Published in final edited form as:
Med Care. 2009 June ; 47(6): 634–641. doi:10.1097/MLR.0b013e31819432ba.

Modeling Ranking, Time Trade-Off and Visual Analogue Scale Values for EQ-5D Health States:

A Review and Comparison of Methods

Benjamin M. Craig, Ph.D.[*],
Assistant Member, Health Outcomes & Behavior Program, Moffitt Cancer Center, Tampa, Florida, and Courtesy Associate Professor, Department of Economics, University of South Florida, Tampa, Florida

Jan J. V. Busschbach, Ph.D., and
Professor, Department of Medical Psychology and Psychotherapy Erasmus MC

Joshua A. Salomon, Ph.D.
Associate Professor of International Health, Department of Population and International Health, Harvard University

Abstract

Background—There is rising interest in eliciting health state valuations using rankings. Due to their relative simplicity, ordinal measurement methods may offer an attractive practical alternative to cardinal methods, such as time trade-off (TTO) and visual analog scale (VAS). In this paper, we explore alternative models for estimating cardinal health state values from rank responses in a unique multi-country database. We highlight an estimation challenge pertaining to health states just below perfect health (the 'non-optimal gap') and propose an analytic solution to ameliorate this problem.

Methods—Using rank, a standardized protocol developed by the EuroQol Group, TTO and VAS responses were collected for 43 health states in eight countries: Slovenia, Argentina, Denmark, Japan, Netherlands, Spain, United Kingdom, and United States, yielding a sample of 179,431 state responses from 11,483 subjects. States were described using the EQ-5D system, which allows for three different possible levels on five different dimensions of health. We estimated conditional logit and probit regression models for rank responses. The regressions included 17 health-state attribute variables reflecting specific levels on each dimension and counts of different levels across dimensions. This flexible specification accommodates previously published valuation models, such as models applied in the United Kingdom and United States. In addition to fitting standard conditional logit and probit models, which assume equal variance across health states (homoskedasticity), we examined a heteroskedastic probit model that assumes no variance for the two points anchoring the scale ("optimal health" and "dead") and relaxes the equal-variance assumption for all other states. Rank-based predictions for the 243 unique states defined by the EQ-5D system were compared to predictions from conventional linear models fitted to TTO and VAS responses.

Results—By construction, the TTO and VAS models assume no variance around the anchoring states of "optimal health" and "dead." Mimicking this assumption in the probit rank models helps dissolve the 'non-optimal gap.' For all other states, variances in TTO and VAS were negatively associated with mean values, which contradict the assumption of homoskedasticity. Estimated health state values from the heteroskedastic probit model for the ranking data were highly correlated with predictions from both TTO and VAS models for the 243 EQ-5D states. Between VAS and rank-

[*]Corresponding Author, Moffitt Cancer Center, 12902 Magnolia Drive, MRC-CANCONT, Tampa, FL 33612-9416; Phone: (813) 745-6710; Fax: (813) 745-6525; benjamin.craig@moffitt.org

based estimates, Lin's rho, a measure of agreement, was over 0.98 with a mean absolute difference of 0.028. Corresponding measures of agreement between TTO and rank estimates were 0.96 and 0.12, which is similar to the agreement between VAS and TTO.

Conclusions—Rank-based valuation techniques, which offer advantages of flexibility, generalizability, and ease of administration, may be attractive substitutes for TTO and VAS in the measurement of societal values for health outcomes.

Keywords

Rank; Quality of Life; EQ-5D; Time Trade-off

INTRODUCTION

The use of ranking methods in eliciting values for health states (1-4) may facilitate data collection where cardinal methods such as time trade-off (TTO) and visual analogue scale (VAS) are not feasible, for example, in populations with limited literacy and numeracy (5). Even in highly-educated populations, the relative flexibility and ease of administration of ranking methods may make this approach an attractive alternative to standard utility measurement techniques. For instance, the EuroQol group is currently exploring the possibility of using ranking to estimate health-state values in a new 5-level version of the widely used EQ-5D classification system (6-8). Unlike standard gamble and time tradeoff techniques, the use of ranking responses has not been formalized within a decision theoretic framework for constructing quality-adjusted life years (QALYs). However, if rank-based estimates are similar to those from common cardinal methods, ranking and other ordinal measurement approaches may offer useful alternatives for deriving weights for QALYs.

A key challenge in estimating health-state values from ranking data is locating the values on the 0 to 1 scale needed for QALYs. The distance between the top anchor for the scale and the next best health state—the 'non-optimal gap'—presents three methodological problems (3). First, in most measurement protocols, state of "perfect health" or "optimal health" marks the upper bound of a segmented scale, while random utility models—which underlie strategies for analyzing ordinal health-state comparisons—assume a continuous and unbounded scale. Because ranks are not scale-based, they are not subject to the compression of values in cardinal measures caused by proximity to the upper bound. Second, optimal health is logically dominant over all other states, and this dominance is evidently clear to most respondents. Few subjects report ties or rank optimal health below any other health states. This pattern complicates the estimation of values for states near optimal health since models for ranked data rely on differences in orderings to infer distances in cardinal values. Finally, standard models for analyzing ranking data, as the conditional logit regression model, typically assume that variance in health-state values is constant across states, including the states that anchor the scale ("optimal health" and "dead"). If this assumption is violated—for example, because values for mild states vary less than those for severe states—then distances between mild states will be overestimated in models that assume constant variance. Furthermore, allowing variance around anchor state values is inconsistent with TTO and VAS approaches, which take these values to be fixed.

To address these problems, we consider alternatives to the conditional logit regression model in analyses of health-state rankings in a large, multi-country dataset. In addition to comparing alternative model specifications for ranking data, we also compare rank-based predictions with predictions from VAS and TTO responses for the 243 unique states defined by the EQ-5D system.

THEORY

Random utility models (RUMs) provide a theoretical framework for health state valuation studies. The utility of a health state, *j*, for an individual, *i*, is represented by:

$$U_{ij}=\mu_j+\varepsilon_{ij} \tag{1}$$

where the state-specific component, μ_j, depends only on state attributes, and the error term, ε_{ij}, represents randomness in health-related utility, either due to fundamental variability (i.e. variation in utility of a given state across individuals) or stochastic error (i.e., variation in an individual's report on utility). This RUM implies that the probability a particular state has higher utility than another for a specific individual is given by:

$$\Pr\left(U_{ij}>U_{ik}\right)=\Pr\left(\varepsilon_{ij}-\varepsilon_{ik}<\mu_j-\mu_k\right) \tag{2}$$

Because the probability that state *j* is preferred to state *k* depends on the difference in error terms, additive individual fixed effects are not identifiable. An individual's rankings would be identical if a constant were added to all of her state-specific utilities. On the other hand, whereas additive individual effects "cancel out" in comparisons of two health states, multiplicative effects may not. Some individuals may respond with greater error than others; at the limit, an individual's orderings may be virtually uncorrelated with the ordering of the means. In this study, we assume homogeneity in individual-specific multiplicative effects.

In previous research, we built on McFadden's seminal work on discrete choice (9), and assumed that the randomness term, ε_{ij}, comes from a type 1 extreme value distribution (EV-1), $\Pr(\varepsilon \leq t) = \exp(-\exp(-t))$. Under this specification, the probability of a particular pairwise ordering depends solely on the two relevant state-specific components, μ_j and μ_k.

$$\Pr\left(U_{ij}>U_{ik}\right)=\frac{e^{\mu_j-\mu_k}}{1+e^{\mu_j-\mu_k}} \tag{3}$$

or, equivalently,

$$\ln\left(\frac{\Pr\left(U_{ij}>U_{ik}\right)}{1-\Pr\left(U_{ij}>U_{ik}\right)}\right)=\mu_j-\mu_k \tag{3a}$$

The difference between two independent EV-1 errors is logistically distributed, which offers advantages of computational simplicity, parsimony, and robustness of logit estimation (10, 11). Equation 3a shows that the difference between two state-specific components equals the log-odds of choosing one over the other.

Two disadvantages of this formulation include the assumption of constant variance across health states (homoskedasticity) and the slight asymmetry of the EV distribution. Alternative distributions may well provide better characterizations of the variance around health-state values. For example, the randomness term, ε_{ij}, may be normally distributed (i.e., following a symmetric bell-curve). A symmetric distribution allows the model to produce equivalent results under a complete reversal of order ("palindromic invariance"). This is intuitively appealing since inferences from a particular rank ordering should not vary depending on whether numbers were assigned to ranks from lowest to highest or highest to lowest. Furthermore, the normal

specification implies the familiar probit model (12), which lends itself more easily to a heteroskedastic formulation:

$$\Pr\left(U_{ij}>U_{ik}\right)=\Phi\left(\frac{\mu_j-\mu_k}{\sqrt{\sigma_j^2+\sigma_k^2}}\right) \tag{4}$$

or, equivalently,

$$\sqrt{\sigma_j^2+\sigma_k^2}\Phi^{-1}\left(\Pr\left(U_{ij}>U_{ik}\right)\right)=\mu_j-\mu_k \tag{4a}$$

where σ_j^2 represent the *j*th state's variance. We compare this heteroskedastic probit model to the standard conditional logit model.

METHODS

Data

In this study, we pool several country-specific data sets provided by members of the EuroQol Group. All studies presented in this secondary analysis (Table 1) are based on the MVH-study protocol. This protocol, first developed and applied in the United Kingdom, has been described in detail elsewhere (13-15). Details about the small differences in the national replication studies that followed the original UK study are also available (3). As we want to focus on the comparison of ordinal and cardinal measures, we only included studies of the EuroQol Group that had both rank and TTO responses based on the original MVH protocol.

In all of the country studies, health states are characterized using the EQ-5D system, which comprises a set of scores on five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with three possible levels on each. A vector of these five scores may be used as shorthand in identifying specific health states. For instance, the abbreviation 21122 represents a health state with some problems in walking, no problems with self-care, no problems with performing usual activities, moderate pain, and moderate anxiety.

The Measurement and Valuation of Health (MVH) Protocol

The original MVH-protocol describes a face-to-face interview with several sections. First, respondents are asked to describe their health using the EQ-5D system. Next, respondents order 15 or more cards each describing a health state, including the anchoring states of 11111 and "immediate death". Respondents are instructed to assume each health state persists for 10 years, followed by death. Following the ranking exercise, the subjects are asked to place each card on the EQ-VAS.

In the TTO exercise that follows, respondents decide whether ten years in the health state is preferred to 'immediate death.' If so, a series of trade-offs are presented to determine the number of years in optimal health, *x*, that is equivalent to 10 years in the state presented. If the respondent prefers `immediate death,' the interviewer presents an alternative series of trade-offs to determine the period (<10 years) in the health state, *y*, followed by the time in optimal health, (10-*y*), that together are equivalent to `immediate death.' Thus, the TTO exercise produces either an *x* or *y* response for each health state depending on whether the state is regarded as better or worse than death.

Exclusion of Respondents

In our pooled dataset, respondent data were excluded for a particular method based on any of the following four criteria: (1) only one or two states were valued (other than 11111, "immediate death", and "unconscious"); (2) all states were given the same value; (3) all states were valued worse than "immediate death"; or (4) the VAS response of "immediate death" was higher than the response for optimal health. In addition, respondents were excluded from the rank sample if they ranked death equivalent to optimal health (1.5%) and from the VAS sample if they reported a difference of less than 3 points on a 101-point interval scale (1.3%). These five criteria resulted in exclusion of 3.7% of the respondents from the rank sample (N=414), 4.5% from the VAS sample (N=470), and 1.5% from the TTO sample (N=173). No further logical consistency criteria were applied.

Pre-Estimation Rescaling of VAS and TTO Responses

For use in QALYs, VAS responses are rescaled using anchor state responses.

$$VAS^{*}_{i,j}=\frac{VAS_{i,j}-VAS_{i,Dead}}{VAS_{i,Optimal}-VAS_{i,Dead}} \tag{5}$$

where $VAS_{i,j}$ is the raw VAS valuation by respondent i for state j, $VAS_{i,Optimal}$ is the valuation of the best possible EQ-5D state (11111), and $VAS_{i,Dead}$ is the valuation of "immediate death."

For TTO responses, different transformations are required for states rated as either better than or worse than being dead. For better-than-dead states the transformation is:

$$TTO^{*}_{i,j}=x_{i,j}/10 \tag{6}$$

where $x_{i,j}$ is the (better-than-dead) TTO response for state j by respondent i.

For worse-than-dead states, various transformations are possible. The alternative that is theoretically consistent with the TTO question transforms responses as follows:

$$TTO^{*}_{i,j}=-\left(10-y_{i,j}\right)/y_{i,j} \tag{7}$$

where $y_{i,j}$ is the (worse-than-dead) TTO response for state j by respondent i.

Based on Equation (7), negative TTO values can be large compared to positive values for better-than-dead states. In the case of the MVH-protocol, the largest negative value is -19 or -39, depending on whether tradeoffs are recorded in six-month or three-month units (16, 17). When calculating mean values, large negative values could easily dominate the much smaller positive values. Some researchers have suggested that negative values be transformed into the negative one to zero scale. Dolan (13) used the following transformation:

$$TTO^{*}_{i,j}=-\left(10-y_{i,j}\right)/10 \tag{8}$$

while Shaw and colleagues (18) divided the negative values obtained by applying Equation (7) by a constant representing the largest possible negative value allowed by that study's protocol:

$$TTO^{*}_{i,j} = -\left(10 - y_{i,j}\right) / \left(39 \times y_{i,j}\right) \tag{9}$$

While both Dolan and Shaw transformations contradict theory, we transformed all worse-than-dead TTO values in this study using Dolan's method with the interest of comparability to the prevailing practice in previous work.

VAS and TTO approaches assume the anchor values do not vary. On the other hand, rank-based models based on homoskedastic specifications (e.g. conditional logit) assume the same nonzero variance for these anchor states as for all other states. As a result, the latter models should produce higher values for optimal health and lower values for "immediate death" compared to TTO and VAS models. This difference may explain the presence of the non-optimal gap found by Craig, Busschbach and Salomon (3). On the other hand, heteroskedastic models can restrict the variability of anchor state values, and produce results more similar to TTO and VAS-based estimations.

Linear Regression for VAS and TTO Values

We estimate values from VAS and TTO using ordinary least squares regression. Because the purpose of this paper is prediction and not proof of statistical significance, the inclusion of random effects—which would produce more conservative standard errors around estimated coefficients—is not necessary.

Conditional Logit Regression for Rankings

Under the EV-1 error distribution assumption and no ties in rank, the likelihood for an observed set of rank responses is:

$$L = \prod_{i \in N} \prod_{j \in K_i} \frac{\exp\left(\mu_j - \mu_c\right)}{\sum_{k \in K_i} s_{ijk} \exp\left(\mu_k - \mu_c\right)} \tag{10}$$

where j does not equal k and s_{ijk} is an indicator of the j^{th} state's dominance over the k^{th} state for the i^{th} individual.

For identification, the model requires one state-specific component, μ_c, be constrained to zero. This constraint is equivalent to dividing the numerator and denominator by $\exp(\mu_c)$ and allows for identification of the remaining difference parameters. A convenient choice for rescaling purposes is to use optimal health (11111) as the zero anchor. A practical consequence of this choice is that the conditional logit model is specified on a "disutility" scale, such that higher numbers map to worse health-state values.

As for the other QALY scale anchor, the coefficient for immediate death has an important role but an ungainly premise. If the coefficient is unity, disutility is EV-1 distributed on an inverted QALY scale. Otherwise, disutility is distributed in proportion to EV-1 on the inverted scale. In theory, the coefficient represents log odds that "immediate death" has greater disutility than optimal health. Intuitively, this log odd should equal infinity; however, such an infinite proportionality would violate the QALY scale. In practice, the coefficient on immediate death is identified primarily by the proportions of respondents rating relatively severe states as being worse than dead, rather than by direct comparisons between optimal health and death. For this reason, previous analyses using conditional logit models for empirical ranking datasets have estimated finite coefficients for immediate death and transformed all model coefficients so the

estimated coefficient for death equals 1 on the disutility scale (2-4). We adopt this same approach in the present study.

Exploded Probit Regression for Rankings

In contrast to the conditional logit, the exploded probit estimation expands ranks into a series of pair wise comparisons. Attributes for states *j* and *k* are included in the likelihood of each pair:

$$L=\prod_{i\in N}\prod_{j\in K_i}\prod_{\substack{k\in K_i\\ k<j}}\Phi\left(\frac{-1^{s_{ijk}}\left(\mu_k-\mu_j\right)}{\sqrt{\sigma_j^2+\sigma_k^2}}\right) \tag{11}$$

where s_{ijk} is an indicator of the j^{th} state's dominance over the k^{th} state for the i^{th} individual, and σ_j^2 is the j^{th} state's variance. We may further define $\sigma_j^2=\exp\left(\alpha+\beta_j\right)$ in order to constrain the estimated variances to be positive. Because optimal health and "immediate death" anchor the scale, their components are constrained to be one and zero, respectively.

Three probit specifications are examined. In the first specification, all state-specific variances are assumed equal, including anchor variances. This homoskedastic specification is similar to the conditional logit model, which has the same number of parameters. In the second specification, we constrained the variances of the QALY scale anchors to be zero. As a result, all pairwise comparisons between optimal health and "immediate death" are logically ranked, contributing no information to the estimation. Aside from anchor-specific heteroskedasticity, we constrained the non-anchor variances to be equal. The third specification relaxes the assumption of equal variances in non-anchor states. Likelihood ratio tests are used to compare these three alternative specifications. We also compare results from the second and third probit models to TTO and VAS predictions, because they share a common scale with fixed anchors.

Ties in Rank

The probit and logit likelihood functions in Equations (10) and (11) were shown in the absence of ties in rank. To handle ties, we adopt the approach suggested by Efron (19), which replaces each observed tie with two synthetic observations reflecting two possible unambiguous orderings of the given pair of states, each assigned 50% weight in the likelihood.

State-specific Component Variables

In all of the regression models for VAS, TTO, and ranking data described above, the state-specific component, μ, is modeled as a function of 18 indicator variables: two level indicators for each of the five dimensions, six indicators for the seven possible combinations of any 1s, 2s, or 3s, a count of 2s squared, and a count of 3s squared. In addition to these variables, the rank-based regressions include an indicator for "immediate death," so predicted values can be anchored on the QALY scale after estimation. An indicator for optimal health is included in the rank-based models, but its coefficient is constrained (as described in the above sections). Both the United States and United Kingdom models published previously are nested within this specification (13,18).

Comment on Software

All database work was conducted on SAS 9.1. All regressions were conducted using Stata 10 (20,21). The rologit command was used for the conditional logit estimation. The exploded

probit estimations required tailored code that incorporated ties in rank using Efron's method (19).

RESULTS

Distributional Assumptions

First, we examined variance specifications from the conditional logit and exploded probit models. If health-state values were distributed EV-1, the estimated coefficient for "immediate death" should be 1; however, the estimated coefficient is 10.28 (95% CI 10.17, 10.39), which clearly rejects the hypothesis that (dis)utility on the QALY scale is distributed EV-1. Using the probit specification, we considered whether utility is distributed standard normal (i.e., whether the variance equals 1). In our first probit model, estimated variance is 0.0161 (95% CI 0.0160, 0.0163), clear evidence against standard normality. These findings confirm that health state values on the QALY scale have smaller variances than those described by EV-1 and standard normal distributions.

Comparing the first two probit specifications (i.e., variable anchors and constant anchors), use of constant anchors increased the maximum predicted value (0.687 to 0.750) and reduced the minimum predicted value (-0.054 to -0.083) compared to variable anchors. The revised scale with constant anchors is comparable to those used by the cardinal measures. For example, the gap between optimal health and the mild health states, found in previous work (3), is likely attributable, in part, to the choice of a scale with variable anchors (i.e., conditional logit). In addition to anchor states having no variance, non-anchor states may have different variances. These findings on anchor-related compression argue against using the conditional logit model with rescaling based directly on observed rankings of optimal health and death, and in favor of the more flexible exploded probit.

Comparing the second and third probit specifications, a likelihood ratio test rejects homoskedasticity in non-anchor states at a significance level of 0.05 (4046.75 $\chi^2(49)$; p-value<0.001). While we can statistically reject homoskedasticity in non-anchor states, the difference in predicted values between the second and third specifications is small (i.e., mean absolute difference = 0.004). Only the predicted value of 'pits' (33333) changes by more than 0.01, and its value decreases from -0.083 to -0.119.

To better understand state-specific heteroskedasticity, we estimated the state-specific means and variances of the adjusted TTO and VAS responses for the 49 non-anchor states represented in the dataset for comparison to the heteroskedastic probit estimates (Figure 1). The patterns from cardinal measures suggest boundary effects in mild states, as exhibited in the negative correlation between means and standard deviations. The estimates from the heteroskedastic probit do not indicate such a relationship. There may be little predictive advantage from allowing state-specific variances in the probit model.

Ordinal and Cardinal Values for the 243 EQ-5D states

Table 2 describes the coefficients of the linear TTO and VAS models as well as the coefficients from the second and third probit models. Table 3 lists correlations and measures of agreement between ordinal and cardinal values. Using the coefficients from the heteroskedastic probit, Figure 2 further illustrates these relationships.

All predicted values are highly correlated, with Pearson's rho and Spearman's rho well above 0.95. The correlation between the rank and VAS predictions is around 0.99. Drawing a 45 degree line (0,0) to (1,1) in Figure 2, we observe VAS values and rank-based values strongly agree. Lin's rho, a measure of precision and accuracy, is over 0.98 (Table 3). The mean difference between TTO and all other predictions is substantial, over 0.11 on the QALY scale;

however, the mean absolute difference between VAS and rank-based predictions is small, around 0.028. The difference between TTO and other predictions is greatest in poorer health states (Figure 2).

DISCUSSION

This paper introduces a novel method for analyzing health state ranking data and produces an international value set based on eight countries, including five in Europe, one in Asia, one in South America, and one in North America. Key questions for future research concern sample sizes and design components needed for rank-based valuation of health states. The demand for such methodological work is increasing, because of the recognition that TTO exercises may be impractical for developing countries and arbitrary rescaling of TTO negative responses needs to be addressed. Furthermore, new descriptive systems (e.g., a 5-level EQ-5D) are being developed that require valuation studies to inform country-specific health policy making. Cardinal valuation techniques are limited to trade offs in thermometer units (VAS) or in intervals of time (TTO) or risk (standard gamble). Rank based techniques are more flexible and can incorporate a nearly limitless range of tradeoffs.

The strength of agreement between VAS and probit predictions ameliorates a major shortcoming of previous rank-based techniques, relating to the location of estimates on the 0 to 1 QALY scale, and the large gap that can arise between optimal health and the next best state when relying on rankings of these states to estimate distance. The use of constant anchors in the probit model brings the analysis of ranking data in line with approaches for analyzing VAS or TTO data, and shrinks the “non-optimal gap” considerably. We acknowledge, however, that some concerns persist about relying solely on ranking data to define distance between the best and second-best states. Inferences about this distance—whether modeling rankings with constant anchors or not—relies on either inversions or ties in the ordering of the two best states, yet optimal health logically dominates all other states. For distances between pairs of non-anchor states, inversions or ties can result from three different phenomena: heterogeneity across individuals in values for the two states, differences between the two states that fall below a minimal threshold for detection, or errors. For pairs of states that only include the top anchor, the latter two apply. Thus, treating the two types of orderings as if they arise from the same measurement process may not be appropriate. Some component of the non-optimal gap may remain even when using constant anchors. Nevertheless, the empirical finding in this paper that the use of constant anchors produces a close alignment between rank-based and VAS-based estimates indicates that this modeling strategy, in the absence of exogenous information to help define the scale, marks a clear advance over previous methods.

Sources of Randomness

The heteroskedastic probit estimation provides evidence that statistically rejects the assumption of state-specific homoskedasticity; however, the relationship between mean and variance among the health state distributions is not clearly defined. It is important to recognize that the basic structure of estimation models for ranking data makes it difficult to identify the systematic relationship between means and variances of health states that appears in TTO and VAS responses (Figure 1). This is because the ranking of health states will be relatively insensitive to many types of monotonic transformation of the underlying scale. As a simple illustration, imagine marking points along an elastic band representing cardinal valuations for a series of states. Pulling on one end of the band will expand the distances between states near this end, without changing the overall ranking of states. Similarly, compression or expansion of variances in health states that relate systematically to their locations on the scale will largely preserve rank orderings in population-level datasets, which means simultaneous identification

of distances on a latent cardinal scale and variances in distributions of individual values on this scale are difficult.

A larger issue is the randomness in rankings that may result from experimental design factors, societal heterogeneity, or individual-specific randomness in addition to inter-state covariance. For example, some respondents may answer with greater measurement error than others due to illiteracy, numeracy, or poor salience. Previous work on logical inconsistencies shows that exclusion of individuals who did not adequately complete a measurement task can significantly change estimates (22). Furthermore, some subjects may better understand a given descriptive system with the evaluation of each state or get bored as the exercise progresses, producing order-dependent variance. Either through practice, education or interest, measurement-related error can cause states to appear more similar, resulting in compression of health state values. By controlling for logical consistency and order of evaluation directly within the estimator, future estimators may reduce compression bias.

In addition to measurement error, some subjects may be more consistent in their preferences than others. Individual-specific heteroskedasticity may be incorporated in rank models, but would likely be difficult to distinguish from measurement error; one possible remedy would be to include individual covariates reflecting cognitive function in the variance regression. Besides individual-specific randomness, levels of agreement about the values of particular health states may vary across different groups of people. While mean differences across countries in TTO and VAS responses have been considered previously (5,23), we are not aware of studies that have considered cross-national differences in variance. The heteroskedastic probit described here could easily accommodate such analysis..

Summary

TTO has been the dominant choice-based method for health state valuation and the only valuation technique that accounts for variation in time. It has been criticized because of its greater cognitive burden, its constant proportionality assumption (24), and its rescaling of "worse than dead" responses to a unit interval (16). Previous research has also shown that subjects provide more logically inconsistent responses using TTO than rank, likely due to its complexity (22). Our results support the use of a simpler valuation exercise. We provide a novel estimation technique built from empirically tested assumptions and a strong theoretical foundation in random utility. The prediction results strongly agree with those of VAS and are representative of an international sample of eight countries. More research is needed to design valuation studies that best utilize these rank-based innovations.

REFERENCES

1. Kind, P. Applying paired comparisons models to EQ-5D valuations - deriving TTO utilities from ordinal preferences data. In: Kind, P.; Brooks, R.; Rabin, R., editors. EQ-5D concepts and methods: a developmental history. Springer; Rotterdamn, Netherlands: 2005. p. 201-220.
2. Salomon JA. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. Population health metrics 2003;1:12. [PubMed: 14687419]
3. Craig BM, Busschbach JJV, Salomon JA. Keep it simple: ranking health states yields values similar to cardinal measurement approaches. Journal of Clinical Epidemiology. 2008Forthcoming
4. McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. Journal of health economics 2006;25:418–431. [PubMed: 16499981]
5. Augustovski, F.; Velázquez, A.; Irazola, V. EQ-5D social values in the Argentine population; 5th World Congress: Investing in Health; Barcelona, Spain: International Health Economics Association. 2005;

6. Janssen MF, Birnie E, Bonsel GJ. Quantification of the level descriptors for the standard EQ-5D three-level system and a five-level version according to two methods. Qual Life Res 2008;17:463–473. [PubMed: 18320352]
7. Janssen MF, Birnie E, Haagsma JA, et al. Comparing the standard EQ-5D three-level system with a five-level version. Value Health 2008;11:275–284. [PubMed: 18380640]
8. Pickard AS, Kohlmann T, Janssen MF, et al. Evaluating equivalency between response systems: application of the Rasch model to a 3-level and 5-level EQ-5D. Medical care 2007;45:812–819. [PubMed: 17712251]
9. McFadden, D. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P., editor. Frontiers of Econometrics. Academic Press; New York, New York, USA: 1974.
10. Maydeu-Olivares A, Bockenholt U. Structural equation modeling of paired-comparison and ranking data. Psychological methods 2005;10:285–304. [PubMed: 16221029]
11. Maydeu-Olivares A, Bockenholt U. Modeling subjective health outcomes: top 10 reasons to use Thurstone's method. Medical care 2008;46:346–348. [PubMed: 18362812]
12. Goldberger, AS. Econometric Theory. John Wiley & Sons; New York, New York, USA: 1964.
13. Dolan P. Modeling valuations for EuroQol health states. Medical care 1997;35:1095–1108. [PubMed: 9366889]
14. Gudex, C. Report of the Centre for Health Economics. University of York; York, United Kingdom: 1994. Time Trade-Off User Manual: Props and Self-Completion Methods.
15. Kind P, Dolan P, Gudex C, et al. Variations in population health status: results from a United Kingdom national questionnaire survey. BMJ (Clinical research ed 1998;316:736–741.
16. Hawthorne, G.; Richardson, J. In: Aaronson, N.; Sprangers, M., editors. Negative Utility Scores: theoretical and practical difficulties with an essential component of utility instruments; 8th Annual Conference of the International Society for Quality of Life Research; Amsterdam, The Netherlands. 2001;
17. Patrick DL, Starks HE, Cain KC, et al. Measuring preferences for health states worse than death. Med Decis Making 1994;14:9–18. [PubMed: 8152361]
18. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Medical care 2005;43:203–220. [PubMed: 15725977]
19. Efron B. The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association 1977;72:557–565.
20. StataCorp. StataCorp LP; College Station, Texas, USA: 2007.
21. SAS. SAS Institute Inc.; Cary, NC, USA: 2007.
22. Craig BM, Ramachandran S. Relative risk of a shuffled deck: a generalizable logical consistency criterion for sample selection in health state valuation studies. Health economics 2006;15:835–848. [PubMed: 16532509]
23. Badia X, Roset M, Herdman M, et al. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. Med Decis Making 2001;21:7–16. [PubMed: 11206949]
24. Craig BM. The duration effect: a link between TTO and VAS values. Health economics. 2008

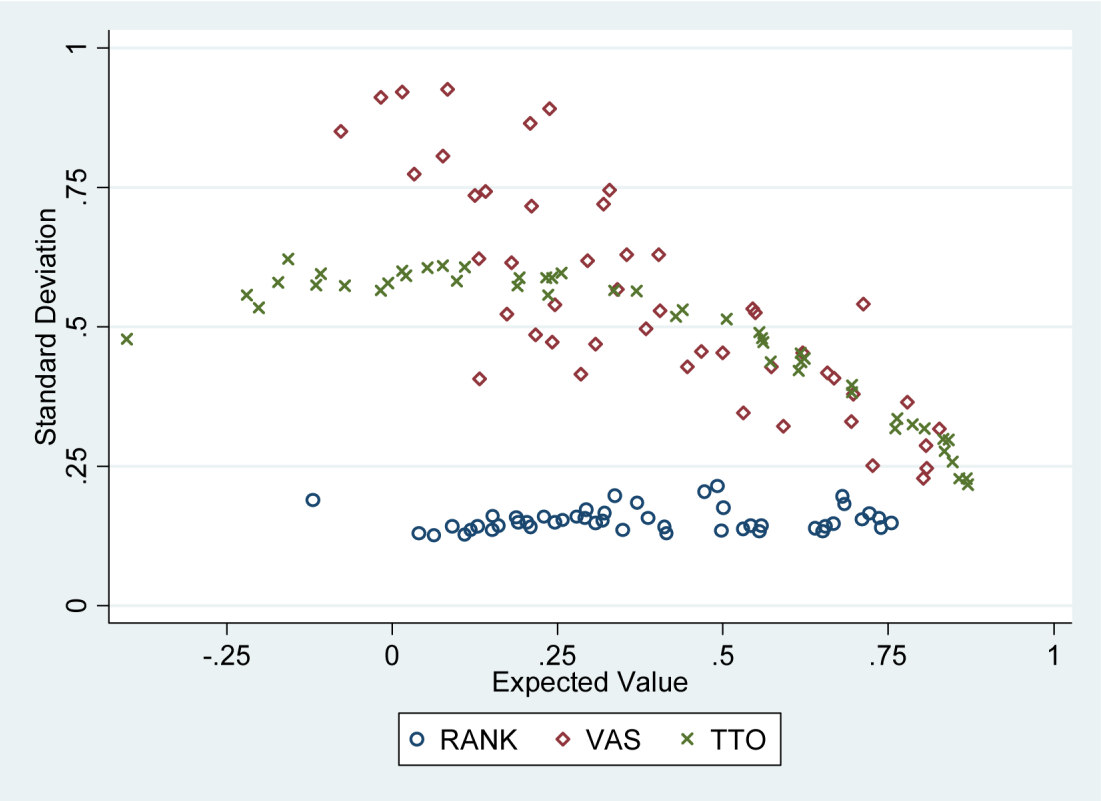

Figure 1.
The Relationship between Predicted Value and Standard Deviation for 49 Hypothetical EQ-5D States

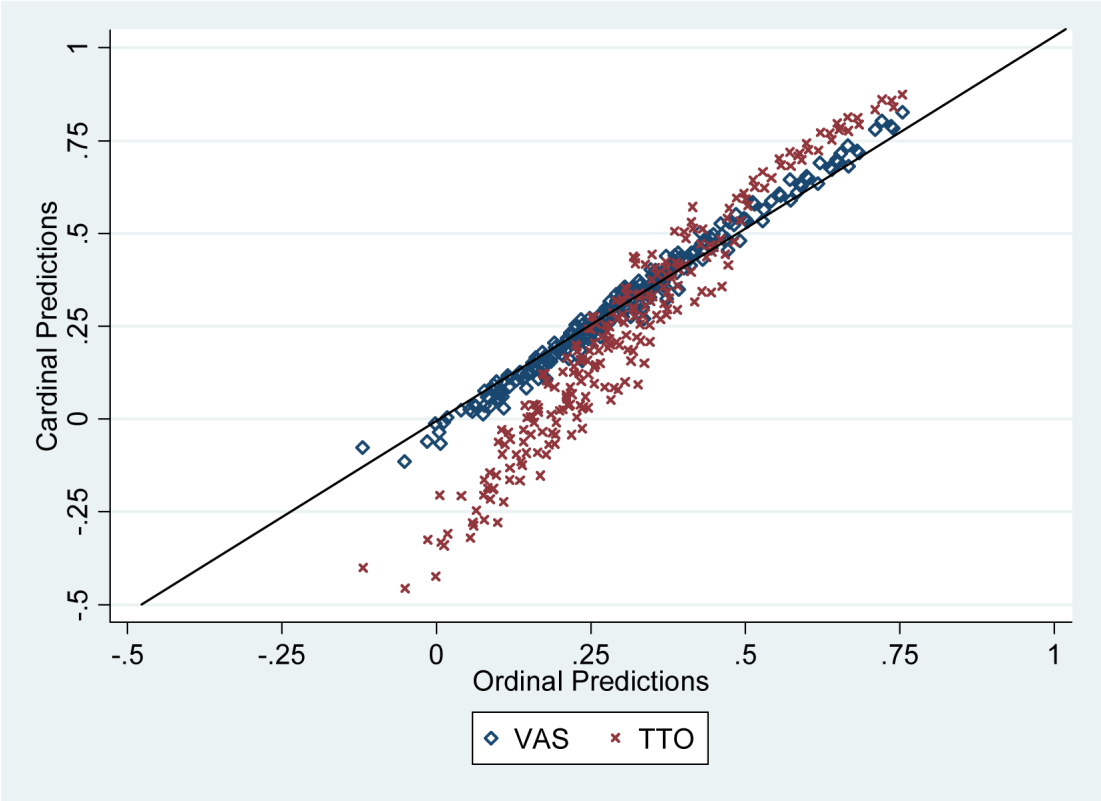

Figure 2.
Relationship between Cardinal and Ordinal Predicted Values for 243 EQ-5D States

Table 1

Number of Respondents and Responses by Method and Country

	US	UK	Japan	The Netherlands	Denmark	Spain	Argentina	Slovania
Original Sample	4025	3395	607	309	1332	979	611	225
Analytical Sample								
Rank	3969	3337	553	205	1196	978	609	222
VAS	3853	3289	607	204	1251	NA	609	221
TTO	3987	3355	551	304	1325	968	611	209
Average Number of Complete Responses per Respondent								
Rank	14.98	14.99	18.97	19.00	16.00	15.00	13.97	14.68
VAS	14.81	14.97	14.99	19.00	16.00	NA	14.03	14.65
TTO	14.83	14.81	18.80	19.00	16.00	14.89	13.97	14.67

* The study from Spain did not collect VAS responses; therefore, its respondents were only included in the rank and TTO analytical samples.

Table 2

Coefficient Estimates by Valuation Method*

	TTO (N=11,069)		VAS (N=10,034)		Homoskedastic Probit (N=11,310)		Heteroskedastic Probit (N=11,310)	
	Coef.	**p-value**	**Coef.**	**p-value**	**Coef.**	**p-value**	**Coef.**	**p-value**
Mobility, 2	-0.053	<0.01	-0.121	<0.01	-0.088	<0.01	-0.098	<0.01
Self-care, 2	-0.080	<0.01	-0.143	<0.01	-0.096	<0.01	-0.110	<0.01
Usual Activities, 2	-0.039	<0.01	-0.097	<0.01	-0.053	<0.01	-0.065	<0.01
Pain, 2	-0.073	<0.01	-0.140	<0.01	-0.066	<0.01	-0.080	<0.01
Anxiety/Depression, 2	-0.055	<0.01	-0.134	<0.01	-0.069	<0.01	-0.083	<0.01
Mobility, 3	-0.400	<0.01	-0.285	<0.01	-0.219	<0.01	-0.225	<0.01
Self-care, 3	-0.274	<0.01	-0.206	<0.01	-0.148	<0.01	-0.155	<0.01
Usual Activities, 3	-0.150	<0.01	-0.158	<0.01	-0.090	<0.01	-0.097	<0.01
Pain, 3	-0.402	<0.01	-0.237	<0.01	-0.146	<0.01	-0.153	<0.01
Anxiety/Depression, 3	-0.281	<0.01	-0.212	<0.01	-0.127	<0.01	-0.134	<0.01
Number of 2s squared	-0.003	0.19	0.010	<0.01	0.001	<0.01	0.003	<0.01
Number of 3s squared	0.006	0.02	0.009	<0.01	-0.002	<0.01	-0.001	0.24
Only 2s (i.e., 22222)	0.944	<0.01	0.822	<0.01	0.757	<0.01	0.769	<0.01
Only 3s (i.e., 33333)	0.963	<0.01	0.799	<0.01	0.686	<0.01	0.661	<0.01
Only 1s and 2s (no 3s)	0.916	<0.01	0.913	<0.01	0.802	<0.01	0.816	<0.01
Only 2s and 3s (no 1s)	0.884	<0.01	0.873	<0.01	0.712	<0.01	0.738	<0.01
Only 1s and 3s (no 2s)	0.809	<0.01	0.683	<0.01	0.616	<0.01	0.626	<0.01
1s, 2s, and 3s	0.795	<0.01	0.811	<0.01	0.658	<0.01	0.678	<0.01

Table 3

Correlation and Agreement between Predicted TTO, VAS and Rank-based Values

		Homoskedastic rank model		Heteroskedastic rank model	
	TTO & VAS	**TTO & Rank**	**VAS & Rank**	**TTO & Rank**	**VAS & Rank**
Correlation					
Pearson's rho	0.972	0.971	0.992	0.969	0.992
Spearman's rho	0.970	0.965	0.989	0.963	0.990
Agreement					
Lin's rho	0.846	0.790	0.981	0.794	0.982
Mean absolute difference	0.112	0.128	0.028	0.127	0.028