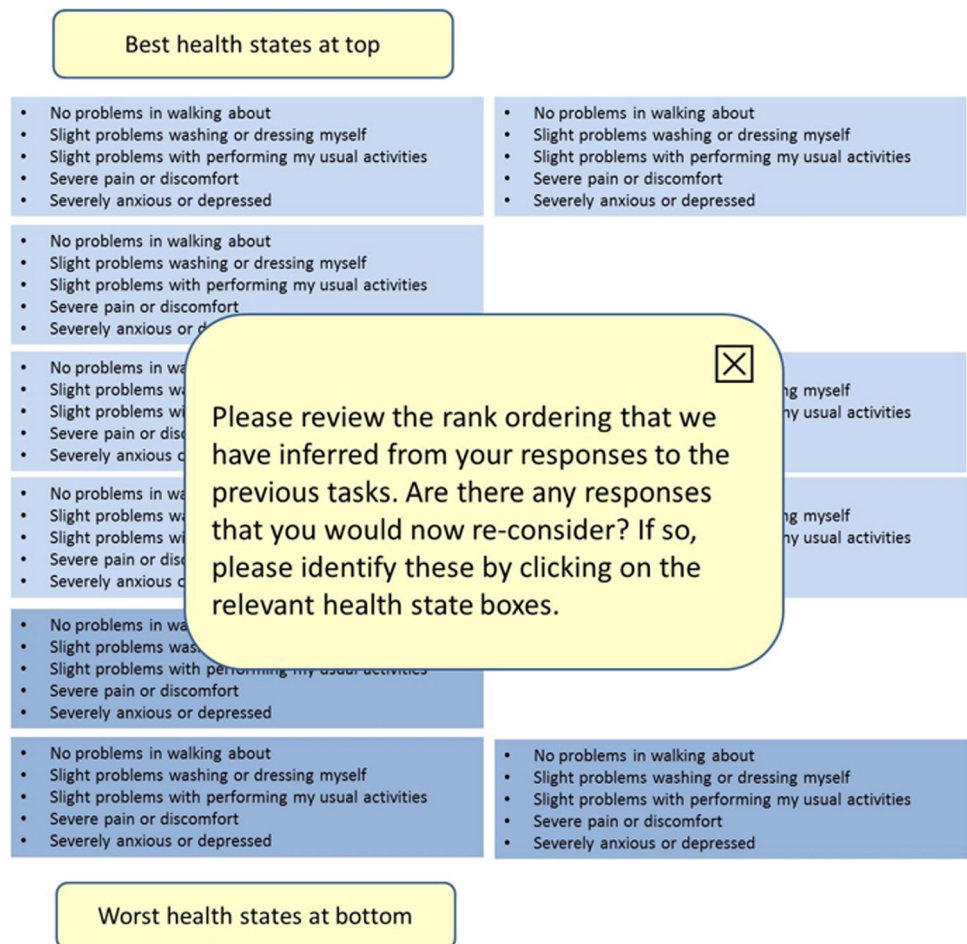
state B on at least one dimension and no worse than state B on any other dimension. Elicited values were considered logically inconsistent if a state was assigned a value indicating it was better than a dominating state.

The impact of the FM in terms of duration of interview (in minutes) for the different elements of the C-TTO tasks was also assessed, and we report timings for the whole study sample and by participants who did and did not exclude states during the FM.

2.2 Statistical Analysis

Background characteristics for the study samples were described using means and standard deviations for continuous variables and frequencies and percentages for categorical covariates. Statistically significant differences at the 5% level in the distribution of socioeconomic characteristics between participants who did and did not exclude states during the FM were evaluated using Pearson’s chi-squared (χ^2) test. The distribution of observed C-TTO utility values were explored across respondents overall and by severity index, defined as the sum of the levels in a

Fig. 1 Screenshot of feedback module of composite time trade-off (C-TTO) responses (reproduced with permission from EuroQol Group)



particular state (e.g., state 22222 had a severity index of 10).

A hybrid model that combined both C-TTO and DCE data was used to estimate the value set in Hong Kong, thereby maximizing the usage of data collected. This modelling strategy has been used successfully in previous EQ-5D-5L valuation exercises [29, 32]. Briefly, this model estimates a set of coefficients from a unique likelihood function obtained by multiplying the likelihood functions of a distribution for the C-TTO data by the likelihood function of a distribution for DC data [29, 43, 44]. A conditional logit was assumed for DC data together with a normal distribution with values censored at -1 (Tobit model) assumed for C-TTO data. We censored C-TTO observations at -1 (lowest possible value in the C-TTO task) given that, theoretically, valuations for worse than dead states were in the range $(-\infty, 0)$. However, the lead-time TTO approach used in this study did not allow the elicitation of values below -1 . Censoring values at -1 when modelling EQ-5D-5L C-TTO data has been implemented in recent valuation studies in the UK and the Netherlands [32, 34] and is considered good practice [45]. The dependent variable of the hybrid model includes both C-TTO and DC responses. The C-TTO responses were defined as 1 minus the C-TTO observed values for a given health state to indicate disutilities and therefore coefficients expressed as utility decrements. The DC responses included in the dependent variable were binary outcomes 0/1 indicating the respondent's choice to each pair of EQ-5D-5L states. The hybrid model used a rescaled parameter θ that assumed that the C-TTO model coefficients were proportional to the DC model coefficients. We used cluster estimation to acknowledge that, for each participant included in the models, ten C-TTO and seven DC responses were available.

The model specification was a hybrid main effects (20-parameter model) consisting of four dummies for each EQ-5D-5L dimension using level 1 as the reference. Dummies were constructed to represent the utility decrement of moving from the reference (level 1) to any of the remaining levels (levels 2, 3, 4, and 5). As an example, we created four dummies (MO2–MO5), indicating the utility decrement of moving from level 1 to level 2 (MO2), moving from level 1 to level 3 (MO3), and so on. The same set of dummy variables was defined for each of the remaining dimensions: self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD).

Previous EQ-5D-5L valuation studies have reported different distributions of C-TTO values between severe and less severe states [29, 46]. In general, people tend to agree more about valuations of less severe states than about more severe ones, resulting in a heteroscedastic error term when modelling C-TTO data. We tested for homoscedasticity of the error term using a separate Tobit model for the C-TTO data as described by Cameron and Trivedi [47]. We

corrected for heteroscedasticity and modelled the variance of the error term as a function of EQ-5D-5L levels for each domain including a constant. We also evaluated the impact of including a constant term in the main effects models as this has received attention in recent work [48]. The decision over whether or not to include a constant in the final value set was based on whether that term was statistically significant.

Performance of the final model for the Hong Kong value set was determined by the logical consistency of parameters and goodness of fit. Coefficients were considered logically consistent if values from logically worse health states were lower than those from logically better health states (all coefficients positive in ascending order in our model outputs). This was reflected by estimated coefficient values in descending order.

We included all DC data in the modelling exercise. However, we excluded C-TTO observations based on criteria 1 (respondents with a positive slope on a regression between his/her C-TTO values and the severity of the states indicating that the participant provided higher utility values for poorer health states on average) and criteria 2 (respondents who valued all states equally except non-traders, i.e., subjects who valued all states as 1).

All analyses were carried out in Stata MP [49]. Hybrid models were estimated in Stata using the `hyreg` command [45].

2.3 Assessing the Impact of the FM on Modelling Results

The final value set for Hong Kong was estimated excluding states selected by participants during the FM. This was a normative approach because there was no reason to keep C-TTO responses participants excluded in the FM. However, we compared this final value set with a value set that did not exclude states identified during the FM. Although such a comparison was meaningless in terms of model selection for a final value set, it provided a useful “what-if” situation for if the FM had not been available. Hence, we estimated models including all states (without FM information) or excluding states (with FM information) and compared the model goodness of fit using the Akaike information criteria divided by the sample size (AIC/n) [50]. Predicted mean values of the 86 health states included in the C-TTO tasks were also compared with and without FM information.

2.4 Comparison with Other Asian Country-Specific EQ-5D-5L Value Sets

We calculated and compared predictions for the 3125 health states using the final Hong Kong EQ-5D-5L value set with predictions from currently published EQ-5D-5L

value sets in China, Indonesia, South Korea, and Japan [33, 51–53]. Predictions were compared using kernel density functions for the 3125 states.

3 Results

3.1 Characteristics of Study Sample

Figure 1 in the Electronic Supplementary Material (ESM) presents a flow diagram of participants who completed the survey. A total of 1033 Hong Kong residents aged ≥ 18 years participated in the survey; 19 dropped out at the beginning of the C-TTO task. Withdrawal reasons included fatigue, the time consumed, and the complexity of the task. As the responses from these 19 respondents were limited, they were excluded from any further analysis, leaving 1014 respondents with complete information included in the study.

Table 1 presents the background characteristics of the study sample in comparison with the Hong Kong general population. The mean age of the sample was 46 years, and 600 (59%) were female. In total, 813 (77%) individuals attended at least secondary school, 583 (58%) were married, and 435 (43%) were in paid employment. Around 70% of participants had experienced serious illness themselves or in a relative or cared for others. Compared with the general population, our sample had fewer respondents aged 45–54 years but slightly more aged 55–64 years, and fewer individuals were in paid employment. Overall, our study sample was reasonably representative of the Hong Kong population in terms of sex, educational attainment, and marital status. The distribution of self-reported EQ-5D-5L descriptive and VAS for the overall sample and different age groups is presented in Table 1 in the ESM. Most participants in the sample (95%) reported either no or slight problems in all domains. The mean \pm standard deviation (SD) of the VAS was estimated at 82.72 ± 11.77 for the overall sample. The mean \pm SD and median total interview time in minutes was estimated at 40 ± 14 and 37, respectively.

3.2 Responses from the Composite Time Trade-Off and Discrete-Choice Experiment Data

Figure 2 in the ESM depicts the distribution of observed C-TTO utility values across all respondents. The histogram showed a concentration of values at 1 (health states equal to full health) and -1 (lowest possible value in the C-TTO task). The spike at -1 indicated that, in 16% of the C-TTO tasks, the lead time was exhausted and the participant indicated that dying now was equivalent to living 10 years in full health followed by 10 years in the impaired state.

This result supported our choice to censor C-TTO values at -1 when modelling the C-TTO data as part of the hybrid model. The distribution of observed C-TTO utility values by severity index is presented in Fig. 3 in the ESM. The series of histograms shows that less severe states enjoyed less variability across participants than did more severe health states, where disagreement across respondents was greater. This translated into a problem of heteroscedasticity of the error term that was corroborated by rejecting the null hypothesis of equal variances in the test of homoscedasticity (critical value = 4561 with $p = 0.000$). No DC data were excluded from any modelling analyses, but C-TTO responses from 15 interviews were excluded from any subsequent analysis after applying the exclusion criteria (Fig. 1 in the ESM).

3.3 Quality Control

There was no need to interrupt the study or drop data as all interviewers met the minimum quality criteria. Additional details about this quality control are available upon request.

3.4 FM

Table 1 reports the background characteristics of participants who excluded [$n = 340$ (34%)] or did not exclude ($n = 674$ (65%)] states during the FM. There were statistically significant differences between the two groups in terms of age ($\chi^2 = 19.3$; $p < 0.01$), highest educational attainment ($\chi^2 = 11.5$; $p < 0.01$), personal experience with serious illness ($\chi^2 = 12.1$; $p < 0.01$), and whether the participant lived with a diagnosed chronic condition ($\chi^2 = 10.9$; $p < 0.05$). Participants who selected health states during the FM were younger, had post-secondary or degree qualifications, did not have personal experience with serious illness, and did not live with a chronic condition.

A total of 9990 states (after exclusion criteria) were used in the C-TTO section of the survey across all participants, and 515 (5%) C-TTO states were selected and excluded during the FM. Hence, data from the remaining 9475 states were included in the estimation of the value set. Of the 340 participants who excluded at least one state during the FM and had at least one inconsistent state before using the tool, 285 (84%) removed the inconsistency or inconsistencies (Table 2). Conversely, 140 (21%) of the 674 participants who did not exclude any state during the FM had at least one inconsistency that was retained in the estimation sample. Table 2 also shows that, in most cases, the number of states excluded during the FM was one or two.

Although participants who excluded states during the FM spent an additional 2 minutes in that section at the end of the C-TTO task, the overall length of the C-TTO section

Table 1 Background characteristics of study samples compared with the Hong Kong general population^a

	Hong Kong population ^b (%)	Study sample (n = 1014)	Participants not excluding states during FM (n = 674)	Participants excluding states during FM (n = 340)
Female gender	55.0	600 (59.0)	405 (60.1)	195 (57.4)
Age group (years)				
15/18–24 ^c	12.1	166 (16.4)	101 (15.0)	65 (19.1)
25–34	16.7	173 (17.1)	100 (14.8)	73 (21.8)
35–44	17.5	173 (17.1)	108 (16.0)	65 (19.1)
45–54	18.6	119 (11.7)	86 (12.8)	33 (9.7)
55–64	17.2	223 (22.0)	159 (23.6)	64 (18.8)
≥65	17.9	160 (15.8)	120 (17.8)	40 (11.5)
Highest education attainment				
Primary and below	20.0	201 (19.8)	150 (22.3)	51 (15.0)
Secondary/sub-degree	57.8	615 (60.7)	408 (60.5)	207 (60.9)
Post-secondary/degree	22.2	198 (19.5)	116 (17.2)	82 (24.1)
Marital status				
Single	30.1	322 (31.8)	197 (29.2)	125 (36.8)
Married	58.4	583 (57.5)	402 (59.6)	181 (53.2)
Divorced/separated	5.1	52 (5.1)	33 (4.9)	19 (5.6)
Widow	6.4	57 (5.6)	42 (6.2)	15 (4.4)
Employment status ^d				
Full-time student	16.9	109 (10.7)	65 (9.6)	44 (12.9)
Retired/homemaker/unemployed	31.8	470 (46.3)	328 (48.7)	142 (41.8)
Employer/full-time/part-time	51.3	435 (42.9)	281 (41.7)	154 (45.3)
Experience with serious illness				
Personal	–	262 (25.8)	197 (29.2)	65 (19.1)
Relatives	–	397 (39.2)	253 (37.5)	144 (42.4)
Caring for others	–	501 (49.4)	325 (48.2)	176 (51.8)
Diagnosed with chronic condition				
None	–	707 (69.7)	450 (66.8)	257 (75.6)
1	–	194 (19.1)	137 (20.3)	57 (16.8)
2	–	80 (7.9)	59 (8.8)	21 (6.2)
≥3	–	33 (3.3)	28 (4.2)	5 (1.5)

FM feedback module

^aData are presented as % or n (%)

^b2016 population by-census – summary results [65]

^cData for Hong Kong Census refer to population aged ≥15 years and for study sample aged ≥ 18 years

^d2011 population census—main report [66]

was similar between those who did and did not exclude states during the FM (Table 2 in the ESM).

3.5 The EQ-5D-5L Value Set in Hong Kong

The last two columns of Table 3 show that utility decrements for each EQ-5D-5L level associated with each potential value set had the correct sign and magnitude (e.g., was logically consistent). A non-significant constant was obtained when the term was introduced and was excluded from the final value set. Consequently, the final selected model from which to derive utility decrements in Hong

Kong is reported in the last column of Table 3, and Table 3 in the ESM reports how to use this value set in practice. The lowest possible estimated value in that model for the health state 55555 was estimated at -0.8637 .

3.6 Assessing the Impact of the FM on Modelling Results

The coefficients of estimating a hybrid model without the FM are also reported in Table 3. The hybrid model without constant and with the FM has a slightly better goodness of fit, estimated at 1.504 compared with the hybrid model

without constant and without the FM: goodness of fit estimated at 1.543. Hence, a marginal benefit in terms of goodness of fit was seen for the hybrid model with the FM, but predictions between the two models were similar (Fig. 4 in the ESM).

3.7 Comparison with Other Asian Country-Specific EQ-5D-5L Value Sets

Figure 2 shows the kernel density function of the 3125 predicted health states in Hong Kong and other published Asian countries. The value set for Indonesia provided predictions that were remarkably similar to those of the Hong Kong value set. A larger proportion of health states considered worse than death (negative values) was observed in the Hong Kong and Indonesia value sets (36%) compared with China (10%) and Japan and South Korea (0.1%).

4 Discussion

In this study, we estimated an EQ-5D-5L value set using a sample of the Hong Kong general Chinese population that excluded health states using an FM at the end of the C-TTO tasks. The sample was representative of the general Hong Kong population in terms of sex, educational attainment,

marital status, and most age groups but not for employment status. Overall, the proportion of C-TTO states excluded during the FM was low (5%). Limited qualitative information collected using free-text at the end of the FM for 71 participants suggested they found the tool useful to identify logically inconsistent states, but they also expressed difficulty understanding the task given the lack of reality of some states. Excluding states identified during the FM did not have a major impact in the model coefficients and predictions compared with not excluding states, but it was associated with slightly better goodness of fit. In addition, the FM reduced the number of inconsistent C-TTO responses used in the final estimation sample.

The use of a feedback process to give participants the opportunity to reflect and think about their preferences in valuation exercises is not novel in the health preference research literature. Existing studies have demonstrated that preferences can vary after a reflection or deliberation exercise [54, 55]. Therefore, it can be seen that the FM used in this study was a process for individuals to reflect about their own C-TTO responses. In our study, we also provided a theoretical rationale for the use of the FM from a behavioral economics viewpoint and hypothesized that cognitive skills and participant characteristics explained how individuals interacted with the tool. Participants who excluded states during the FM differed from those who did not: They were younger, had post-secondary or degree

Table 2 Number of participants using the feedback module and impact on the number of inconsistent health states

Number of inconsistent health states without FM	Number of states excluded during FM	Number of inconsistent health states with FM			
		None	1	2	>2
None	None	534	NA	NA	NA
	1	121	NA	NA	NA
	2	58	NA	NA	NA
	>2	14	NA	NA	NA
1	None	NA	100	NA	NA
	1	38	20	NA	NA
	2	7	5	NA	NA
	>2	7	0	NA	NA
2	None	NA	NA	21	NA
	1	7	8	6	NA
	2	11	1	0	NA
	>2	4	1	0	NA
>2	None	NA	NA	NA	19
	1	3	3	3	1
	2	6	5	0	0
	>2	9	2	0	0

Figures in bold indicate the number of participants who did not exclude health states during the FM ($n = 674$)

FM feedback module, NA not applicable

Table 3 Hybrid main effects model results without and with feedback module

	Without feedback module				With feedback module			
	Tobit: c-TTO responses cLogit: DC responses		Tobit: c-TTO responses cLogit: DC responses (no constant)		Tobit: c-TTO responses cLogit: DC responses		Tobit: c-TTO responses cLogit: DC responses (no constant)	
	Coefficient (SE)	<i>p</i> value	Coefficient (SE)	<i>p</i> value	Coefficient (SE)	<i>p</i> value	Coefficient (SE)	<i>p</i> value
MO2	0.111 (0.009)	0.000	0.111 (0.009)	0.000	0.110 (0.01)	0.000	0.109 (0.01)	0.000
MO3	0.177 (0.012)	0.000	0.177 (0.012)	0.000	0.183 (0.013)	0.000	0.182 (0.013)	0.000
MO4	0.359 (0.015)	0.000	0.359 (0.015)	0.000	0.372 (0.016)	0.000	0.371 (0.016)	0.000
MO5	0.505 (0.018)	0.000	0.505 (0.018)	0.000	0.529 (0.019)	0.000	0.529 (0.019)	0.000
SC2	0.088 (0.009)	0.000	0.089 (0.009)	0.000	0.087 (0.009)	0.000	0.087 (0.009)	0.000
SC3	0.113 (0.012)	0.000	0.113 (0.012)	0.000	0.114 (0.012)	0.000	0.113 (0.012)	0.000
SC4	0.268 (0.015)	0.000	0.268 (0.015)	0.000	0.272 (0.016)	0.000	0.271 (0.016)	0.000
SC5	0.338 (0.016)	0.000	0.338 (0.016)	0.000	0.352 (0.017)	0.000	0.352 (0.017)	0.000
UA2	0.073 (0.009)	0.000	0.074 (0.008)	0.000	0.068 (0.008)	0.000	0.067 (0.008)	0.000
UA3	0.098 (0.012)	0.000	0.098 (0.012)	0.000	0.095 (0.013)	0.000	0.094 (0.012)	0.000
UA4	0.230 (0.012)	0.000	0.230 (0.012)	0.000	0.234 (0.013)	0.000	0.234 (0.013)	0.000
UA5	0.277 (0.013)	0.000	0.277 (0.013)	0.000	0.282 (0.014)	0.000	0.282 (0.014)	0.000
PD2	0.081 (0.008)	0.000	0.081 (0.008)	0.000	0.076 (0.008)	0.000	0.076 (0.008)	0.000
PD3	0.150 (0.012)	0.000	0.150 (0.012)	0.000	0.148 (0.013)	0.000	0.147 (0.013)	0.000
PD4	0.299 (0.015)	0.000	0.299 (0.015)	0.000	0.308 (0.015)	0.000	0.307 (0.015)	0.000
PD5	0.343 (0.016)	0.000	0.343 (0.016)	0.000	0.355 (0.017)	0.000	0.354 (0.017)	0.000
AD2	0.085 (0.009)	0.000	0.086 (0.009)	0.000	0.081 (0.009)	0.000	0.080 (0.009)	0.000
AD3	0.143 (0.013)	0.000	0.144 (0.013)	0.000	0.140 (0.013)	0.000	0.140 (0.013)	0.000
AD4	0.290 (0.015)	0.000	0.290 (0.015)	0.000	0.294 (0.015)	0.000	0.293 (0.015)	0.000
AD5	0.340 (0.016)	0.000	0.340 (0.016)	0.000	0.348 (0.016)	0.000	0.348 (0.016)	0.000
Const.	0.001 (0.005)	0.826	–	–	–0.002 (0.005)	0.739	–	–
nC-TTO	8407		8407		7943		7943	
n _{censored} C-TTO	1583		1583		1532		1532	
nDC	7098		7098		7098		7098	
AIC/ <i>n</i>	1.543		1.543		1.504		1.504	
U(21111)	0.8879		0.8886		0.8919		0.8906	
U(12111)	0.9107		0.9114		0.9144		0.9133	
U(11211)	0.9257		0.9264		0.9339		0.9328	
U(11121)	0.9183		0.9189		0.9254		0.9244	
U(11112)	0.9139		0.9145		0.9209		0.9199	
U(55555)	–0.8024		–0.8028		–0.8642		–0.8637	

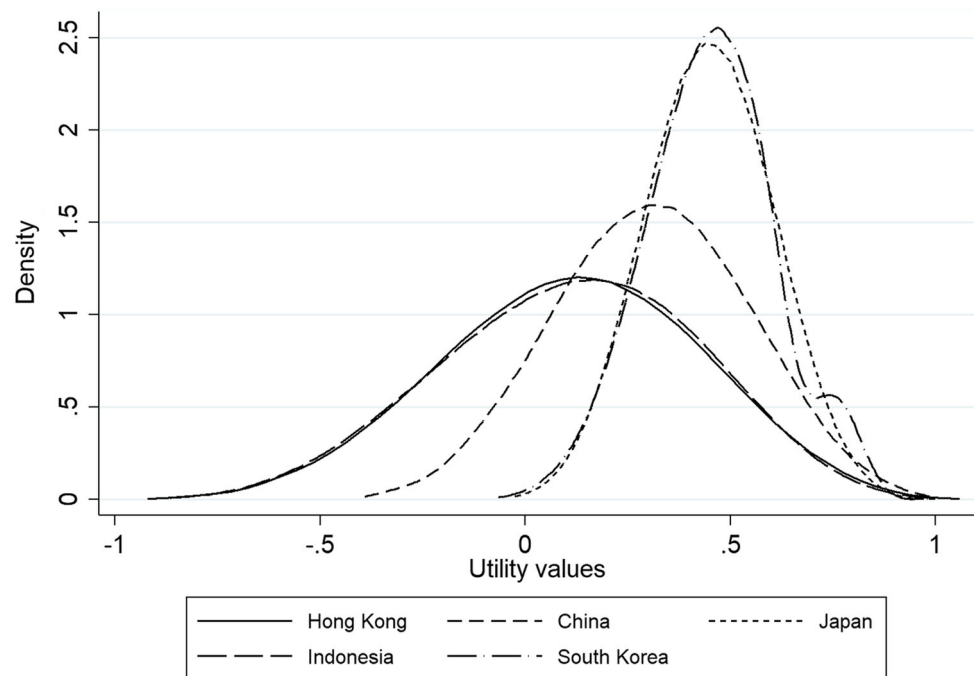
AD anxiety/depression, *AIC* Akaike information criteria, *cLogit* multinomial conditional logit, *C-TTO* composite time trade-off, *DC* discrete choice, *MO* mobility, *nC-TTO* number of uncensored C-TTO observations, *n_{censored}C-TTO* number of censored C-TTO observations, *nDC* number of discrete choice observations, *PD* pain/discomfort, *SC* self-care, *SE* standard error, *U()* utility associated with health state, *UA* usual activities

qualifications, did not have personal experience with serious illness, and did not live with a chronic condition. Therefore, our results seem to suggest that those with higher cognitive skills were more likely to use the FM. This systematic difference is important, and future EQ-5D-5L valuation exercises should bear this in mind. In particular, our results should be incorporated in the training package for interviewers using the EQ-VT. A note of

caution: Our measure of cognitive skills was based on the proxy variable of highest educational attainment and not an explicit measure, and this should be seen as a limitation of our analysis.

A hybrid main effects model using both C-TTO and DC data without the constant coefficient was selected as the final value set for Hong Kong. The use of additional covariates to inform the model using either first-order

Fig. 2 Kernel density function of predicted values for all 3125 health states for EQ-5D-5L value sets in Asian countries



interactions or other related terms has been widely practiced in EQ-5D-3L valuation exercises [56]. The inclusion of interactions in EQ-5D valuation studies was normally informed by previous evidence of significant interactions in other valuation studies and/or imposed by the researchers [57–60]. However, a recent simulation study has warned that, to allow identification and estimation of interactions between dimensions and levels, the design of the study should include main effects and the expected interactions [61]. The authors of the work concluded that most EQ-5D-3L valuation exercises conducted to date lack sufficient coverage of the EQ-5D space to allow estimation of interactions. The experimental design implemented in the current EQ-5D-5L valuation exercise was based on main effects only for both the C-TTO and the DC tasks [8]. Consequently, modelling of health states for Hong Kong was based on a main effects specification without interactions or additional terms. An important aspect of the inclusion of interaction effects in any econometric model is largely ignored in applied work: The identification of interaction effects to test in an econometric model should occur before the data are collected and not after to minimize the role of chance in the findings. This has recently been debated by researchers describing the pros and cons of the use of analysis plans in econometric analyses [62, 63].

This study was conducted at the same time as other EQ-5D-5L valuation exercises were being or had been conducted in different Asian countries. We compared predictions of the Hong Kong value sets with those from published EQ-5D-5L value sets in China, Japan, South

Korea, and Indonesia. Health preferences of the Cantonese population in Hong Kong could be similar to the general population in these countries. However, Hong Kong has a population epidemiology and healthcare system that differs from those in these other settings [64]. Besides, although Hong Kong is a special administrative region of China, its cultural diversity suggests views on health between the two jurisdictions are likely to differ. We found that predictions between the Hong Kong and Indonesian value sets were very similar but those from the remaining countries differed. Hong Kong and Indonesia employed the same valuation protocol, which included an FM, and modelling technique to derive their value sets (Table 4 in the ESM). Future research should investigate whether the differences observed with the other Asian countries were due to the use of a different protocol version, modelling methods, or cultural factors.

5 Conclusion

An FM for C-TTO responses was a useful tool that improved the quality of the data used when estimating an EQ-5D-5L value set in Hong Kong. The final value set can now be used to measure the impact of interventions on health-related quality of life in economic evaluations for resource allocation in this jurisdiction.

Acknowledgements The authors are grateful to the EQ-VT support team for facilitating and supporting the EQ-VT software and monitoring the quality control during the data collection. We thank the district officers from 18 districts throughout Hong Kong for their help

with recruitment. We are also grateful to Mr. Dicken Chan, Ms. Nicole Huang, Ms. Yeung Yeung Pan, Mr. Ringo Sze, and Mr. Sky Chan for their assistance during data collection. Finally, we are grateful to all the Hong Kong residents who participated in the survey. Without their participation and engagement, the study would not have succeeded.

Author contributions ELYW secured funding for the study, planned and led the project, wrote the manuscript jointly with ORA, and revised and approved the final version for submission. JMRG supervised data collection, conducted the statistical analysis, and revised and approved the final manuscript for submission. AWLC and AYKW monitored and supervised the data collection, clarified data-related issues, and revised and approved the final manuscript for submission. ORA planned and led the project, supervised data collection, conducted the statistical analysis, wrote the manuscript with ELYW, and revised and approved the final version for submission.

Compliance with Ethical Standards

Funding This study was supported by the Health and Medical Research Fund from the Food and Health Bureau of Hong Kong, Hong Kong Special Administrative Region (SAR) (Grant number: HMRF11120491). The study sponsors had no role in the analysis and interpretation of data, writing of the manuscript, or decision to submit the article for publication.

Conflict of interest Oliver Rivero-Arias and Juan Manuel Ramos-Goñi are members of the EuroQol Group. Eliza LY Wong, Annie WL Cheung, and Amy YK Wong have no potential conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. EuroQol G. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16:199–208.
2. Kind P, Brooks RG, Rabin R. EQ-5D concepts and methods: a developmental history. Dordrecht: Springer; 2005.
3. Janssen MF, Lubetkin EI, Sekhobo JP, Pickard AS. The use of the EQ-5D preference-based health status measure in adults with Type 2 diabetes mellitus. *Diabet Med*. 2011;28(4):395–413. doi:10.1111/j.1464-5491.2010.03136.x.
4. Pickard AS, Wilke CT, Lin H-W, Lloyd A. Health utilities using the EQ-5D in studies of cancer. *PharmacoEconomics*. 2012;25(5):365–84. doi:10.2165/00019053-200725050-00002.
5. Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technol Assess*. 2014. doi:10.3310/hta18340.
6. Mulhern B, Mukuria C, Barkham M, Knapp M, Byford S, Soeteman D, et al. Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D. *Br J Psychiatry*. 2014;205(3):236–43. doi:10.1192/bjp.bp.112.122283.
7. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–36. doi:10.1007/s11136-011-9903-x.
8. Oppe M, Devlin NJ, van Hout B, Krabbe PFM, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health*. 2014;17(4):445–53. doi:10.1016/j.jval.2014.04.002.
9. Devlin N, Krabbe P. The development of new research methods for the valuation of EQ-5D-5L. *Eur J Health Econ*. 2013;14(Issue 1 Suppl):1–3.
10. Robinson A, Spencer A. Exploring challenges to TTO utilities: valuing states worse than dead. *Health Econ*. 2006;15(4):393–402. doi:10.1002/hec.1069.
11. Augustovski F, Rey-Ares L, Irazola V, Oppe M, Devlin NJ. Lead versus lag-time trade-off variants: does it make any difference? *Eur J Health Econ*. 2013;14(Suppl 1):S25–31. doi:10.1007/s10198-013-0505-0.
12. Salomon J. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. *Popul Health Metr*. 2003;1(1):12.
13. McCabe C, Brazier J, Gilks P, Tsuchiya A, Roberts J, O'Hagan A, et al. Using rank data to estimate health state utility models. *J Health Econ*. 2006;25(3):418–31. doi:10.1016/j.jhealeco.2005.07.008.
14. Ramos-Goñi JM, Oppe M, Slaap B, Busschbach JJV, Stolk E. Quality control process for EQ-5D-5L valuation studies. *Value Health*. 2017. doi:10.1016/j.jval.2016.10.012.
15. Ludwig K, von der Schulenburg J-MG, Greiner W. Valuation of the EQ-5D-5L with composite time trade-off for the German population—an exploratory study. *Health Qual Life Outcomes*. 2017;15(1):39. doi:10.1186/s12955-017-0617-9.
16. Lenert LA, Cher DJ, Goldstein MK, Bergen MR, Garber A. The effect of search procedures on utility elicitation. *Med Decis Mak*. 1998;18(1):76–83. doi:10.1177/0272989X9801800115.
17. Arnesen T, Trommald M. Are QALYs based on time trade-off comparable? A systematic review of TTO methodologies. *Health Econ*. 2005;14(1):39–53. doi:10.1002/hec.895.
18. Augestad LA, Stavem K, Kristiansen IS, Samuelsen CH, Rand-Hendriksen K. Influenced from the start: anchoring bias in time trade-off valuations. *Qual Life Res*. 2016;25:2179–91. doi:10.1007/s11136-016-1266-x.
19. Edelaar-Peeters Y, Stiggelbout AM, Van Den Hout WB. Qualitative and quantitative analysis of interviewer help answering the time tradeoff. *Med Decis Mak*. 2014;34(5):655–65. doi:10.1177/0272989X14524989.
20. Al Sayah F, Johnson JA, Ohinmaa A, Xie F, Bansback N, Canadian EQDLVSG. Health literacy and logical inconsistencies in valuations of hypothetical health states: results from the Canadian EQ-5D-5L valuation study. *Qual Life Res*. 2017;26(6):1483–92. doi:10.1007/s11136-016-1495-z.
21. Brañas-Garza P, Smith J. Cognitive abilities and economic behavior. *J Behav Exp Econ*. 2016;64:1–4. doi:10.1016/j.socec.2016.06.005.
22. Agranov M, Caplin A, Tergiman C. Naive play and the process of choice in guessing games. *J Econ Sci Assoc*. 2015;1(2):146–57. doi:10.1007/s40881-015-0003-5.
23. Agranov M, Potamites E, Schotter A, Tergiman C. Beliefs and endogenous cognitive levels: an experimental study. *Games Econ Behav*. 2012;75(2):449–63. doi:10.1016/j.geb.2012.02.002.
24. Schulz JF, Fischbacher U, Thöni C, Utikal V. Affect and fairness: dictator games under cognitive load. *J Econ Psychol*. 2014;41:77–87. doi:10.1016/j.joep.2012.08.007.

25. Benito-Ostolaza JM, Hernández P, Sanchis-Llopis JA. Do individuals with higher cognitive ability play more strategically? *J Behav Exp Econ.* 2016;64:5–11. doi:10.1016/j.socec.2016.01.005.
26. Kiss HJ, Rodriguez-Lara I, Rosa-García A. Think twice before running! Bank runs and cognitive abilities. *J Behav Exp Econ.* 2016;64:12–9. doi:10.1016/j.socec.2015.01.006.
27. Corgnet B, Espín AM, Hernán-González R, Kujal P, Rassenti S. To trust, or not to trust: cognitive reflection in trust games. *J Behav Exp Econ.* 2016;64:20–7. doi:10.1016/j.socec.2015.09.008.
28. Lohse J. Smart or selfish—when smart guys finish nice. *J Behav Exp Econ.* 2016;64:28–40. doi:10.1016/j.socec.2016.04.002.
29. Ramos-Goni JM, Pinto-Prades JL, Oppe M, Cabases JM, Serano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Med Care.* 2017;55(7):e51–8. doi:10.1097/MLR.000000000000283.
30. Augustovski F, Rey-Ares L, Irazola V, Garay OU, Gianneo O, Fernandez G, et al. An EQ-5D-5L value set based on Uruguayan population preferences. *Qual Life Res.* 2015;. doi:10.1007/s11136-015-1086-4.
31. Xie F, Pullenayegum E, Gaebel K, Bansback N, Bryan S, Ohinmaa A, et al. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Med Care.* 2016;54(1):98–105. doi:10.1097/MLR.000000000000447.
32. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Econ.* 2017. doi:10.1002/hec.3564.
33. Kim SH, Ahn J, Ock M, Shin S, Park J, Luo N, et al. The EQ-5D-5L valuation study in Korea. *Qual Life Res.* 2016;. doi:10.1007/s11136-015-1205-2.
34. Versteegh MM, Vermeulen KM, Evers SMAA, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. *Value Health.* 2016;19(4):297–510. doi:10.1016/j.jval.2016.01.003.
35. Devlin N, Buckingham K, Shah K, Tsuchiya A, Tilling C, Wilkinson G, et al. A comparison of alternative variants of the lead and lag time TTO. *Health Econ.* 2013;22(5):517–32. doi:10.1002/hec.2819.
36. Devlin NJ, Tsuchiya A, Buckingham K, Tilling C. A uniform time trade off method for states better and worse than dead: feasibility study of the ‘lead time’ approach. *Health Econ.* 2011;20(3):348–61. doi:10.1002/hec.1596.
37. Ryan M, Gerard K, Amaya-Amaya M. Using discrete choice experiments to value health and health care. Dordrecht: Springer; 2008.
38. Stolk EA, Oppe M, Scalone L, Krabbe PF. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. *Value Health.* 2010;13(8):1005–13. doi:10.1111/j.1524-4733.2010.00783.x.
39. Krabbe PF, Devlin NJ, Stolk EA, Shah KK, Oppe M, van Hout B, et al. Multinational evidence of the applicability and robustness of discrete choice modeling for deriving EQ-5D-5L health-state values. *Med Care.* 2014;52(11):935–43. doi:10.1097/MLR.000000000000178.
40. Hensher DA, Rose JM, Greene WH. Applied choice analysis. 2nd ed. Cambridge: Cambridge University Press; 2015.
41. Alloway TP, Passolunghi MC. The relationship between working memory, IQ, and mathematical skills in children. *Learn Individ Differ.* 2011;21(1):133–7. doi:10.1016/j.lindif.2010.09.013.
42. Alloway TP, Alloway RG. Investigating the predictive roles of working memory and IQ in academic attainment. *J Exp Child Psychol.* 2010;106(1):20–9. doi:10.1016/j.jecp.2009.11.003.
43. Oppe M, van Hout B, editors. The optimal hybrid: experimental design and modeling of a combination of TTO and DCE. Euro-Qol Group Proceedings, 2013.
44. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Med Decis Mak.* 2015;35(3):328–40. doi:10.1177/0272989X14559542.
45. Ramos-Goñi J, Craig AM, Oppe M, Van Hout B. Combining continuous and dichotomous responses in a hybrid model. Rotterdam: EuroQol Research Foundation; 2016.
46. Feng Y, Devlin NJ, Shah KK, Mulhern B, van Hout B. New methods for modelling EQ-5D-5L value sets: an application to english data. *Health Econ.* 2017. doi:10.1002/hec.3560.
47. Cameron AC, Trivedi PK. Microeconometrics using Stata. College Station: Stata Press; 2009.
48. Shah K, Mulhern B, Longworth L, Janssen MF. An empirical study of two alternative comparators for use in time trade-off studies. *Value Health.* 2016;19(1):53–9. doi:10.1016/j.jval.2015.10.012.
49. StataCorp. Stata statistical software release 14. College Station, TX: StataCorp LP2014.
50. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control.* 1974;19(6):716–23.
51. Luo N, Liu G, Li M, Guan H, Jin X, Rand-Hendriksen K. Estimating an EQ-5D-5L value set for China. *Value Health.* 2017;20(4):662–9. doi:10.1016/j.jval.2016.11.016.
52. Purba FD, Hunfeld JAM, Iskandaryah A, Fitriana TS, Sadarjoen SS, Ramos-Goni JM, et al. The Indonesian EQ-5D-5L value set. *Pharmacoeconomics.* doi:10.1007/s40273-017-0538-9 (Epub 10 July 2017).
53. Shiroywa T, Ikeda S, Noto S, Igarashi A, Fukuda T, Saito S, et al. Comparison of value set based on DCE and/or TTO data: scoring for EQ-5D-5L health states in Japan. *Value Health.* 2016;19(5):648–54. doi:10.1016/j.jval.2016.03.1834.
54. Karimi M, Brazier J, Paisley S. The effect of reflection and deliberation on health state values. Sheffield: University of Sheffield; 2016.
55. Robinson S, Bryan S. Does the process of deliberation change individuals’ health state valuations? An exploratory study using the person trade-off technique. *Value Health.* 2013;16(5):806–13. doi:10.1016/j.jval.2013.03.1633.
56. Xie F, Gaebel K, Perampaladas K, Doble B, Pullenayegum E. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. *Med Decis Mak.* 2014;34(1):8–20. doi:10.1177/0272989X13480852.
57. Dolan P. Modeling valuations for EuroQol health states. *Med Care.* 1997;35(11):1095–108.
58. Johnson JA, Luo N, Shaw JW, Kind P, Coons SJ. Valuations of EQ-5D health states: are the United States and United Kingdom different? *Med Care.* 2005;43(3):221–8.
59. Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, et al. Estimating an EQ-5D population value set: the case of Japan. *Health Econ.* 2002;11(4):341–53. doi:10.1002/hec.673.
60. Bansback N, Tsuchiya A, Brazier J, Anis A. Canadian valuation of EQ-5D health states: preliminary value set and considerations for future valuation studies. *PloS One.* 2012;7(2):e31115. doi:10.1371/journal.pone.0031115.
61. Viney R, Norman R, King MT, Cronin P, Street DJ, Knox S, et al. Time trade-off derived EQ-5D weights for Australia. *Value Health.* 2011;14(6):928–36. doi:10.1016/j.jval.2011.04.009.
62. Olken BA. Promises and perils of pre-analysis plans. *J Econ Perspect.* 2015;29(3):61–80. doi:10.1257/jep.29.3.61.
63. Coffman LC, Niederle M. Pre-analysis plans have limited upside, especially where replications are feasible. *J Econ Perspect.* 2015;29(3):81–97. doi:10.1257/jep.29.3.81.
64. Kong X, Yang Y, Gao J, Guan J, Liu Y, Wang R, et al. Overview of the health care system in Hong Kong and its referential significance to mainland China. *J Chin Med Assoc.* 2015;78(10):569–73. doi:10.1016/j.jcma.2015.02.006.

-
65. Census and Statistics Department. 2016 Population by-census: Summary results. Hong Kong: Census and Statistics Department; 2017. p. 1–164. <http://www.statistics.gov.hk/pub/B11200942016XXXXB0100.pdf>.
66. Census and Statistics Department. 2011 Population census main report: Volume I. Hong Kong: Census and Statistics Department; 2012. p. 1–309. <http://www.statistics.gov.hk/pub/B11200592012XXXXB0100.pdf>.