The Episodic Random Utility Model Unifies
Worse Than Death and Better Than Death TTO Responses in Health State Valuation

ABSTRACT

Objective: To introduce an episodic random utility model (RUM) that remedies concerns with worse than death (WTD) responses and unifies all discrete-choice approaches in health state valuation.

Methods: After reviewing the theoretical differences between the episodic RUM with the more common instant RUM, we demonstrate the empirical implications of these differences by estimating the TTO values of 42 states from the seminal United Kingdom Measurement and Valuation in Health (MVH) study, using ordinary least squares (OLS). For this comparison, the instant values are estimated with and without Dolan's transformation of WTD responses. Secondly, we motivate and demonstrate an exploded probit model that theoretically and econometrically incorporates both TTO and rank responses.

Results: By construction, instant RUM predictions are less than or equal to episodic RUM predictions, because its estimator differentially magnifies downward errors in WTD responses. The Dolan transformation of WTD responses causes TTO predictions to be more similar to the episodic RUM predictions, yet they are not equivalent (i.e., absolute mean difference = 0.192). Unlike instant predictions, the episodic RUM predictions demonstrate strong concordance with rank-based predictions (Lin's rho=0.88), and combining the TTO and rank responses improves the identification of health state values, decreasing the average width of confidence intervals from 0.203 to 0.041.

Conclusions: Episodic RUM expands upon the theoretical basis and econometrics underlying health state valuation in ways that suggest a re-examination of current TTO evidence. Furthermore, valuation surveys may build from the episodic RUM, and by incorporating multiple responses under a single estimator, this study may reduce sample size requirements.

Keywords: Rank, Quality of Life, EQ-5D, Time Trade-off

INTRODUCTION

Health state valuation studies lack a coherent theoretical framework to deal with the worse than death (WTD) responses for the estimation of quality adjusted life years (QALYs). As already indicated by Torrance in 1982, the value of better than death (BTD) states is bounded by the value of optimal health (1.00) and the value of WTD states may be as large as minus infinity [1]. In time trade-off (TTO) studies, the conventional approach to QALY estimation entails an average of bounded positive responses and large negative numbers based on WTD responses. These mean estimates are coherent with a specific random utility model (RUM). However, in practice the lower limit of the value space is bounded in measurement. For instance, in Dolan, the lowest possible value was -39 while Tsuchiya and colleagues restricted with their method negative values to -19 [2, 3]. Although such a practical boundary reduces the influence of WTD responses, the influence of the large negative numbers was still so massive that the face validity of the predictions was questioned: the raw mean predictions seem too low to be in concordance with the underlining construct of quality of life.

Confronted with this threat to face validity, researchers typically chose to impose a boundary of negative one on health state values. This suggestion of Torrance from 1982 applies for all valuation studies in QALY research, not just TTO and standard gamble (SG) [1, 2, 4]. The boundary of negative one reduces the influence of negative outliers and gives an appealing mirror image for the valuations space above zero. Nevertheless, from early on, critics have warned that there is no theoretical justification for the value -1.00, which means that the truncated scale may not really represent 'utility' [5].

In pursuit of an explanation underlying this negativity, studies reported that respondents find it more difficult and make more errors in estimating negative values than estimating positive values, especially in the TTO tasks [6]. These psychometric complications are reflected in the higher variance in negative values the low discriminating power of negative values and the ‘gap-effect’; the discontinued scale around the value of death [7-9]. While the evidence on the influence of state-specific heteroskedasticity is mounting, there is not yet a clear and coherent framework for combining BTD and WTD TTO responses in view of this heteroskedasticity.

Recently there has been considerable interest in estimating health state values suitable for QALY calculations from ranking exercises [6, 7, 10]. Ranking is seen as a relatively easy valuation method, like the visual analogue scale (VAS), and shown to render predictions that are concordant (if not identical) to VAS predictions [7]. The advantage of ranking versus VAS is the absence of response spreading and context effects and a well develop theoretical foundation in Item Response Theory [11]. Moreover, unlike VAS, ranking is a choice-based approach, which provides a basis for its merger with economic oriented choice-based methods with weaker psychometric properties, like TTO and SG. A drawback for both VAS and ranking is their unclear relation to health state values on the QALY scale, a relation which is better described for TTO and SG.

A theoretically driven model that reduces the difference between a psychometrically strong method (e.g., ranking) and a method with a strong link to utility theory (e.g., TTO) has the potential to revolutionize the field of health state valuation. This model would increase the 'convergent validity' of related psychometric and econometric methods, and therefore, enhances the 'construct validity' of these methods [12]. In the absence of a 'gold standard' in health state valuation, such an increase in convergent validity is a way forward to better understanding the latent construct of quality of life and its assessment. Furthermore, if such a model reduces dependence on arbitrary deviations from utility theory, such as negating the use of ad hoc corrections of WTD responses in the QALY paradigm, then the model would promote face validity. Lastly, such a model might further improve upon the validity of QALYs by integrating the benefits of psychometric and econometric methods under a single statistical estimator.

In this paper, we introduce an episodic random utility model (RUM) as such a theoretical framework. This model not only allows for the comparisons between rank and TTO predictions within a common model, it resolves key econometric and psychometric issues that inhibit TTO-based valuation. In introducing this model, the difficulties with the face validity of WTD responses are addressed in a way that is theoretically coherent and improves upon the convergent validity between TTO and rank-based predictions. The remaining negative values under the episodic RUM predictions also have stronger face validity than the conventional mean estimates with and without the introduction of an ad hoc bound of negative one. The initial TTO results in this paper are estimated using a relatively simple non-parametric estimator; however, maximum likelihood estimation is required to integrate TTO and ranked responses under a single estimator. For purposes of illustration, the conventional and episodic RUM are estimated using the United Kingdom Measurement and Valuation of Health (MVH) study data [2, 13, 14].

METHODS

Episodic and Instant Random Utility Models (RUMs)

The utility of a health state, j, over time, t, for individual, i, is random and may be represented by either:

$$U_{ij}(t) = \mu_j t + \varepsilon_{ij} \qquad or \qquad U_{ij}(t) = \mu_j t + \varepsilon_{ij} t \qquad (1)$$

The first model can be called an episodic RUM as the error, ε_{ij}, is independent of duration (e.g., the 10 year time frame in the UK MVH protocol). The second model, which can be called an instant RUM, is the theoretical basis underlying the conventional approach to TTO valuation studies. It assumes that error in episode utility is duration dependent. Specifically, the magnitude of error is proportional to the duration of the episode; therefore, more time in state j coincides with more error in the valuation. While this difference between episodic and instant RUMs is subtle, it has multiple econometric implications. This is particularly true for WTD TTO responses, where the respondent chooses the amount of time in state j. By changing the time spent in state j the amount of error changes under the instant RUM and the difference between episodic and instant RUM becomes relevant. Before discussing these implications, it is necessary to review the basis of random utility models and the QALY scale.

On the QALY scale, the state-specific component, μ_j, is bounded between one and minus infinity, where one represents optimal health and zero represents dead. The error term, ε_{ij}, represents randomness in health-related utility due either to individual or state-specific variability. Because optimal health and dead anchor the scale, the error term for optimal health and immediate death are constant and zero. In other words, for either the episodic or instant RUM, the utility of dead for any duration is zero (i.e., $U_{dead}(t) = 0$) and the utility of optimal health for any duration equals the duration (i.e., $U_{optimal}(t) = t$) [7].

Both models assume constant proportionality, which suggests that the expected utility of a health state is proportional to its duration, *t*, and that expected error is zero. State-specific components and errors may depend on the duration (e.g., $\mu_j(t)$) [15, 16]. Questions concerning duration effects in health state valuation are outside the scope of this paper and left to be examined in future work.

TTO RUM Estimation
TTO responses may inform the estimation of either the episodic or the instant RUM state-specific components, μ_j. Therefore, both models are shown in each of equations 2 through 5. As part of the TTO task, respondents provide one of two possible responses for each hypothetical health state, *j*. If ten years in the health state, *j*, is better than dead (BTD), the respondent determines the duration in optimal health, t_1, such that:

$$U_{ij}(10) = U_{ioptimal}(t_1) \quad \Rightarrow \begin{cases} t_1 = \mu_j 10 + \varepsilon_{ij} & \textit{Episodic RUM} \\ t_1 / 10 = \mu_j + \varepsilon_{ij} & \textit{Instant RUM} \end{cases} \tag{2}$$

The interpretation of the BTD response, t_1, is for all intensive purposes equivalent under episodic and instant RUMs, because the amount of time in the state *j* is equal regardless of response (i.e., ten years). On the other hand, if ten years in the health state, *j*, is worse than dead (WTD), the respondent determines the duration in optimal health, t_2, such that:

$$U_{ij}(10 - t_2) + U_{ioptimal}(t_2) = U_{idead}(10) \quad \Rightarrow \begin{cases} -t_2 = \mu_j (10 - t_2) + \varepsilon_{ij} & \textit{Episodic RUM} \\ -t_2 / (10 - t_2) = \mu_j + \varepsilon_{ij} & \textit{Instant RUM} \end{cases} \tag{3}$$

The interpretation of these WTD responses, t_2, differs greatly between the episodic and instant RUM estimates. For the estimation of episodic RUMs using equations 2 and 3, imagine a regression without a constant. In equation 2 the dependent variable is t_1 and the independent variable is 10 and in equation 3 the dependent variable is $-t_2$ and the independent variable is $10-t_2$. The state-specific component, μ_j, is just the coefficient of this simple regression. The instant RUMs in equations 2 and 3 show that the instant component estimator is just the sample mean of the transformed responses.

$$\hat{\mu}_j = \begin{cases} \left(\sum_{BTD} 10 t_{1i} + \sum_{WTD} - t_{2i}(10 - t_{2i}) \right) / \left(\sum_{BTD} 10^2 + \sum_{WTD} (10 - t_{2i})^2 \right) & \textit{Episodic RUM} \\ \frac{1}{N} \left(\sum_{BTD} t_{1i} / 10 + \sum_{WTD} t_{2i} / (t_{2i} - 10) \right) & \textit{Instant RUM} \end{cases} \tag{4}$$

Both estimators are non-parametric, and they are equivalent if the sample includes only BTD responses, t_1. The central difference between them is in how they treat the WTD responses. To illustrate, equation 5 translates the estimator into a function of sample parameters,

$$\hat{\mu}_j = \begin{cases} \left(p_{BTD}\bar{d}_1 + p_{WTD}\left(-\bar{d}_2\left(1-\bar{d}_2\right)+s^2\right)\right)/\left(p_{BTD} + p_{WTD}\left(\left(1-\bar{d}_2\right)^2 + s^2\right)\right) & \textit{Episodic RUM} \\ p_{BTD}\bar{d}_1 + p_{WTD}\left(\frac{1}{N_{WTD}}\sum_{WTD} d_{2i}/\left(d_{2i}-1\right)\right) & \textit{Instant RUM} \end{cases} \quad (5)$$

Where d is $t/10$, s^2 is the sample variance of d_2, and p is the proportion of the sample that is either BTD or WTD. If the sample only includes WTD responses (i.e., $p_{WTD} = 1$) and the variance of d_2 is small, the difference between the episodic and instant estimators becomes the difference between an odds based on averages and an average of odds. Also noteworthy is the fact that because d_2 is bounded between zero and one; therefore, it may be appropriate to constrain its 95% confidence interval within those bounds. If so, the greatest possible variance of d_2 is around 0.0625, however the bound depends on the sample mean: $Min\left\{\left(\frac{\bar{d}_2}{1.96}\right)^2, \left(\frac{1-\bar{d}_2}{1.96}\right)^2\right\}$.

In equation 5, the instant RUM estimator is a weighted sum of fractions. If respondent, i, erroneously increased d_2, this error would be magnified in the respondent's fraction. If the scale allowed, this error might cause the fraction to equal negative infinity, because the error simultaneously increases the numerator and decreases the denominator, making a large contribution to the sum. On the other hand, the episodic RUM estimator is a fraction of weighted sums making it more stable to the addition or subtraction of bounded values.

In an effort to improve the face validity of the instant RUM predictions, Dolan replaced the negative respondent values with $-d_2$, while Shaw and colleagues divided the negative values by a constant representing the minimum value in the transformed TTO range (i.e., 39) [2, 17]. Each transformation attenuates the magnifying effects in the fraction, but these arbitrary approaches can not be nested within either the instant or episodic RUMs, or within any other utility theory [5].

Mixing TTO, Rank and RUM

While TTO estimation does not require further specification to produce consistent results, it may be more efficient to assume that the errors are normally distributed for the estimation of the episodic RUM. This assumption allows for maximum likelihood estimation and, more importantly, the merger of rank and TTO responses under a single estimator.

Craig, Busschbach and Salomon demonstrated that ranks can be decomposed into a series of pair-wise comparisons for rank-based health state valuation using an exploded probit model[7, 18]. Because in all EQ-5D rank responses within the UK MVH-protocol all hypothetical states last ten years (i.e., t does not vary), their model estimates agrees with either the episodic or instant RUMs. Under the assumption of normally distributed errors, the probability of dominance for each pair-wise comparison in the rank responses is represented by:

$$\Pr\left(U_{ij}\left(t_j\right) > U_{ik}\left(t_k\right)\right) = \Pr\left(\varepsilon_{ij} - \varepsilon_{ik} < \mu_j t_j - \mu_k t_k\right) = \Phi\left(\frac{\mu_j t_j - \mu_k t_k}{\sqrt{\sigma_j^2 + \sigma_k^2}}\right) \quad (5)$$

The exploded probit estimation can predict the state-specific components and variances. While health states clearly have different expected utilities, Craig, Busschbach and Salomon showed that differences in variances (i.e., $\sigma_j \neq \sigma_k$) have little effect on the predicted values[7]. Therefore, in this paper, we estimated a homoskedastic probit model using rank responses and predict values for 42 health states on the QALY scale with fixed anchors for comparison with the OLS episodic RUM predictions using TTO responses.

The rank-based estimator (equation 5) required slight modification to incorporate both the episodic RUM and TTO responses. In rank responses, ties occur when a respondent considers two or more states to be equivalent. TTO responses can also be described as equivalences between two hypothetical scenarios (equations 2 and 3). Therefore, TTO responses can be incorporated into the same exploded probit model using the Efron's method for ties in rank responses[19]. Specifically, the probability of a TTO response is:

$$\Pr\left(\varepsilon_{ij} = \mu_j x_j - y_j\right) = \Phi\left(\frac{y_j - \mu_j x_j}{\sqrt{\sigma_j^2}}\right)^{0.5} \Phi\left(\frac{\mu_j x_j - y_j}{\sqrt{\sigma_j^2}}\right)^{0.5} \quad (6)$$

where y equals t_1 if BTD and $-t_2$ if WTD and x equals 10 if BTD and $(10-t_2)$ if WTD. This is equivalent to a simple linear regression without a constant and an assumption of normally distributed errors with state-specific variances. The central advantage of the exploded probit is that the estimator can accommodate both TTO and rank responses.

Caution is warranted, when merging responses from different valuation techniques into a single estimator. While the estimation of state-specific components, μ_j, may benefit greatly from the added information, it remains unclear whether the TTO variance is equal to the variance found in rank responses. Completion of the TTO task entails a greater cognitive burden for respondents, which may result in greater errors. In the combined estimator, a separate variance parameter is included for rank responses, which describes the difference between the method-specific variances.

In combining TTO and rank responses within a single estimation, we increase the power of valuation studies that explore preference of respondents using both TTO and rank responses. This is the case in most valuation studies done on the basis of the MVH protocol. A problem might be that there are more ranked pairs than TTO responses. To impose balance across methods, the pair-wise comparisons were assigned a reduced weight, equal to the respondent's number of hypothesized non-anchor states over the respondent's number of pair-wise comparisons. As a result, each respondent's set of decomposed rank responses received the same weight in the maximum likelihood estimation as their set of TTO responses, and the estimator accounts for both sources of information equitably.

RESULTS

United Kingdom Measurement and Valuation of Health (MVH) Study

In 1993 the University of York administered 3395 interviews with a response rate of 64%, and collected values of 42 EQ-5D health states, and "unconscious[2, 13, 14]. The MVH protocol describes a face-to-face interview that can be separated into several sections. First, the respondents are asked to describe their own health using the EQ-5D descriptive system. Then the respondents rank 15 cards each describing a health state. This set of 15 health state cards always includes the anchor states, optimal health (11111) and immediate death. The respondents are instructed to assume that the duration of the health state was 10 years and afterwards life ended. After the ranking exercise, the subjects are asked to place each card on the EQ-VAS, often referred to as the EuroQol "thermometer". After the EQ-VAS valuation section, the deck of health state cards is reshuffled, and 13 health states are valued using the TTO method: the two missing states are 11111 and 'immediate death' as these states cannot be valued directly using standard TTO, because they serve to anchor the TTO scale. The TTO-interview is complemented by a visual aid - a TTO-probe board that graphically displays the difference in life years between health states. As previously described, the TTO task produces either t_1 or t_2 responses, each of which describes a compensating amount of time in the optimal health state.

For the TTO and rank analytical sample (N=3,333 and 3,355, respectively), respondents were excluded for a particular method (1) if only one or two states were valued (other than 11111, "immediate death", and unconscious); (2) if all states were given the same value; or (3) if all states were valued worse than "immediate death". In addition, respondents were excluded from the rank sample if they ranked death equivalent to optimal health. These four criteria motivated the exclusion of 1.8% of the rank respondents and 1.2% of the TTO respondents.

Comparison between Instant and Episodic RUM

Figure 1 illustrates the relationship between the episodic and instant RUM predictions using TTO responses for the 42 EQ-5D hypothetical health states included in the UK MVH study. As described in equation 4, the instant RUM estimates are sample means. These means are presented with and without Dolan's transformation of negative values to a bound of -1.00. The unadjusted means are substantially less than the episodic RUM predictions, depending on the quantity of WTD responses. This intuitively illustrates the effect of summing fractions, instead of taking a fraction of sums.

Figure 1 further illustrates that the adjusted means based on Dolan's transformation of negative responses is linearly related to the episodic RUM predictions[2]. Based on the correlations shown in Table 1, episodic and instant predictions are highly correlated, above 98% for Pearson's rho or Spearman's rho. However, unadjusted means poorly agree with the episodic predictions (Lin's rho=0.140). On the other hand, the adjusted means moderately agree (Lin's rho=0.827). The predictions rendered through the arbitrary correction of negative responses are still substantially different from the episodic RUM predictions (average absolute difference=0.193). While we recognize that the Dolan transformation improves concordance between the instant and episodic predictions, why settle for an instant model with an ad hoc adjustment when you can have the real thing?

Episodic RUMs using TTO, Rankings, and Both Responses
Table 1 further describes the relationship between the predictions from the three episodic RUM estimations. The episodic predictions are highly correlated, but what is more interesting is that episodic TTO predictions show stronger agreement with episodic rank-based predictions (Lin's rho = 0.884) than instant TTO predictions with Dolan's transformation of negative responses (Lin's rho=0.827). This suggests that rank responses provide similar information to TTO responses in the estimation of episodic RUM. This means that the convergence validity between the two methods is improved more by a theoretical coherent model, rather than by an ad hoc boundary of -1.00. This, in turn, increases the construct validity of both the ranking and TTO estimates for health state valuation.

Figure 2 illustrates patterns in the relationship between episodic predictions. TTO based predictions are slightly higher for mild states and lower for states near death, which suggests the potential for duration dependence in health state valuation[15, 16]. Future analysis may parameterize the duration effect and estimate the extent of adaptation; nevertheless, this result suggests strong concordance between predicted values.

Lastly, we compared the 95% confidence intervals between episodic RUM predictions using TTO responses and intervals using both TTO and rank responses. Among the 42 states, the TTO confidence intervals of 28 states overlap with their dual response counterparts. Most of these intervals (18) have their dual response counterparts nested within, suggesting concordance. In terms of interval width, the average width of the TTO interval is 0.203, ranging from 0.135 to 0.282; the average width of the dual response interval is 0.041, ranging from 0.339 to 0.214, suggesting that the use of both responses decrease the standard error in health state value predictions by nearly 80% .

DISCUSSION

In this paper, we introduce the episodic RUM, a parsimonious approach to health state valuation that is both theoretically and econometrically consistent. The findings suggest a re-analysis of all TTO survey data, and the merger of TTO and rank responses under a unified QALY estimator. To better understand this conclusion, we delineate the three major contributions of the episodic RUM.

The first contribution is the theoretical realization that under the conventional TTO approach, known as instant RUM, the error scale in WTD and BTD responses are different by construction. As shown in equation 1, BTD error is divided by ten and WTD error is divided by a number less than ten. Therefore, the instant RUM inflates the error of WTD responses, causing them to become more influential on the estimator and pulling the estimates down. Dolan's transformation of WTD responses ($-t_2$ / 10) inadvertently causes the error scale to be equivalent, but the predictions are no longer consistent[2]. On the contrary, the episodic RUM assigns the same error scale, regardless of responses type, and produces consistent results.

The second contribution is in convergent validity [12]: the episodic RUM predictions from the TTO responses strongly agree with predictions from the rank responses. In fact, this strength of agreement is larger than the agreement between rank predictions and the instant RUM predictions with the Dolan transformation of WTD responses. The results confirm that ranking and TTO are closely related, suggesting that the strengths of both methods can be combined: the sound psychometric foundations and feasibility of ranking and the face validity of the TTO as it relates closely to the QALY paradigm. Furthermore, this evidence on the promise of the episodic RUM demonstrates that Dolan's arbitrary correction of negative responses is outmoded.

The third contribution is more practical. Under the assumption of normal errors, the episodic RUM implies an exploded probit estimator that integrates rank and TTO responses and therefore increases the power of valuation studies considerably. We demonstrate that this is feasible and decreases the standard errors of the state value predictions. By merging a psychometrically strong instrument (i.e., ranks) with discrete choice data based on utility theory (i.e., TTO), predictions are more robust. However, we also recognize the appeal of the nonparametric episodic RUM estimator (equation 5).

Summary

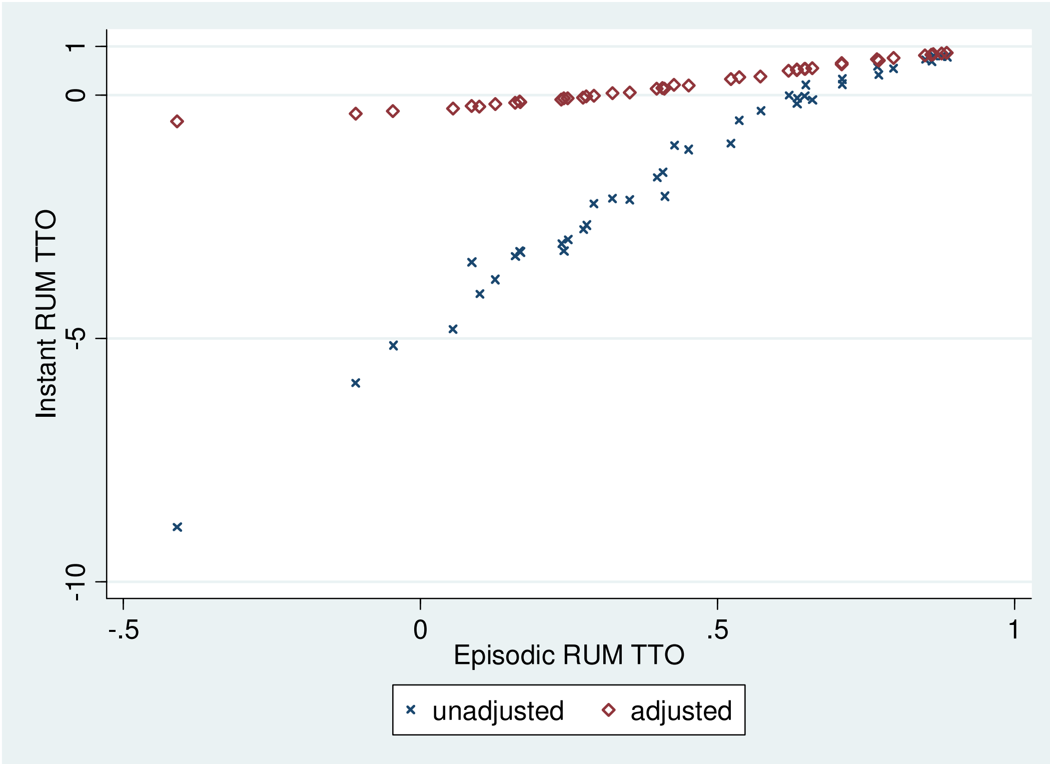
Because the instant RUM changes the error scale by type of response, because arbitrary corrections of WTD responses produce aberrant results, and because the exploded probit allows the integration of TTO, rank, SD and other discrete choice responses in a theoretically and econometrically consistent manner, the episodic RUM may replace all currently available approaches to health state valuation. In more practical terms, future valuation studies (e.g., EQ-5D-5L) can be statistically powered using the episodic RUM. A next step might be to re-estimate each country-specific valuation sets using the episodic RUM and to further examine duration effects in components and errors.

REFERENCES

1. Torrence GW: **Multi-attribute utility theory as a method of measuring social preferences for health states in long-term care. .** In *Values and Long Term Care.* Edited by Kane Rl, Kane. RA. Lexington, Massachusetts: Lexington Books; 1982: 127-156
2. Dolan P: **Modeling valuations for EuroQol health states.** *Medical Care* 1997, **35:**1095-1108.
3. Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, Hamashima C, Hisashige A, Tamura M: **Estimating an EQ-5D population value set: the case of Japan.** *Health Econ* 2002, **11:**341-353.
4. Brazier J, Roberts J, Deverill M: **The estimation of a preference-based measure of health from the SF-36.** *J Health Econ* 2002, **21:**271-292.
5. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA: **Measuring preferences for health states worse than death.** *Med Decis Making* 1994, **14:**9-18.
6. Craig BM, Ramachandran S: **Relative risk of a shuffled deck: a generalizable logical consistency criterion for sample selection in health state valuation studies.** *Health Econ* 2006, **15:**835-848.
7. Craig BM, Busschbach JJV, Salomon JA: **Ranking, Time Trade-Off and Visual Analogue Scale Values for EQ-5D Health States** *Under Review* 2008.
8. Busschbach JJV, Weijnen T, Nieuwenhuizen M, Oppe S, Badia X, Dolan P, Greiner W, Kind P, Krabbe P, Ohinmaa A, et al: **A comparison of EQ-5D time trade-off values obtained in Germany, The United Kingdom and Spain.** In *The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective.* Edited by Brooks R, Rabin R, Charro Fd. Netherlands: Kluwer Academic Publishers; 2003: 143-165
9. Stalmeier PF, Busschbach JJ, Lamers LM, Krabbe PF: **The gap effect: discontinuities of preferences around dead.** *Health Econ* 2005, **14:**679-685.
10. McCabe C, Brazier J, Gilks P, Tsuchiya A, Roberts J, O'Hagan A, Stevens K: **Using rank data to estimate health state utility models.** *J Health Econ* 2006, **25:**418-431.
11. Krabbe PF, Stalmeier PF, Lamers LM, Busschbach JJ: **Testing the interval-level measurement property of multi-item visual analogue scales.** *Qual Life Res* 2006, **15:**1651-1661.
12. Nunnally JC: *Psychometric Theory.* 2nd edn. New York, New York: McGraw-Hill Book Company; 1978.
13. Gudex C: **Time Trade-Off User Manual: Props and Self-Completion Methods.** In *Report of the Centre for Health Economics*. York, United Kingdom: University of York; 1994.
14. Kind P, Dolan P, Gudex C, Williams A: **Variations in population health status: results from a United Kingdom national questionnaire survey.** *Bmj* 1998, **316:**736-741.
15. Craig BM: **The duration effect: a link between TTO and VAS values.** *Health Econ* 2008.
16. Stalmeier PF, Lamers LM, Busschbach JJ, Krabbe PF: **On the assessment of preferences for health and duration: maximal endurable time and better than dead preferences.** *Med Care* 2007, **45:**835-841.
17. Shaw JW, Johnson JA, Coons SJ: **US valuation of the EQ-5D health states: development and testing of the D1 valuation model.** *Med Care* 2005, **43:**203-220.

18. Goldberger AS: *Econometric Theory.* New York, New York, USA: John Wiley & Sons; 1964.
19. Efron B: **The efficiency of Cox's likelihood function for censored data.** *Journal of the American Statistical Association* 1977, **72:**557-565.

Figure 1 Comparison of Instant and Episodic RUM TTO Estimates for 42 EQ-5D Health States, with and without the Negative One Ad Hoc Boundary Adjustments



* Adjustment of the TTO responses is based on Dolan's transformation of negative responses. The episodic RUM has a lowest value of just above -0.50, while the negative values of the instant RUM can be as low as -9.00.

Figure 2 Comparison of Episodic RUM Estimates using Single and Both Responses for 42 EQ-5D Health States

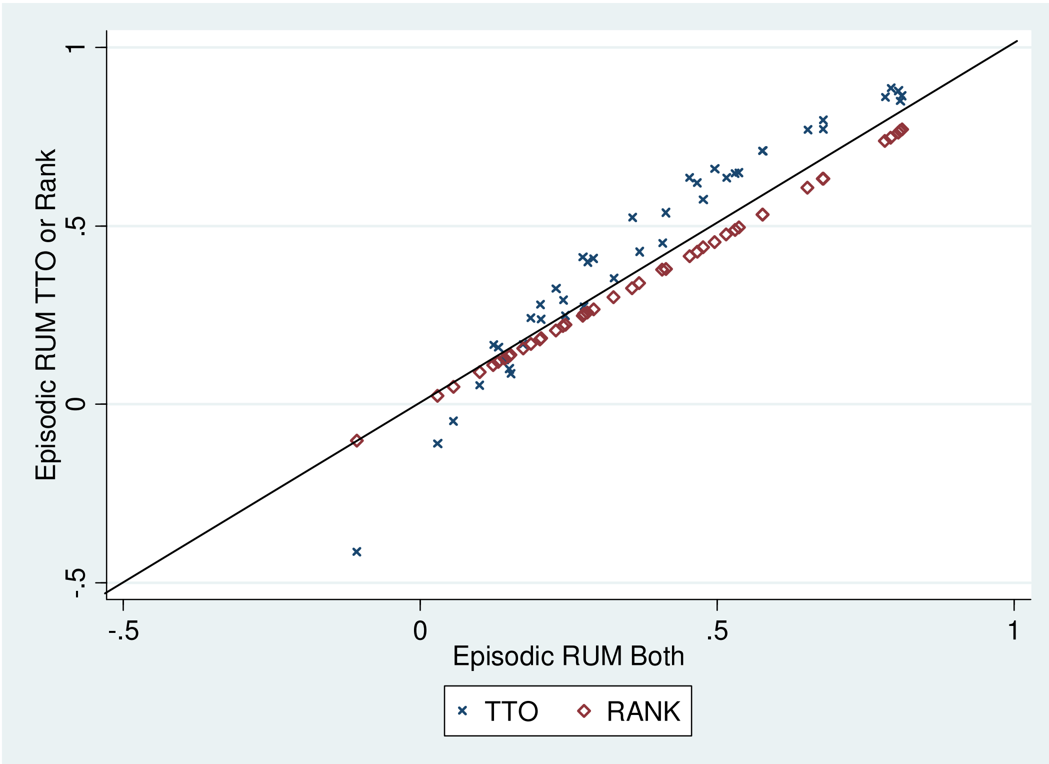


Table 1 Correlation and Agreement between Predicted Values for 42 EQ-5D States

	Comparison between Episodic RUM Estimates using TTO responses and…			
	Instant RUM Estimates using…		Episodic RUM Estimates using…	
	Unadjusted TTO Responses	Adjusted TTO Responses*	Rank Responses	TTO & Rank Responses
Correlation				
Pearson's rho	0.982	0.985	0.968	0.972
Spearman's rho	0.994	0.999	0.986	0.987
Agreement				
Lin's rho	0.140	0.827	0.884	0.925
Mean absolute difference	2.110	0.193	0.115	0.090

* Adjustment these TTO responses is based on Dolan's transformation of negative responses.