VALUE IN HEALTH 22 (2019) 370–376

Available online at www.sciencedirect.com

ScienceDirect

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

Preference-Based Assessments

Valuation of Health States Considered to Be Worse Than Death—An Analysis of Composite Time Trade-Off Data From 5 EQ-5D-5L Valuation Studies

Mihir Gandhi, MSc [1,2,3,*], Kim Rand, PhD [4,5], Nan Luo, PhD [6]

[1]Department of Biostatistics, Singapore Clinical Research Institute, Singapore; [2]Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore; [3]Tampere Center for Child Health Research, University of Tampere and Tampere University Hospital, Tampere, Finland; [4]Health Services Research Centre, Akershus University Hospital, Lørenskog, Norway; [5]Department of Health Management and Health Economics, University of Oslo, Oslo, Norway; [6]Saw Swee Hock School of Public Health, National University of Singapore, Singapore

ABSTRACT

Objectives: To evaluate the discriminative ability of negative values measured in 5-level EuroQol 5-dimensional questionnaire (EQ-5D-5L) value set studies. **Methods:** This is a secondary analysis of EQ-5D-5L value set studies from Singapore, the Netherlands, China, Thailand, and Canada in which health state values were elicited from a general population sample using a composite time trade-off (TTO) method. Mean values were calculated for health states with same severity. The association between the mean values and severity was evaluated using Pearson correlation (*r*). A linear mixed model using severity as the fixed effect was fitted for values. The analyses were performed separately for positive values (from a conventional TTO for health states considered "better than death") and negative values (from a lead time TTO for health states considered "worse than death"). **Results:** In Singapore (N = 1000; negative values 32.6%), the mean decreased with severity from 0.89 to 0.21 for positive values and increased with severity from −0.98 to −0.89 for negative values. The correlation between values and severity was much lower for negative values ($r = -0.016$) than for positive values ($r = -0.614$). The coefficient of severity in the linear mixed model for negative values was much smaller (coefficient = −0.009; pseudo-R^2 < 0.001) compared with the model for positive values (coefficient = −0.041; pseudo-R^2 = 0.337). Results using data sets from the other countries were similar. **Conclusions:** Negative values are not associated with severity of health states in EQ-5D-5L valuation studies, suggesting poor discriminative ability of the lead time TTO method in valuing health states considered worse than death.

Keywords: EQ-5D, TTO, utility, valuation, worse than death

Introduction

When asked to value hypothetical health states using the time trade-off (TTO) methods, such as in a valuation study for a health utility instrument, respondents frequently indicate that 1 or more of the presented states are worse than death (referred to as "WTD health states" hereafter). With the conventional utility scale in which 1 indicates a preference equal to full health and 0 indicates a preference equal to that of immediate death, WTD health states are represented as negative values. Using currently available valuation methods, utility values of WTD health states are more difficult to estimate than those considered better than death (referred to as "BTD health states" hereafter).[1] First, WTD health states are rarely experienced and typically less familiar than BTD health states to respondents and are therefore more difficult to imagine and contextualize by laypersons. Second, and more important, valuation tasks for WTD health states are more complex and less transparent than those for BTD health states. Taking the TTO procedure adopted by the widely used Measurement and Valuation of Health protocol[2] as an example, the WTD task involves asking the respondent to imagine living in a WTD health state for x years ($x < 10$) followed by full health for $10 - x$ years,

* *Address correspondence to*: Mihir Gandhi, MSc, Department of Biostatistics, Singapore Clinical Research Institute, #02-01, Nanos, 31 Biopolis Way, Singapore 138669, Singapore.
E-mail: mihir.gandhi@scri.edu.sg

https://doi.org/10.1016/j.jval.2018.10.002

followed by death. This composite life is compared with immediate death, and the length of time in the WTD health state is varied until indifference is reached. Furthermore, negative values derived from the TTO procedure are typically rescaled to the range of −1 to 0 using a mathematical function, which was criticized for lack of validity.[3,4] A study looking at data from the UK and US 3-level EuroQol 5-dimensional questionnaire (EQ-5D) valuation studies observed that the mean of negative values was unrelated to state severity.[5]

Recently, a new TTO variant called *composite TTO* (cTTO) was developed and recommended for valuation of 5-level EQ-5D (EQ-5D-5L) health states.[6] For BTD health states, it is identical to the Measurement and Valuation of Health protocol for valuation of BTD 3-level EQ-5D health states (hereafter referred to as "conventional TTO"). For WTD health states, it uses lead time TTO. Originally proposed by Robinson and Spencer,[7] lead time TTO adds a period of full health (lead time such as 10 years) to both presented alternative lives. The lead time TTO was found to generate very low values for BTD EQ-5D-5L health states,[8] and the procedure was adopted only for WTD health states in the EQ-5D-5L valuation protocol.[9] Nevertheless, the performance of using the lead time TTO method as implemented in the EQ-5D-5L valuation studies is not formally assessed, especially its discriminative ability to differentiate health states of varying severity.

The objective of this study was to critically scrutinize negative values using the lead time TTO collected in EQ-5D-5L valuation studies from Asian (Singapore, China, and Thailand) as well as Western countries (the Netherlands and Canada)[10,11] and to assess sensitivity to the severity of the health states under investigation, both within individuals and for the study population as a whole.

Methods

This study drew data from the EQ-5D-5L valuation studies in Singapore, the Netherlands, China, Thailand, and Canada. All studies followed the data collection protocol developed by the EuroQol Group to interview 1000 or more members of the local public face-to-face in a cross-sectional survey. All interviews were conducted by trained interviewers using the EuroQol Valuation Technology (EQ-VT) computer program running from a laptop[8] and by following standard interviewer scripts developed in local languages.

The Valuation Interview

Each respondent was asked to value 10 EQ-5D-5L health states using cTTO. Health states considered BTD and WTD were valued using conventional TTO and lead time TTO, respectively. A detailed description of the cTTO-based valuation task can be found elsewhere.[8,9] Briefly, the task aims to identify the point of preferential indifference for BTD health states between 2 alternatives: 10 years of life in the health state for valuation, followed by death, and a shorter life ($x \leq 10$ years) in full health, also followed by death. With a defined utility value of 1 for full health, the utility value of a BTD health state was calculated as $x/10$. For states considered to be WTD, a lead time of 10 years was added to the 10 years of life in the health state to be valued to elicit a negative utility value. The utility value of a WTD health state was calculated as $(x - 10)/10$, which was bounded at −1 and 0. In each interview, the interviewer used state "in a wheelchair" to explain the concept of cTTO to participants before proceeding to the 10 formal valuation tasks. The EQ-VT program was designed to value selected EQ-5D-5L health states also using discrete choice experiment (data not used in the present study).

EQ-5D-5L Health States

The EQ-VT–based valuation studies included 86 EQ-5D-5L health states for valuation using cTTO, including the 5 most mildly impaired health states (ie, 21111, 12111, 11211, 11121, and 11112), state 55555, and 80 other states of varying severity. The 86 health states were grouped into 10 blocks, all of which contained 1 mildest health state, state 55555, and 8 block-unique health states. Each participant was randomly assigned 1 block for cTTO valuation. As usual, the 5-digit numbers represent EQ-5D-5L health states, with each digit representing the level of problems for 1 dimension in order of mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Thus, 11111 represents no problems on any dimension, whereas 55555 represents extreme problems on all 5 dimensions.

Statistical Analyses

The final data sets used to estimate EQ-5D-5L value sets for Singapore (number of respondents [N] = 1000), China (N = 1298), Thailand (N = 1205), the Netherlands (N = 989), and Canada (N = 1073) were used. For illustration purpose, the key results for 1 Asian country (Singapore) and 1 Western country (the Netherlands) are presented in the "Results" section. The results of the other countries are presented in the Appendix in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2018.10.002.

Each valuation data set was partitioned into 2 data sets, one containing all the positive values (including the "0" value) reflecting values of health states considered BTD and valued using the conventional TTO method and the other containing all the negative values reflecting values of health states considered WTD and valued using the lead time TTO method, for performing the following hypothesis-testing analysis separately.

First, we hypothesized that for both the BTD and WTD groups, health state values should be negatively correlated with overall severity of the health states as measured by the number of deviations from full health. To test this hypothesis, we calculated the mean utility values for health states with the same severity distance for each participant. Severity distance was calculated as the difference between the sum of severity levels (1 = no problem, 2 = slight problem, 3 = moderate problem, 4 = severe problem, and 5 = extreme problem) for each of the 5 dimensions of a health state and full health (11111), that is, a value between 0 and 20. The Pearson product-moment correlation coefficient (r) between mean health state value and severity distance was estimated using individual-level data. This hypothesis was also assessed visually at the level of individual study participants by calculating the mean utility values by severity distance for each individual health state and plotting a line chart to observe whether utility values tended to decrease with increasing overall severity. Furthermore, we fitted a linear mixed model (LMM) for utility values with the severity distance as the fixed effect and participant-specific intercepts as random effects to assess the association between utility values and severity, controlling for intraparticipant correlation among utility values (eg, some participants tend to give very low or high values for all assigned health states).

Second, we investigated the extent to which negative utility values were distinguishable from random noise. The null hypothesis was that there was no association between health states and assigned values, conditional on the respondent considering the state in question to be WTD. To statistically test this hypothesis, we compared the proportion of variance in negative utility values explained by health states to variance explained when permuting observed health states over the assigned values. This was done in 4 steps:

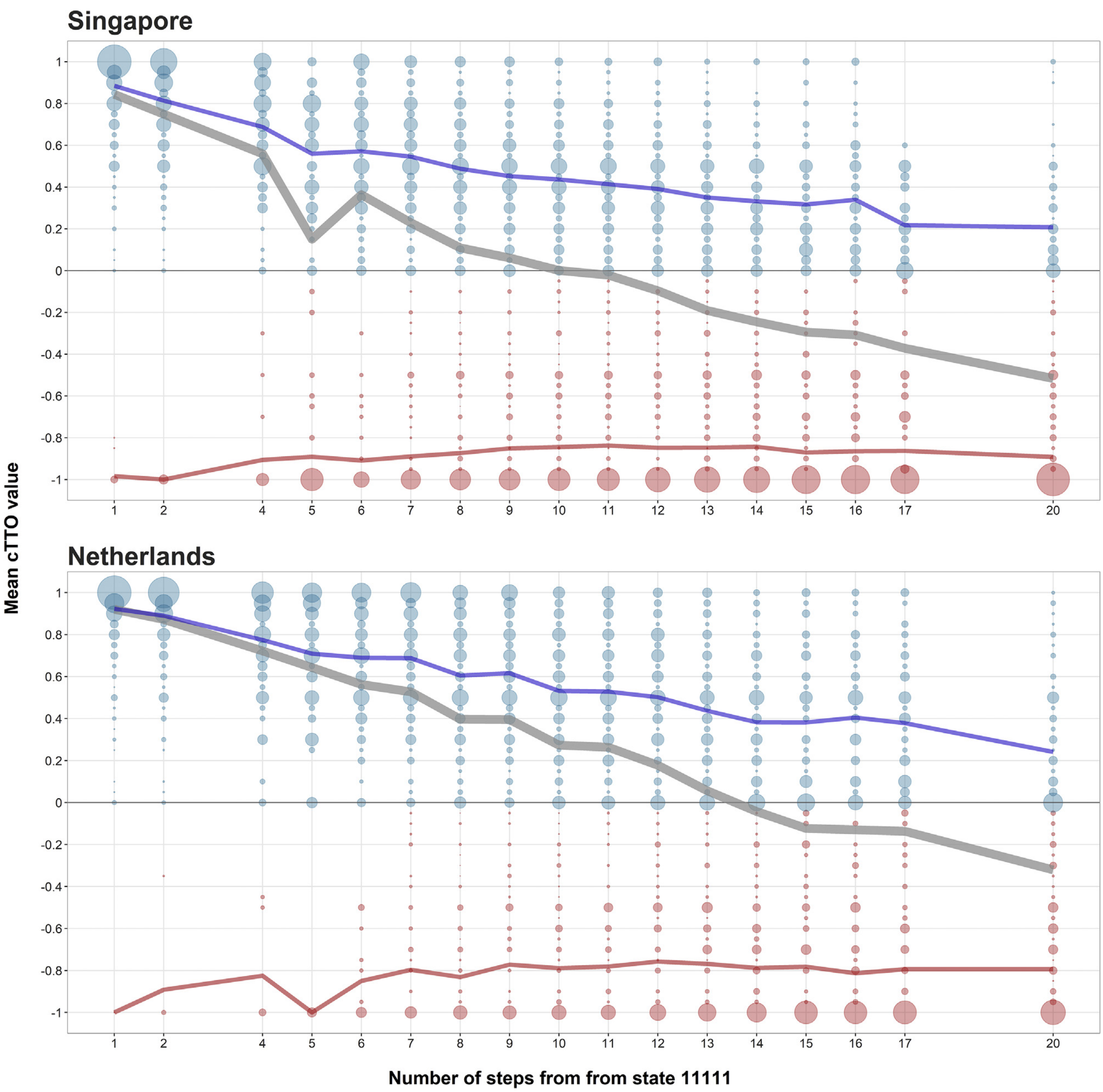

Fig. 1 – Mean utility values over severity distance in Singapore and the Netherlands. Severity distance on the *x*-axis is the difference between the sum of severity levels for each of the 5 dimensions of given health state and full health (11111). Circle area equals the proportion of responses to health states of severity distance *x* given at a specified value. The thick gray line is the grand mean, the thin blue line is the mean of positive utility values, and the thin red line is the mean of negative utility values. cTTO indicates composite time trade-off.

1. We fitted a simple linear model to the observed negative values. If *i* is an index over health states (eg, 11211 or 55555), the model had the following form:

$$\text{tto} \sim \beta_i * \text{state}_i + \varepsilon$$

   With this model, there is a single parameter per health state, taking on a value identical to the observed mean for that health state. We then calculated the analysis of variance (ANOVA) sum of squares for the model.
2. We created a pool of 10 000 permutations over the health state index (ie, 10 000 separate samples of the same size), each without replacement, while keeping the assigned values fixed. This reflects a substantial pool of counterfactual observations in which the assigned values were assigned with no regard to the health state. Importantly, the joint distribution of health states and assigned values is retained.
3. For each permutation, we repeated step 1, producing an empirical distribution of ANOVA sum of squares, allowing

Table 1 – Summary of the LMM for utility values on severity distance from the full health in Singapore and the Netherlands.

Variables	Singapore, Coefficient (95% CI)		The Netherlands, Coefficient (95% CI)	
	Based on positive utility values	Based on negative utility values	Based on positive utility values	Based on negative utility values
Severity distance	−0.041 (−0.041 to −0.040)	−0.009 (−0.010 to −0.008)	−0.038 (−0.039 to −0.037)	−0.007 (−0.009 to −0.005)
Intercept	0.861 (0.847 to 0.875)	−0.692 (−0.716 to −0.669)	0.938 (0.924 to 0.953)	−0.639 (−0.673 to −0.606)
Pseudo-R^2	0.377	<0.001	0.306	<0.001

Note. Severity distance (range 0-20) is the difference between the sum of severity levels for each of the 5 dimensions of given health state and full health (11111).
CI indicates confidence interval; LMM, linear mixed model.

calculation of a mean and standard error for this statistic under the null hypothesis.
4. We calculated 2-tailed *P* values for the sum of squares found in step 1.

This procedure gives the probability that the variance in mean values across health states is as large or larger than the observed value, under the null hypothesis of no relationship between values and health states. Note that the procedure does not assume any particular functional form for the relationship between health states and mean values. For graphical representation, we calculated all percentiles and 95% confidence intervals (CIs) for each health state, and superimposed the observed mean values on the permutation-based percentiles and the 95% CIs.

For comparison purposes, the same procedure was used for positive utility values, with the expectation that these would display a clear relationship between health states and mean values, such that positive values cannot be construed as being random, that is, indicating that the positive values display the desired systematic (and valid) information.

All the analyses were carried out using Stata/MP 13.1 (StataCorp, College Station, TX) or R 3.2.3 for Windows (R Foundation for Statistical Computing, Vienna, Austria).

Results

A total of 3260 (33%), 2114 (21%), 3061 (24%), 2076 (17%), and 1136 (9%) negative values were given to 86, 80, 83, 77, and 81 health states in Singapore, the Netherlands, China, Thailand, and Canada, respectively. The number of "−1" value was for 2220 (22%), 1094 (11%), 812 (6%), 366 (3%), and 451 (4%) participants in Singapore, the Netherlands, China, Thailand, and Canada, respectively; among negative values, 93% in Singapore, 89% in the Netherlands, 73% in China, 74% in Thailand, and 89% in Canada were in the range of −0.5 to −1. In contrast, the number of "1" value ranged from 377 (3%) in China to 2458 (20%) in Canada; 58% in Singapore, 69% in the Netherlands, 65% in China, 68% in Thailand, and 66% in Canada were in the range of 0.5 to 1 among positive values. The 55555 health state was valued WTD by 658 (66%), 531 (54%), 828 (64%), 691 (57%), and 270 (22%) participants in Singapore, the Netherlands, China, Thailand, and Canada, respectively. Appendix Figure 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2018.10.002 shows the histogram of utility values of directly valued health states in the 5 countries.

Figure 1 shows means of positive and negative utility values over severity distance for Singapore and the Netherlands. In Singapore, the mean of positive values decreased from 0.89 to 0.21 with increase in severity distance (*r* [positive] = −0.614). In contrast, the mean of negative values increased from −0.98 to −0.89 with increase in severity distance (*r* [negative] = −0.016). Similar trends were observed in the Netherlands (*r* [positive] = −0.553; *r* [negative] = 0.015), China (*r* [positive] = −0.580; *r* [negative] = −0.058), Thailand (*r* [positive] = −0.743; *r* [negative] = −0.182), and Canada (*r* [positive] = −0.519; *r* [negative] = 0.047) (see Appendix Figure 2 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2018.10.002). Furthermore, fewer participants gave positive values for health states with larger severity distance, indicating participants' tendency to assign lower utility value as health state severity increases. This is in alignment with the decreasing trend of mean utility value with increase in severity. In contrast, although the proportion of participants rises for negative values with increase in severity, there was no decreasing trend observed for mean utility value with increase in severity. Appendix Figures 3 and 4 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2018.10.002 show how means of positive and negative utility values changed with severity distance at the individual health state level for all the 5 countries. It gives the impression of no systematic pattern in negative values, whereas positive values seem to have a decreasing trend with increase in severity distance. This indicates suboptimal face validity of negative utility values.

Table 1 presents the results of the LMM for positive and negative values in Singapore and the Netherlands. In Singapore, the coefficient for severity distance in the model for negative values was much smaller (coefficient = −0.009; pseudo-$R^2 < 0.001$)

Table 2 – Summary of ANOVA sum of squares test of association between health state and mean values in Singapore and the Netherlands.

Country	Based on positive utility values			Based on negative utility values		
	Mean (SE)	Observed	P	Mean (SE)	Observed	P
Singapore	1899.78 (1.29)	2171.7	<.001	2439.86 (0.72)	2439.5	.656
The Netherlands	2858.43 (1.45)	3159.2	<.001	1317.98 (0.88)	1317.8	.794

ANOVA indicates analysis of variance; SE, standard error.

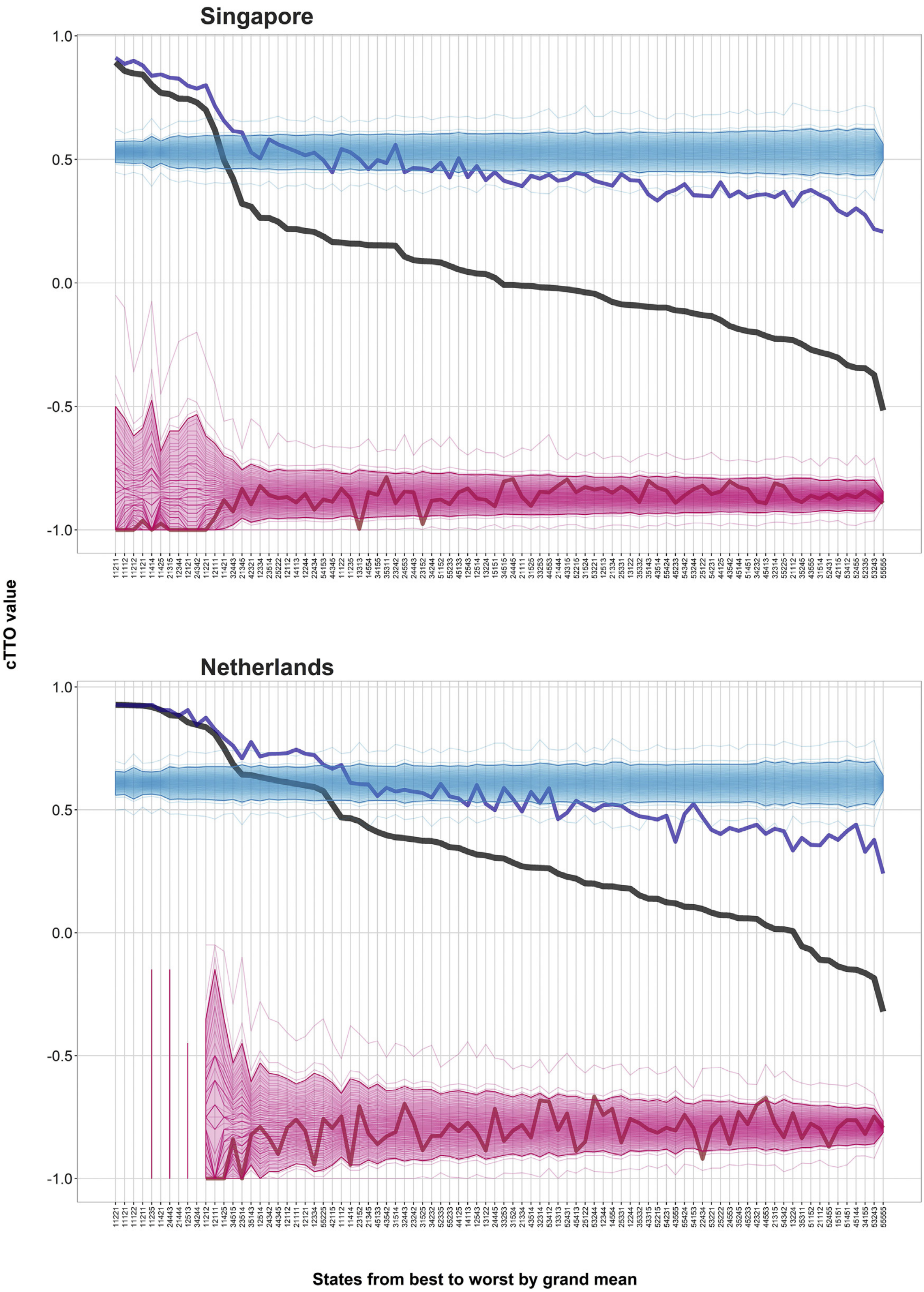

Fig. 2 – Observed means and 95% permutation-based CIs based on the empirical distribution of positive and negative utility values from Singapore and the Netherlands. The thick blue line is the observed mean of positive utility values, and the thick red line is the observed mean of negative utility values. The black line is the observed combined mean of positive and negative utility values. The blue and red regions show 95% permutation-based CIs for positive and negative utility values derived from an empirical distribution of positive and negative utility values, respectively. The thin blue and red lines are 0 to 100 percentiles for each state from the 10 000 permutation-based means for positive and negative utility values, respectively. CI indicates confidence interval; cTTO, composite time trade-off.

compared with that in the model for positive values (coefficient = −0.041; pseudo-R^2 = 0.337). Furthermore, the range of predicted negative values was also narrower (−0.87 to −0.70) compared with that of positive values (0.05 to 0.82). This shows trivial impact of health state severity on negative utility values, whereas positive values were better associated with health state severity, even after removing participant-specific effects. Similar results were observed in the data for the Netherlands and the other countries (see Appendix Table 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2018.10.002).

Table 2 presents the results of the permutation-based ANOVAs to test the likelihood of the observed health state means for positive and negative values under the assumption that health state was not a factor. For both Singapore and the Netherlands, positive values were found to vary substantially by health state, with ANOVA sum of squares several hundred SDs away from those produced in the permutation test. For negative values, however, both were found to be well within the realm of plausibility under the null hypothesis. Similar results were observed for China and Canada, whereas negative values from Thailand were found to be related to health state.

Figure 2 illustrates the overall mean utility value and corresponding 95% CI based on the 10 000 permutations, along with observed mean value of all positive and negative utility values for Singapore and the Netherlands. Similarly, results across the countries show that lack of discriminative ability in negative values is not country-specific (see Appendix Figure 5 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2018.10.002).

Discussion

In this study, we aimed to assess the discriminative ability of utility values derived from the lead time TTO. Our results lent little support to the discriminative ability of negative values elicited using the lead time TTO in 3 Asian and 2 Western EQ-5D-5L valuation studies. First, a high proportion of negative utility values were clustered in the range of −0.5 to −1. Clustering in a specific range may affect the discriminative ability of negative values. Second, there was little or no reduction in negative utility with increase in severity distance, suggesting poor face validity of lead time TTO utility values. This is in contrast to the result for positive utility values from the conventional TTO method. Furthermore, the observed negative values were indistinguishable from random samples from the total pool of negative values (except in Thailand), indicating that they could be construed as noise. From a purely information theoretical perspective, this indicates that there is no practical gain in the collection and analysis of negative values; they appear to be uninformative as to the relative severity of health states. These results suggest suboptimal discriminative ability of negative values, at both individual (the Pearson product-moment correlations) and group levels (the LMM analysis), using the lead time TTO method in EQ-5D-5L valuation studies.

Our findings of poor association between negative lead time TTO values and severity of health states are consistent with the findings from a study[5] of conventional TTO values derived from the UK and US populations. The study showed that the proportion of respondents considering health states to be BTD was strongly related to health state severity, but that values expressed conditional on a health state being valued as WTD were unrelated to health state severity. Together, these findings suggest that issues related to valuation of health states considered to be WTD remain unresolved with the introduction of lead time TTO, and that these issues may not be unique to a particular country or population. There are several possible reasons behind these observations. First, the elicitation methods for health states considered WTD used in conventional TTO and lead time TTO may be too complex or nontransparent, meaning that respondents have preferences for these health states, but are unable to express them properly. Similarly, the transition from BTD to WTD valuation could be confusing, making the task more difficult for respondents.[12] The clustering of negative values between the range of −1 and −0.5 may be related to this confusion. In general, if only part of the scale is used, the face validity of the measurement is not optimal. Also, the use of 10 years as the lead time may not be sufficient because there was a significant floor effect in Singapore and the Netherlands. Second, when identifying a health state as being WTD, we could be moving into a range of “badness” for which specific values have reduced importance to respondents; for an imperfect analogy, it might be worse to be lost at sea in deep waters than in a pond, but not in any way that truly matters. Third, respondents might have ordinal preferences for health states considered WTD (both states A and B are WTD, but state A is clearly worse than B) without being able to express these in terms of cardinal values. If this were the case, the TTO is not suited for WTD valuation.

Depending on what explanation is correct, there are several options available to future researchers. One option would be to extrapolate into negative utility values using models based on positive utility values, for instance, using Tobit regression[13] left-censored at 0. Second, the design of lead time TTO could have made the valuation of WTD states difficult. If this is the case, research on new, easier valuation procedure is needed to improve the design of the current valuation protocol for the EQ-5D-5L.

Although our study suggests that measurement of negative utility values needs improvement, we certainly need to keep the question on whether a health state is BTD or WTD from the participant’s perspective. Nevertheless, the apparent lack of association between health state severity and negative values calls for a discussion as to whether it is useful to ask participants to value a health state if it is considered WTD; the procedure is time-consuming, increases task difficulty for participants, and yields negative values of questionable quality. Although we do not suggest dropping the lead time TTO elicitation for WTD health states at present, we would applaud efforts aimed at improving the valuation of health states considered WTD.

Finally, we acknowledge that the analyses were conducted on existing data sets, which were not designed to collect respondents’ feedback on the rationales behind their valuation responses. Qualitative analyses might help understand respondents’ thought process in valuing health states considered WTD.

Conclusions

Negative utility values are poorly associated with severity of health states, suggesting suboptimal discriminative ability of the lead time TTO method in valuing health states that are considered WTD. Further research is needed to understand respondents’ thought process when valuing WTD health states. Considering the practical costs involved in WTD valuation using the current procedures, juxtaposed with the lack of indication of the usefulness of the resulting values, development of less complex TTO procedures for WTD valuation and modeling without negative values, and studies designed to determine the value of WTD valuation may be warranted.

Acknowledgments

We thank Elly Stolk, Feng Xie, and Junata Pattanaphesaj for providing EQ-5D-5L value set data for the Netherlands, Canada, and Thailand, respectively. We also appreciate Paul Krabbe for helpful comments on earlier versions of this article.

Source of financial support: The authors have no other financial relationships to disclose.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2018.10.002.

REFERENCES

1. Tilling C, Devlin N, Tsuchiya A, et al. Protocols for time tradeoff valuations of health states worse than dead: a literature review. Med Decis Making 2010;30(5):610–9.
2. Gudex C. Time trade-off user manual: props and self-completion methods. In: Gudex C, editor. Occasional Paper Series. York, UK: Centre for Health Economics, University of York; 1994. https://www.york.ac.uk/media/che/documents/papers/occasionalpapers/CHE%20Occasional%20Paper%2020.pdf.
3. Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. Med Care 2007;45(3):238–44.
4. Augestad LA, Rand-Hendriksen K, Kristiansen IS, et al. Impact of transformation of negative values and regression models on differences between the UK and US EQ-5D time trade-off value sets. Pharmacoeconomics 2012;30(12):1203–14.
5. Rand-Hendriksen K, Augestad LA, Dahl FA, et al. A shortcut to mean-based time tradeoff tariffs for the EQ-5D? Med Decis Making 2012;32(4):569–77.
6. Janssen BM, Oppe M, Versteegh MM, et al. Introducing the composite time trade-off: a test of feasibility and face validity. Eur J Health Econ 2013;14(suppl 1):5–13.
7. Robinson A, Spencer A. Exploring challenges to TTO utilities: valuing states worse than dead. Health Econ 2006;15(4):393–402.
8. Oppe M, Devlin NJ, van Hout B, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health 2014;17(4):445–53.
9. Oppe M, Rand-Hendriksen K, Shah K, et al. EuroQol protocols for time trade-off valuation of health outcomes. Pharmacoeconomics 2016;34(10):993–1004.
10. Luo N, Liu G, Li M, et al. Estimating an EQ-5D-5L value set for China. Value Health 2017;20(4):662–9.
11. Wong E, Yeoh E, Slaap B, et al. Validation and valuation of the preference-based health index using EQ-5D-5L in the Hong Kong population. Value Health 2015;18(3):A27.
12. Al Sayah F, Mladenovic A, Gaebel K, et al. How dead is dead? Qualitative findings from participants of combined traditional and lead-time time trade-off valuations. Qual Life Res 2016;25(1):35–43.
13. Scott Long J. Regression Models for Categorical and Limited Dependent Variables. Thousands Oak, CA: Sage Publications, Inc; 1997. p. 7. Advanced Quantitative Techniques in the Social Sciences.