VALUE IN HEALTH ▮ (2017) ▮▮▮–▮▮▮

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Original Research

Evaluation of Split Version and Feedback Module on the Improvement of Time Trade-Off Data

Eliza L.Y. Wong, MPH, PhD[1,*], *Koonal Shah, MSc*[2], *Annie W.L. Cheung, MPhil*[1], *Amy Y.K. Wong, MSc*[1], *Martijn Visser, MSc*[3], *Elly Stolk, PhD*[4,5]

[1]*The Jockey Club School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China;* [2]*Office of Health Economics, London, UK;* [3]*Medical Psychology and Psychotherapy, Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands;* [4]*Institute of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands;* [5]*EuroQol Executive Office, Rotterdam, The Netherlands*

ABSTRACT

Background: EQ-5D-5L valuation studies previously reported many inconsistent responses in time trade-off (TTO) data. A number of possible elements, including ordering effects of the valuation tasks, mistakes at the sorting question, and interviewers' (learning) effects, may contribute to their inconsistency. **Objectives:** This study aimed to evaluate the effect of two modifications on consistency of TTO data in The Netherlands (NL) and Hong Kong (HK): (1) separating the valuation of the Better than Dead (BTD) and Worse than Dead (WTD) states; and (2) Implementation of feedback (FB) module by offering an opportunity to review TTO responses. **Methods:** A cross-over design with two study arms was used to test the effect of the modifications. In each jurisdiction, six interviewers were involved where half the interviewers started using the standard version, and the other half started with the split version. Each version was switched after every 25 (NL) or 30 (HK) interviews until 400 interviews were completed. **Results:** In the NL and HK, 404 and 403 respondents participated, respectively. With the use of the FB module, the proportion of respondents with inconsistent responses was lowered from 17.8% to 10.6% ($P < 0.001$) in NL and from 31.8% to 22.3% ($P = 0.003$) in HK. The result of separating the valuation of BTD and WTD states was not straightforward because it reduced the inconsistency rate in NL but not in HK. **Conclusions:** The results support implementation of the FB module to promote the consistency of the data. The separation of the BTD and WTD task is not supported.

Keywords: EQ-5D-5L, health-related quality of life, health preference, composite time trade-off, feedback module, utility measurement, split version, Hong Kong, The Netherlands.

A standard protocol and interviewer training materials developed by the EuroQol Group can be used to create value sets for the EQ-5D-5L questionnaire uniformly across different jurisdictions with the aid of a computer-assisted personal interview software —EuroQol Valuation Technology (EQ-VT) [1,2]. For the valuation, some consider (or at least can imagine) certain health problems to be worse than being dead. However, the valuation of WTD states is controversial, and a number of competing methods for obtaining WTD values have been proposed [3–6]. Composite time trade-off (cTTO), which has been presented as one of the primary health state valuation approaches, uses the 10 years lead-time TTO with a ratio of 1:1 [2,7]. Prior to the initiation of cTTO, there was no "theoretical" lower boundary on the utilities for states WTD. If the person considers the health states equivalent to "full health," the utility value is 1, and for states "equal to dead," the value is 0. For states WTD, this implies that the utilities could approach $-\infty$ [8]. By transforming utilities for health states WTD and to compare the effect of bounding the negative values, the choice for –1 was motivated by an equal range for positive and negative values [3]. To assess the cognitive burden of BTD/WTD, a number of debriefing questions were included at the end of 10 cTTO tasks to evaluate whether the instructions were clear to respondents [7]. Scales of the debriefing statements, together with the average number of steps in the iterative process, were analyzed, and the results confirmed the validity of WTD as a measurement and the feasibility of cTTO.

Unfortunately, the first series of EQ-5D-5L valuation studies have reported several common quality issues in their cTTO data. One of these issues is a large number of inconsistent responses observed [9–13]. In the descriptive system of EQ-5D-5L, a *health state* is defined by taking one level from each of the five different dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). Each health state is described by a five-digit single index. For example, the index for having "no problems" in all five dimensions of EQ-5D-5L would be "11111," and the index for having "extreme problems in all dimensions would be

* ***Address correspondence to:*** Eliza L.Y. Wong, School of Public Health, Prince of Wales Hospital, Shatin, Hong Kong, China.
E-mail: lywong@cuhk.edu.hk

1098-3015$36.00 – see front matter

https://doi.org/10.1016/j.jval.2017.10.013

"55555." Logically, all health states "dominate" the worst state (index 55555). A response is defined as inconsistent if a health state that is better in at least one dimension and no worse in all others, receiving a lower value than the state it logically dominates. However, at least 20% of respondents valued one or more health states as being worse illogically in each of those valuation studies [9–13]. Similarly, about 10% of respondents gave lower values to very mild health states than they did to more severe and logically worse health states. Moreover, it was recognized that about half (8%) of the inconsistencies with regard to the worst state involved large utility differences (>0.5) [14]. It has been considered that these observed problems might (at least partially) reflect implementation issues: ordering effects of the valuation tasks of the state considered as BTD or WTD, mistakes at the "sorting question" identifying whether a state was considered BTD or WTD, interviewees' strategic behavior, and interviewers' learning effects. Although real-time data monitoring using quality check software (QC tool), three practice cTTO tasks in addition to one standard example the state of being "in a wheelchair," and confirmatory pop-ups have been included in the standard protocol, the issue of inconsistency of TTO data still exists. Because this can be resolved, at least partially, an update of the protocol for EQ-5D-5L valuation studies is warranted.

Perfect consistency may not be expected because the differences between EQ-5D-5L health states are subtle and respondents are likely to be uncertain about the values that they provide [15]. However, the proportion of respondents whose responses contained severe inconsistencies has been considered unacceptably large and is unlikely to be the product of random error or uncertainty *per se*. As a result, the individual level cTTO data were closely scrutinized. Here, it was observed that severe inconsistencies could often be traced to inconsistent behavior at the sorting question, guiding respondents to value of the states considered BTD or WTD in the cTTO task [7]. It leads to the hypothesis that the order of the valuation of states considered BTD/WTD would influence the response. In addition, the BTD/WTD sorting question may possibly be "error prone" if respondents mistakenly give a –1 response while intending to give a 0 response, or vice versa. It may just simply be inherently difficult to answer, hence it invokes inconsistent responses, which, in turn, quickly result in severe inconsistencies because the distribution of responses on the BTD and WTD parts of the scale are different between the two comparable health states (full health vs BTD/WTD states). It further supports another hypothesis that making mistakes at the sorting question for BTD/WTD valuation may have contributed to the rate of inconsistency.

This study investigated the effect of two modifications on the implementation: (1) separating the valuation of the BTD and WTD states by moving all WTD valuations to the end of the cTTO task, and (2) implementation of a feedback (FB) module by offering respondents the opportunity to review their responses and take the wrong ones out, if any. Both modifications were tested on the consistency of TTO data in NL and HK.

Methods

Study Design

The differences between the standard version of protocol [2] and a modified version of cTTO task (hereafter referred to as "split version") were explored, with a crossover study design in NL and HK. In each jurisdiction, a stable team of six interviewers was recruited, and three interviewers were randomly assigned to two groups to conduct interviews. Block sampling (25 and 30 interviews for NL and HK, respectively) was adopted to minimize the selection bias of respondents in the two groups. Group A started with the standard version (control arm) for the first 25 (NL) or 30 (HK) interviews and then switched to the split version, which involved separating the BTD and WTD tasks within the cTTO exercise (experimental arm). Group B started with the split version of cTTO exercise and then switched to the standard version, in the same manner as group A. The two study arms were switched again after completing further 25 and 30 interviews until they had both reached the study's target sample of 400. Randomizing individual respondents to either the standard version or the split version was considered an alternative design. However, it was not opted for because of possible spillover effects between the methods that could occur if interviewers muddled up the protocols by switching between the two methods simultaneously. After collecting TTO values by using the standard version or the split version of the cTTO task, the FB module was offered to all respondents, giving them the opportunity to indicate changes in both the control arm and the experimental arm. The study design is shown in Figure 1.

Control arm

In the control arm, respondents completed the health state valuation tasks by using the standard protocol [1,2], including multiple training and quality control components. The training task aimed to make sure that the respondents understood the concept of TTO. The interviewer first showed how TTO works using as an example the state of being "in a wheelchair" and three practice tasks where the respondent was asked to value health states of varying severities. For monitoring the protocol compliance and the performance of interviewers, QC tool was implemented. The valuation task was completed in a single sitting, regardless of whether the respondent indicated that the health state under evaluation was BTD or WTD.

Experimental arm

The experimental arm differed from the control arm in that the valuation task of WTD states was postponed until after all of the states had been checked and all of the BTD states had been evaluated. Separation of the valuation of the BTD and WTD states implied that the respondents initially indicated that a health state was WTD in the cTTO task while postponing the assessment (lead time TTO) of WTD states after completion of all of the health states by using conventional cTTO. When a respondent considered a health state WTD, a pop-up would appear, asking the respondent to confirm that immediate death was preferred over having to live in that health state for 10 years and indicating that follow-up questions (lead time TTO task) would be presented later. The interview would be continued to evaluate health states considered BTD. After completion of all the valuation for the BTD states, the lead time TTO task was again explained by using the standard example—the state of being "in a wheelchair"; and the states considered WTD were evaluated by using the lead time TTO task. The respondents could reconsider their initial classification of states as WTD—that is, they would rather live for 20 years in total (10 years in full health followed by 10 years in the impaired state) than live only 10 years in full health. This response brought them back to the valuation of BTD tasks.

Feedback module

The FB module was presented after completing the cTTO task in both study arms. FB presented the implied rank ordering of health states derived from the respondent's cTTO responses. All 10 health states were presented as vignettes (five bullet points, one for each dimension of health), consistent with the way the health states had been presented in the preceding cTTO exercise. They were shown on one screen sorted in order of obtained TTO values, with the states given the highest value presented on top and the state given the lowest value at the bottom. Health states