NIH Public Access
Author Manuscript
J Clin Epidemiol. Author manuscript; available in PMC 2010 March 1.

Published in final edited form as:
J Clin Epidemiol. 2009 March ; 62(3): 296–305. doi:10.1016/j.jclinepi.2008.07.002.

Keep it simple:

Ranking health states yields values similar to cardinal measurement approaches

Benjamin M. Craig[a,*], **Jan J.V. Busschbach**[b], and **Joshua A. Salomon**[c]

[a]Health Outcomes & Behavior, Moffitt Cancer Center, Tampa, FL, USA

[b]Department of Medical Psychology and Psychotherapy, Erasmus MC, Rotterdam, The Netherlands

[c]Department of Global Health and Population, Harvard School of Public Health, Boston, MA, USA

Abstract

Objectives—To examine the relationship between ordinal and cardinal valuation of health states.

Study Design and Setting—We analyzed rank, visual analog scale (VAS), and time trade-off (TTO) responses for 52 health states defined using the EQ-5D classification system developed by the EuroQol Group. We analyzed 179,431 responses from 11,483 subjects in eight countries: Slovenia, Argentina, Denmark, Japan, Netherlands, Spain, United Kingdom, and United States. We first compared responses across methods by frequency of ties and values below dead. Ordinal associations between methods were evaluated using Spearman's correlation and Kendall's tau. Next, we estimated numerical values from rank responses using country-specific conditional logit models. After anchoring predicted values on a common scale, we further investigated the cardinal relationships between rank, VAS, and TTO-based values using Pearson's rho and quadratic regression.

Results—For each country, rank responses are less likely than TTO responses to be tied and to indicate that states are worse than dead. In all countries, rank responses show a strong linear correlation with both TTO (Pearson's rho = 0.88-0.99) and VAS (rho = 0.91-0.98) responses. However, rank-based values imply greater decrements in health for mild states than cardinal values.

Conclusions—Illiteracy and innumeracy can hinder implementation of complex preference elicitation techniques in diverse settings and populations. These results indicate that ranking exercises may provide an attractive alternative for health-state valuation.

Keywords

Rank; Quality of life; EQ-5D; Time trade-off; Visual Analog Scale; Ranking

1. Introduction

A range of techniques have been proposed for eliciting valuations of different health states, but there has been no consensus on a single preferred method. Although ordinal data collection techniques, such as ranking of health states, have been included in a number of health-state valuation exercises, rank responses are not typically used to estimate cardinal valuations needed for economic evaluations. There exists a strong methodological foundation, however, for estimating cardinal values from ordinal information originating in psychology but

© 2008 Elsevier Inc. All rights reserved.

*Corresponding author. Moffitt Cancer Center, 12902 Magnolia Drive, MRC-CANCONT, Tampa, FL 33612-9416, USA. Tel.: 813-745-6710; fax: 813-745-6525. *E-mail address:* benjamin.craig@moffitt.org (B.M. Craig).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

commonly applied in areas as diverse as consumer marketing [1], political science [2], transportation research [3], and environmental economics [4].

The potential advantages of ordinal data collection approaches include relative ease of comprehension and administration and greater reliability corresponding to reduced measurement error. Particularly in settings or subpopulations in which educational attainment and numeracy are limited, an ordinal measurement strategy may have considerable practical advantages over more commonly applied techniques such as the standard gamble and time trade-off (TTO), which may place a greater cognitive burden on respondents and demand a relatively high degree of abstract reasoning. For example, valuation of health states using a standard gamble may be incomprehensible to children, who might, on the other hand, become actively engaged in the ordering of health states as a card-sorting exercise.

The notion of inferring cardinal values from ordinal choices has its origins in the pioneering work of Thurstone [5], whose law of comparative judgment provides the starting point for an array of adaptations that have followed, for example, the Bradley-Terry model [6]. Application of Thurstone's paired comparison approach to estimate health valuations was first proposed by Fanshel and Bush [7] in one of the earliest examples of a time-based health index model. Kind [8] offered another early precedent in a comparison of Thurstone and Bradley-Terry models for scaling the sleep dimension of the Nottingham Health Profile. More recently, Kind has considered Thurstone scaling of EQ-5D values from the original Measurement and Valuation of Health (MVH) study, which collected valuations in the United Kingdom for 44 states using visual analog scale (VAS) and TTO, but also included a preliminary ordinal ranking task [9].

What's new

Rank-based health-state values estimated using a conditional logit model show a strong linear correlation with both TTO and VAS values.

Ranking exercises may provide an attractive alternative to TTO for health-state valuation, particularly in diverse settings and populations.

Thurstone's law of comparative judgment may be extended in formal models that capture the relationships between rankings and attributes of the alternatives in the choice set. Salomon [10] proposed the use of a random utility approach to modeling health-state valuations from data on ordinal rankings and presented a first application using conditional logit models applied to the existing MVH data set from the United Kingdom. The results in this first application suggested that the information content in aggregate-level ordinal data may be surprisingly similar to that elicited through more traditionally recommended valuation techniques such as the TTO.

In this study, we further explore the potential of rank-based approaches to estimating cardinal valuations in a multicountry context. Using a unique data set from eight countries following a standardized data collection protocol, we examine the correspondence between TTO, VAS, and rank responses, and between the health-state values estimated from these responses.

2. Data and methods

The EQ-5D descriptive system classifies a health state as a vector of scores on five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The scores on the dimensions are "1" for the best level and "3" for the worst level. As shorthand, health states are typically described by listing the scores on the dimensions in the order noted above. For instance, a health state with some problems in walking, no problems with self-care, no problems with performing usual activities, moderate pain, and moderate anxiety is abbreviated as 21122 [11].

All studies presented in this secondary analysis are based on the so-called MVH-study protocol [12]. This protocol, originated by the MVH group of the Centre for Health Economics in York, has been described in detail elsewhere [13-15]. Below we will summarize the protocol and chronologically note minor differences in the national replication studies. As we aimed to compare ranking, VAS, and TTO, we excluded German and Zimbabwean MVH studies because they did not record results from a ranking exercise.

2.1. The MVH protocol

The original MVH protocol describes a face-to-face interview that can be separated into several sections. First, respondents are asked to describe their own health using the EQ-5D descriptive system. Then respondents rank 15 hypothetical health states, selected from the universe of 243 unique states defined by the EQ-5D system, and with the aid of printed index cards, one per state. This set of 15 health states always includes the states 11111 (the best health state) and immediate death. The respondents are asked to assume that the duration of each health state is 10 years, followed by immediate death. After the ranking exercise, the subjects are asked to place the card on a standardized VAS, often referred to as the EQ-VAS or EuroQol "thermometer."

After the EQ-VAS valuation section, the deck of health states is reshuffled and 13 health states are valued using the TTO method; the two excluded states are 11111 and immediate death as these states are used as anchors for the TTO scale and are therefore not valued directly. The TTO interview uses as an aid a TTO-probe board that helps respondents visualize the difference in life years between health states. If the respondent prefers a given health state to immediate death, tradeoffs are made using the board to determine the number of years in optimal health that is equivalent to 10 years in the state presented. The tradeoff is first described in steps of 1 year and then refined into steps of a half-year and a quarter-year depending on study. To allow an even finer discrimination between relatively good states, tradeoffs smaller than 1 year are measured in weeks.

If the respondent prefers death over the health state presented, the interviewer proceeds by asking if the respondent prefers 5 years in the given state followed by 5 years in the optimal state (11111), or again death. If the respondent decides again for death, the period in the given health state is reduced iteratively up to the point where the respondent opts for the short period in the state followed by the balance of the 10 years in the optimal health state. If instead, the respondent chooses time in the given state rather than death, the period in the state is extended iteratively until reaching the indifference point.

Upon computing health-state valuations from TTO responses, the absolute value of negative health states can be very large compared to positive values: in the case of the MVH protocol the minimal value is -19 or -39, depending on whether tradeoffs were recorded in 6-month or 3-month units. These large negative values could easily dominate the much smaller positive values. For this reason, some researchers suggest that negative values be transformed to the 0 to -1 scale [13,16]. This strategy is used in EQ-5D valuation studies in two ways: Dolan [13] replaced the negative values with $(10/x)$-1, where x represents the number of years spent in the best health state (11111), whereas Shaw and colleagues [17] divided the negative values by a constant representing the greatest negative value in the range (i.e., 39). For consistency and comparison across measures in the present study, all worse-than-dead TTO values were transformed using Dolan's transformation.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2.2. Population surveys using the MVH protocol

Researchers from the University of York first applied the protocol in 1993 in a representative sample of the noninstitutionalized adult population of England, Scotland, and Wales. They selected 6,080 addresses from a post-code address file. Of these 6,080 addresses, 756 were outside the scope of the survey, being nonresidential, empty/derelict, untraceable, or even not yet built. This resulted in a total of 5,324 in-scope addresses. A total of 92 trained interviewers completed 3,395 interviews with a response rate of 63.7%, and collected values of 43 EQ-5D health states, “Immediate Death” and the health state “Unconscious” [13,14].

The first replication of the original MVH study, the Cornellà de Llobregat study, was conducted between October and December 1996 in Catalonia, Spain [18,28]. These researchers selected a random sample of members of the general population from different socioeconomic areas in Barcelona using a primary care database. Selection was based on age and sex quota sampling so as to be representative of the Catalan general population. Individuals who were illiterate, cognitively impaired, or who had a severe illness or mental disorder were excluded. Fieldwork was carried out by 11 interviewers, and the final sample constituted 979 respondents out of the 1,930 contacted individuals (50.7%). Although the Cornellà de Llobregat study contained the same set of health states as the original study, the protocol did not include a VAS valuation section for the entire sample. VAS responses were collected from a subsample of 300 respondents, but these data were not available for this analysis.

In 1998, a reduced protocol was used in Japan, including a subset of the 43 states used in the original study. In three Prefectures, Saitama, Hiroshima, and Hokkaido, people aged 20 years and above were sampled for the survey. Tsuchiya et al. [19] interviewed 621 respondents for the collection of TTO and rank values for 17 states, whereas a second team of researchers collected VAS values using a set of 23 EQ-5D states (unpublished). As a result, the respondents did not receive the same set of states in the rank, VAS, and TTO sections: eight states were evaluated using TTO, VAS, and rank methods; 10 states were evaluated with only TTO and rank methods; “Immediate Death” and optimal health were evaluated with only VAS and rank methods; and five states and “Unconscious” were evaluated with only VAS. In addition to increasing the number of states evaluated by each respondent from 15 to 19, the smallest unit of TTO in the Japan study is 6 months instead of 3 months, which makes the lower bound of the TTO values -19 instead of -39.

In 2000, Gudex (one of the researchers in the original UK study) and colleagues replicated the MVH study in Denmark [20]. Instead of using an extensive written protocol, 1,332 respondents were interviewed using a computerized protocol. The visual aid was still the classical TTO-probe board. The researchers noted that the use of the computer was easier than using paper questionnaires, eliminated errors due to paper-computer transfer, and allowed for integrated logical consistency checks. In the Danish study 46 health states were valued, four more than in the original study, and each respondent assessed 16 states, one more than the original study. A total of 4,075 addresses were contacted. One thousand four hundred twenty-one were not at home (after three contact attempts), leaving a net sample of 2,653. One thousand three hundred twenty-one refused to participate. A total of 1,332 completed interviews make up the data used in the study, a completion rate of 50.2%.

In fall 2002, Coons and colleagues replicated the original study using a United States sample and the same EQ-5D states. Because one of the research questions was to investigate ethnic differences in the value of health, minority groups were oversampled in this study. One hundred nine interviewers collected values from 4,048 out of 5,237 respondents using rank, VAS, and TTO methods, representing a 77.3% response rate [17]. Shaw et al. used a slightly different transformation function for values below death as indicated above.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

In a study in the Netherlands, conducted in the summer of 2003, Lamers and colleagues tried to maximize efficiency by integrating the visual aid with the Danish computer interface and combining this with the Japanese low number of health states: 17 EQ-5D states, “Unconscious” and “Immediate Death.” Quota sampling conducted by a marketing firm was used to achieve a sample of 300 respondents from the Rijnmond area between the ages of 18 and 75 years, representative of the Dutch population with regard to age and gender.

Using the central population register in September 2005, the Statistical Office of Slovenia randomly selected a sample of 25 people from each of 40 different Slovenian communities [21]. Professional interviewers visited all 1,000 addresses and successfully contacted 675 subjects of which 450 refused to take part in an interview. The final sample size was 194, or 29% of the subjects that were contacted, because inconsistencies in the responses led to the removal of an additional 31 participants. In line with the MVH protocol, health states were ranked first; however, the respondents were not allowed to give VAS values that were inconsistent with the preceding ranking task. Therefore, the ranks and the ranking of the VAS responses agree perfectly by construction. TTO values were collected subsequently to the VAS administration and were allowed to disagree in rank with the VAS responses.

In 2005, Augustovski et al. in Argentina presented a recent replication of the original study [22]. Like the Spain data, consecutive subjects attending four primary care centers in large and small urban areas were selected using quota sampling according to demographic characteristics. Each third of the sample was randomly assigned a fixed set of states selected from the Japan study, but like the original MVH-study protocol, each respondent was asked to evaluate only 15 states using rank, VAS, and TTO methods. Their 14 interviewers achieved of response rate of 90% in 611 subjects.

2.3. Comparisons between methods

In all studies that included both VAS and ranking, the VAS responses directly followed the ranking of the health states, with the ranked index cards still visible to respondents. This contributed to a high correlation between the ranking and the VAS. After the VAS task, the deck of states was shuffled and TTO responses were collected. This shuffling between VAS and TTO (and thus also between ranking and TTO), contributes to additional random variance between ranks and TTO as compared to rank and VAS.

By country, we compute the average proportion of tied pairs out of all pairs for each method. Cardinal responses are measured on interval scales. The VAS has 101 possible responses and the TTO, with values recorded in quarter-year intervals, has 81 possible responses. The rank responses are measured in terms of relative values; therefore, the maximum number of unique responses depends on the number of states. Rank responses may have fewer ties than cardinal responses, because ranked states may have discordant values that fall within a common interval.

“Immediate Death” is a key health state, because quality-adjusted life years, used as the unit of health outcome in cost-utility analyses, are anchored on its value. For each country, we estimate the average proportion of worse-than-dead values in rank, VAS, and TTO responses to identify whether the ranking of dead differs by method.

VAS and TTO responses were designed to be cardinal measures of utility, but can be transformed into ordinal values for comparison with rank responses. To measure the rank correspondence between responses, we estimated the Kendall’s tau and Spearman’s rho for each respondent. Kendall’s tau is a difference between two probabilities: the probability of drawing a concordant pair of states and the probability of drawing a discordant pair of states. A concordant pair comprises two states that have the same ordering in two measures (e.g., ranking and VAS values). Discordant pairs are marked by discrepancies in ranking across two

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

measures. Suppose that a deck has eight unique, ranked pairs. If seven pairs were concordant and one was discordant, the Kendall's tau would be 0.75 (or 7/8-1/8). Spearman's rho is the Pearson's product-moment correlation coefficient computed from ranks. The tau and rho are each intended to measure the strength of relationships in rank; however, the Spearman's rho measures strength based on the squared differences between ranks and the Kendall's tau measures strength based on the probability of concordant and discordant pairs. In most cases, the correlation estimates are similar. We report the mean estimates within each country-specific sample.

2.4. Transformation and estimations

As in previous valuation studies, VAS and TTO values were transformed to a common scale anchored by "Immediate Death" and optimal health (11111). The rank expected values were estimated using conditional logit regression and a rescaling of estimate values using the same anchors. The methods are further described in detail below.

Using the EQ-VAS, respondents assigned values to health states using a 101-point rating scale from 0 (worst imaginable health state) to 100 (best imaginable health state). The VAS responses were transformed to the common scale by subtracting the VAS response for "Immediate Death" and dividing by the difference between responses of optimal health and "Immediate Death."

The TTO valuation process begins by asking the respondent whether the value of 10 years in a health state is preferred over "Immediate Death." If the state is considered better than "Immediate Death," the valuation process determines the number of years, y_1, such that 10 years in the health state equals y_1 years in the optimal health state (11111). For states worse than "Immediate Death," the unadjusted TTO response were transformed to a 0 to -1 scale using the Dolan 1994 transformation, as described above.

For the country-specific ranking data, we estimated cardinal values for each health state using conditional logit regression. Mathematical details of the model have been described previously [10], so we summarize the logic here. The model assumes that rankings of a set of items are related to latent cardinal values that are distributed around the mean levels for each item. Under this framework, a person may rank an item as being better than another item with a higher mean value due to individual variability or random error. The frequency of these reversals is related to the proximity of the mean values for different items on the latent scale. Mean values that are far apart, in other words, will produce greater agreement in orderings than mean values that are close together.

If the error terms are assumed to be independent, identically distributed, following a type-1 extreme value distribution, the difference between two errors is logistically distributed, which is advantageous due to the computational simplicity, parsimony, and robustness of logit estimation. Identification of the model requires that the mean value for one reference state be assigned a value of zero. For the purposes of rescaling, it is convenient to use "Immediate Death" as the reference state, but estimation of the model is invariant to the choice of reference states given a particular pair of anchors used to define the valuation scale.

Each country-specific data set contains a different mixture of health states. Even though the United States and the United Kingdom examined the same states, the sets of health states evaluated by each respondent differed. Under the assumption of independence from irrelevant alternatives (IIA), the state-specific estimates are set-independent as long as the rescaling of the rank values uses a common set of anchors. We test IIA using the Hausman test, which involves the comparison of maximum likelihood estimates from the full sample to a reduced sample, where one state is randomly selected to be removed from each respondent's data [23].

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

All statistical analyses were undertaken using Stata version 9.2. We note that Stata operationalizes the conditional logit regression model by leveraging the identical partial likelihood of the Cox proportional hazards model with the Efron option for ties in rank.

2.5. Value comparison

Relationships between rank-based, VAS, and TTO values were assessed by individual country using Pearson's rho and quadratic regression. Pearson's rho represents the linear correlation between values, and quadratic regression allows for curvature in the relationships. For convenience in presentation and comparison, we have reversed the value scale, such that 0 is the value of the best possible health state and 1 is the value of being dead.

2.6. Exclusion of respondents

Respondent data were removed for a particular method if only one or two states were valued (other than 11111, "Immediate Death," and "Unconscious"), if all states were given the same value, or if all states were valued worse than "Immediate Death." In cases where the VAS response of "Immediate Death" was higher than the response for optimal health, the respondent data were also removed.

In the TTO responses, the difference between "Immediate Death" and optimal health equals unity on a 40-unit range between 1 and -39; therefore, this difference equals a minimum of 2.5% (or 1/(1+39)) of the response range. For the VAS and rank estimations, we removed respondents who did not report a similar minimum difference between "Immediate Death" and optimal health. Some respondents (1.5%) were removed from the rank sample, because they ranked death equivalent to optimal health, and some respondents (1.3%) were removed from the VAS sample, because they reported a difference of less than 3 points on a 101-point interval scale.

3. Results

From the original sample of 11,483 respondents, the sample selection criteria led to the exclusion of 3.7% of respondents from the rank sample, 4.5% from the VAS sample, and 1.5% from the TTO sample. Table 1 presents the final number of subjects and average responses per subject.

3.1. Do ordinal rankings of states differ across measurement methods?

A central difference between the methods is the proportion of tied pairs out of the number of unique pairs (Table 2). Ties are more common among the TTO responses (3%-19% across countries) than among rank and VAS responses (0%-5% and 1%-8%, respectively). Differences in the number of ties are unrelated to transformation and estimation techniques, suggesting fundamental differences in information collected by method.

Another distinction between the rank and TTO responses is in the rank comparison to "Immediate Death" (Row 8, Table 2). Health states are more likely to be ranked as worse than dead using the TTO responses than the rank responses. Arbitrary transformations that bound worse-than-dead values may diminish the cardinal difference, but they cannot mitigate the ordinal difference between rank and TTO responses. The proportion of worse-than-dead responses in the rank and VAS are similar (Row 8, Table 2).

The evidence on rank correlations suggests that the direct rankings are largely consistent with the orderings implied by VAS and TTO responses (Row 12, Table 2). VAS and ranking responses are more highly correlated, which may be explained at least in part by the sequence of tasks in the data collection process. Note that as described above, in the Slovenian study,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the ranks are inferred from the VAS responses; therefore, the rank and VAS responses are perfectly correlated. The rank correlations between direct rankings and TTO responses, whereas lower than those between ranking and VAS, are nevertheless highby typical benchmarks, with mean tau values ranging from 0.61 to 0.80 and mean rho values ranging from 0.75 to 0.91. The greater fraction of tied pairs in TTO responses compared to the VAS responses contributes to the relatively lower tau estimates between the rank and TTO responses.

3.2. Does the assumption of independence from irrelevant alternatives hold?

To test the IIA assumption of the conditional logit models, we randomly removed one health-state response from each respondent’s data and re-estimated the country-specific models using the reduced sample. By comparing the estimated model coefficients in these results to those of the full sample, we produced country-specific Hausman test results. Based on a 0.05 significance level, we fail to reject IIA for four countries (United States, $P = 0.415$; United Kingdom, $P = 0.315$; Spain, $P = 0.154$; and Argentina, $P = 0.120$), and we reject for the remaining four countries (Japan, $P < 0.01$; Netherlands, $P < 0.01$; Denmark, $P = 0.027$; and Slovenia, $P = 0.003$). Based on these results, we do not find sufficient evidence to reject IIA in general, because the results from the smaller studies may be attributable to the reduction in sample sizes necessary for estimation of the consistent parameters of the Hausman test, not due to violations of IIA.

3.3. Do cardinal valuations of states differ across measurement methods?

Once the values were linearly adjusted (see Appendix table for final estimated values by state, method, and country), over 86% of the variation in VAS and TTO values was captured by the rank-based expected values (Table 3). This high correlation suggests that ranks may provide useful information for estimating cardinal values.

With the simple rescaling used here, the rank-based values imply greater decrements in health for mild states than values based on VAS, and TTO (Figs. 1 and 2). Poor health states, on the contrary, appear more similar in value, except at the extreme where rank-based values portraylower decrements in health compared to VAS and TTO values. The range of rank-based values appears narrower than the range of VAS and TTO values; however, this may be reconciled using a linear rescaling of the rank-based values.

The relationship between rank and VAS values appears linear for all countries (Table 3). For the United States, United Kingdom, Japan, and Netherlands, the relationship between TTO and rank values also appears linear; however, this relationship seems nonlinear for Denmark, Spain, and Argentina. For these three countries, the relationship appears kinked, such that the slope is flatter for states with a large portion of worse-than-death responses and steeper for states with fewer worse-than-death responses.

The Denmark and Argentina quadratic estimation gives the subtle appearance of a nonmonotonic relationship between TTO and rank in mild health states (Table 3). We believe that this is an artifact of the quadratic functional form and its a poor fit to predicted values for mild states in these two countries. Many researchers have predicted the value of all EQ-5D states using data on a subset of states, relying heavily on assumptions relating to additive functional forms in relating valuations to levels on specific domains. In this analysis, we predicted the values of only those states included in the data set. The quadratic regressions were not intended to extend the prediction of values, but to illustrate the possible curvature in the relationships between predicted values.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

4. Discussion

In this paper, we find strong correlations between values inferred from ordinal rankings and those elicited through more complex valuation techniques, such as VAS and TTO. Moreover, the strength of the relationship is consistent across eight different countries included here. Ordinal ranking tasks were originally included in the MVH protocol as a warm-up exercise for the VAS and TTO sections to better prepare respondents to judge the relative location of alternative states, rather than as a means to estimate cardinal valuations. However, the logic of inferring cardinal values from ordinal data dates back nearly a century, and econometric techniques have increasingly explored refinements of estimation methods for ordinal data. Our study contributes to the growing evidence base that ranking may provide an attractive alternative to traditional health valuation methods [9,10,24].

Further investigation of alternative models of rank responses may lead to improved agreement between ordinal and cardinal measures. The assumption of an extreme value error distribution may be replaced with normal distribution assumptions, calling for multinomial probit estimation. Further refinements of the model may also be explored, for example, various ways to relax the constraints used in the basic formulation of the conditional logit model.

An important point to note in interpreting the findings reported here is that the large gap in rank-based values between optimal health and the “mild” EQ-5D states (i.e., 11112, 11121, 11211, 12111, and 21111) relates to the estimation of distances based on reversals in ordering. In the case of 11111, estimation of distances to the next nearest states requires logically inconsistent rankings of these states as being better than 11111. By construction, optimal health cannot be ranked lower than “Immediate Death”; yet, the log odds of this event is the denominator for the rescaling of the rank-based values. Future research is needed to address the “nonoptimality” gap.

Although further investigation is warranted in the use of rankings on their own as valuation techniques, we also might consider the combination of ranks and cardinal responses into a two-stage estimation. Value estimates or value bounds might first be estimated by modeling the rank data and, in a second stage, the estimated values might be refined or updated with the cardinal response data from the VAS, TTO, or other valuation method. This two-stage estimation has the potential to improve precision and lends itself well to a Bayesian framework. It may not be a question of whether to use ranking or TTO responses, but how to use both efficiently. To do so, two incongruencies in response data need to be addressed.

Regardless of the specific country considered in this analysis, TTO responses were more likely to imply states worse than “Immediate Death” than rank responses. Macran and Kind [25] reviewed the difficulties involved with the valuation of dead, yet it remains the inextricable anchor of the quality-adjusted life year scale. Since then, the TTO method has been criticized for its constant proportional tradeoff assumption and the transformation of worse-than-dead states to a unit interval [26]. However, neither of these two criticisms addresses the fact that dead appears more favorably in the TTO responses than in the VAS and ranking responses.

The second incongruency concerns a less well-known finding: TTO responses are more likely to be tied than rank responses. The TTO is assumed to be a cardinal measure, which among other implications suggests a more refined judgment and therefore a greater differentiation between health states with values close to each other. However, the TTO responses are collected using intervals that may obscure the differences, leading to the increased number of ties. The greater cognitive burden of the TTO exercise also may contribute to respondents’ indifference between health states that may be differentiated in direct rankings or VAS responses. Craig and Ramachandran [27] showed that respondents are less logically consistent under the TTO compared to the VAS responses. This may be related to cognitive difficulty with the TTO task, interval scaling of the cardinal responses, or states appearing more similar when using TTO methods than when using VAS methods.

These incongruencies are not criticisms of the TTO in favor of the VAS or ranking methods, but they identify attributes of the TTO theoretical framework that are contradicted by existing evidence. The patterns represent an opportunity to review the location of dead in cardinal measures. We may find that the “pits” (i.e., 33333) is a better anchor for the scale than “Immediate Death” due to heterogeneous religious beliefs concerning death. TTO methods may also be improved to reduce ties, when ranks are known. The Danish work in the computer-aided collection of valuation data may enable such an approach.

Illiteracy and innumeracy can hinder efforts to measure societal preferences over health states, particularly in the developing world. In settings of relatively low educational attainment, common valuation instruments, such as TTO, may not be feasible. Simple health-state ranking techniques may succeed under these circumstances and capture preferences from less numerate respondents. Ranking is a rudimentary process, as it involves sorting items, many of which may be logically ordered. Similar processes of prioritization are performed intuitively by most people every day, for example, when they select an outfit to wear, schedule daily activities, or order a meal from a menu. The high correlation between rank responses and cardinal valuations suggests that ranking exercises may be used more widely to deduce values of health states for application in economic modeling. This trend may in turn open the door to wider empirical study of valuations in diverse communities and respondent groups.

Appendix

Appendix

Expected values by EQ-5D state, method, and country

	Rank								VAS							TTO							
	US	**UK**	**JP**	**NE**	**DE**	**SP**	**AR**	**SL**	**US**	**UK**	**JP**	**NE**	**DE**	**AR**	**SL**	**US**	**UK**	**JP**	**NE**	**DE**	**SP**	**AR**	**SL**
11112	0.30	0.35	0.40	0.37	0.38	0.42	0.53	0.29	0.17	0.19	0.25	0.23	0.22	0.15	0.25	0.15	0.17	0.20	0.20	0.17	0.12	0.09	0.18
11113	0.57	0.61	0.45	0.54	0.71	0.65	0.58	0.44	0.44	0.52		0.47	0.64	0.47	0.50	0.40	0.53	0.28	0.52	0.57	0.33	0.47	0.30
11121	0.32	0.35	0.41	0.35	0.28	0.39	0.57	0.34	0.18	0.19	0.22	0.22	0.18	0.21	0.29	0.13	0.15	0.21	0.13	0.11	0.10	0.11	0.18
11122	0.42	0.49			0.44	0.55	0.60	0.45	0.30	0.33	0.41		0.36	0.28	0.42	0.22	0.26			0.21	0.17	0.17	0.28
11131	0.61	0.67	0.47	0.54	0.72	0.71	0.68		0.58	0.60		0.50	0.74	0.54		0.55	0.65	0.34	0.57	0.56	0.50	0.65	
11133	0.70	0.77	0.52	0.64	0.83	0.81	0.75	0.56	0.67	0.73		0.65	0.94	0.63	0.60	0.68	0.83	0.44	0.77	0.75	0.61	0.69	0.57
11211	0.3	0.35	0.39	0.35	0.30	0.44	0.57	0.30	0.17	0.19	0.12	0.20	0.13	0.21	0.24	0.13	0.13	0.18	0.11	0.11	0.13	0.11	0.12
11212					0.50								0.30							0.19			
11221					0.46								0.29							0.15			
11312	0.60	0.64	0.49	0.52	0.67	0.76	0.73	0.48	0.47	0.50		0.42	0.47	0.37	0.52	0.34	0.41	0.34	0.40	0.25	0.45	0.28	0.37
12111	0.38	0.39	0.41	0.36	0.34	0.46	0.59	0.30	0.19	0.21	0.35	0.22	0.18	0.19	0.27	0.15	0.16	0.18	0.11	0.10	0.14	0.11	0.18
12121	0.47	0.51				0.59			0.33	0.34						0.20	0.25				0.25		
12211	0.44	0.48			0.47	0.60			0.31	0.32			0.25			0.20	0.23			0.16	0.21		
12222	0.54	0.60			0.60	0.71			0.37	0.48			0.52			0.32	0.41			0.31	0.35		
12223	0.64	0.72			0.79	0.80			0.61	0.66			0.87			0.49	0.67			0.59	0.54		
13212	0.66	0.69			0.73	0.83			0.57	0.59			0.66			0.47	0.54			0.38	0.60		
13311	0.68	0.73	0.50	0.56	0.72	0.86	0.81	0.50	0.65	0.65		0.48	0.70	0.71	0.48	0.51	0.58	0.37	0.45	0.40	0.63	0.67	0.47
13332	0.81	0.86				0.97			0.83	0.89						0.83	0.95				0.90		
21111	0.33	0.37	0.43	0.39	0.33	0.43	0.60	0.34	0.17	0.20	0.19	0.27	0.15	0.25	0.29	0.12	0.12	0.21	0.08	0.10	0.12	0.15	0.17
21121					0.42								0.27							0.16			
21133	0.73	0.80			0.88	0.86			0.69	0.81			1.02			0.69	0.86			0.76	0.64		
21221					0.53								0.50							0.21			
21222	0.54	0.61			0.61	0.70			0.44	0.48			0.44			0.30	0.41			0.33	0.34		
21232	0.68	0.75			0.78	0.84	0.76	0.57	0.60	0.69	0.49		0.69	0.59	0.65	0.57	0.77			0.62	0.59	0.65	0.70
21312	0.63	0.67			0.70	0.82			0.47	0.56			0.64			0.37	0.44			0.32	0.49		
21322					0.73								0.66							0.41			
21323	0.71	0.78			0.84	0.89			0.69	0.78			0.91			0.56	0.71			0.66	0.70		
22112	0.52	0.57			0.54	0.67			0.39	0.41			0.45			0.27	0.33			0.25	0.29		
22121	0.50	0.57			0.53	0.66			0.34	0.43			0.43			0.26	0.33			0.22	0.31		
22122	0.54	0.62			0.60	0.70			0.39	0.50			0.50			0.31	0.43			0.30	0.34		
22222	0.61	0.66	0.59	0.60	0.67	0.76	0.73	0.52	0.52	0.54		0.55	0.60	0.51	0.53	0.37	0.45	0.46	0.40	0.36	0.37	0.39	0.39
22233	0.74	0.85			0.90	0.92	0.83	0.63	0.72	0.87	0.61		0.93	0.63	0.71	0.73	0.90			0.85	0.77	0.70	0.73
22322					0.78								0.68							0.40			
22323	0.73	0.82			0.88	0.92	0.85	0.63	0.70	0.85	0.79		1.02	0.64	0.70	0.61	0.78			0.67	0.74	0.65	0.66
22331	0.73	0.82			0.83	0.90			0.69	0.82			0.84			0.68	0.81			0.71	0.79		
22333					0.97								1.02							0.91			
23232	0.77	0.84	0.63	0.72	0.87	0.95	0.87	0.66	0.76	0.82		0.73	0.86	0.72	0.73	0.76	0.87	0.55	0.78	0.76	0.82	0.71	0.71

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

	Rank								VAS							TTO							
	US	**UK**	**JP**	**NE**	**DE**	**SP**	**AR**	**SL**	**US**	**UK**	**JP**	**NE**	**DE**	**AR**	**SL**	**US**	**UK**	**JP**	**NE**	**DE**	**SP**	**AR**	**SL**
23313	0.78	0.85			0.90	0.96			0.77	0.86			0.96			0.70	0.87			0.81	0.85		
23321	0.73	0.80			0.80	0.92			0.69	0.76			0.68			0.58	0.72			0.53	0.78		
32211	0.73	0.78	0.68	0.70	0.80	0.87	0.83	0.72	0.68	0.72	0.60	0.69	0.88	0.62	0.78	0.64	0.72	0.61	0.49	0.63	0.79	0.65	0.79
32223	0.79	0.88	0.77	0.85	0.93	0.96	0.89	0.76	0.86	0.90		0.83	0.98	0.68	0.87	0.77	0.94	0.71	0.77	0.90	0.90	0.63	0.98
32232	0.81	0.87			0.92	0.98			0.88	0.93			1.01			0.80	0.95			0.88	0.93		
32313	0.83	0.89	0.76	0.85	0.93	0.97	0.90	0.81	0.83	0.88		0.85	1.00	0.79	0.84	0.80	0.91	0.72	0.77	0.90	0.89	0.90	0.84
32331	0.84	0.91			0.94	1.00			0.85	0.97			1.00			0.88	1.00			0.93	0.91		
33212	0.80	0.84				0.95			0.78	0.86						0.76	0.83				0.87		
33232	0.86	0.94			0.97	1.03			0.95	0.99			1.07			0.90	1.03			0.98	0.95		
33321	0.84	0.89			0.94	1.02	0.89	0.78	0.85	0.91	0.81		1.05	0.74	0.88	0.84	0.90			0.85	0.98	0.86	1.00
33323	0.89	0.96	0.83	0.96	0.97	1.07	0.97	0.78	0.93	1.01		0.93	1.11	0.89	0.85	0.97	1.08	0.82	0.89	0.88	1.00	0.97	0.99
33333	0.97	1.05	0.87	1.06	1.11	1.15	1.00	0.88	1.05	1.12	0.86	1.02	1.22	1.02	0.92	1.08	1.20	0.89	1.03	1.19	1.03	0.96	1.12
Unconscious	0.97	0.99			1.00	1.16	1.00	0.86	1.00	1.03	1.04		1.16	0.98	0.94	1.07	1.13			1.15	1.16	1.01	1.14

NOTE. For convenience in presentation and comparison the value scales are reversed, such that 0 is the value of the best possible health state and 1 is the value of being dead.

References

[1]. Louviere, JJ.; Hensher, DA.; Swait, JD. Stated choice methods: Analysis and application. Cambridge University Press; Cambridge: 2000.

[2]. Koop G, Poirier DJ. Rank-ordered logit models: an empirical analysis of Ontario voter preferences. J Appl Econom 1994;9:369–88.

[3]. Beggs S, Cardell S, Hausman J. Assessing the potential demand for electric cars. J Econom 1981;16:1–19.

[4]. Adamowicz W, Louviere JJ, Swait JD. Combining stated and revealed preference methods for valuing environmental amenities. J Environ Econ Manage 1994;26:65–84.

[5]. Thurstone LL. A law of comparative judgment. Psychological Review 1927;34:273–86.

[6]. Bradley RA, Terry ME. Rank analysis of incomplete block designs, I. The method of paired comparisons. Biometrika 1952;39:324–45.

[7]. Fanshel S, Bush JW. A health status index and its application to health services outcomes. Oper Res 1970;18:1021–66.

[8]. Kind P. A comparison of two models for scaling health indicators. Int J Epidemiol 1982;11:271–5. [PubMed: 7129741]

[9]. 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; Netherlands: 2005. p. 201-20.

[10]. Salomon JA. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. Popul Health Metr 2003;1:12. [PubMed: 14687419]

[11]. Brooks R with the EuroQol Group. EuroQol: the current state of play. Health Policy 1996;37:53–72. [PubMed: 10158943]

[12]. Szende, A.; Oppe, M.; Devlin, N. EuroQol Group Monographs. Vol. 2. 2007. EQ-5D value sets: inventory, comparative review and user guide; p. 1-102.

[13]. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108. [PubMed: 9366889]

[14]. Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. BMJ 1998;316:736–41. [PubMed: 9529408]

[15]. Gudex, C. Report of the Centre for Health Economics. University of York; 1994. Time trade-off user manual: props and self-completion methods.

[16]. Patrick DL, Starks HE, Cain KC. Measuring preferences for health states worse than death. Med Decis Making 1994;14:9–18. [PubMed: 8152361]

[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;43:203–20. [PubMed: 15725977]

[18]. Badia X, Roset M, Herdman M. Inconsistent responses in three preference-elicitation methods for health states. Soc Sci Med 1999;49:943–50. [PubMed: 10468398]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[19]. Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, et al. Estimating an EQ-5D population value set: the case of Japan. Health Econ 2002;11:341–53. [PubMed: 12007165]

[20]. Wittrup-Jens, KU.; Lauridsen, JT.; Pedersen, KM.; Brooks, R.; Gudex, C. In: Norinder, AL.; Pedersen, KM.; Roos, P., editors. Estimating Danish EQ-5D tariffs using the time trade-off (TTO) and visual analogue scale (VAS) methods; Proceedings of the 18th Plenary Meeting of the EuroQol Group; Lund: EuroQol Group. 2002; p. 257-92.

[21]. Prevolnik Rupel, V. In: Xavier, Badia, editor. The comparison of two Slovenian valuation sets and stability of Slovenian preferences; 23rd Scientific plenary meeting of the EuroQol group; Barcelona, Spain. Barcelona: Health Economics & Outcomes Research, IMS Health. September 14-16, 2006; 2006. p. 513-6.

[22]. Augustovski, F.; Velázquez, A.; Irazola, V.; IHEA International Health Economics Association. EQ-5D social values in the Argentine population; 5th World Congress: Investing in Health-Barcelona; 2005;

[23]. Hausman J, McFadden D. Specification tests for the multinomial log-it model. Econometrica 1984;52:1219–40.

[24]. McCabe 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;25:418–31. [PubMed: 16499981]

[25]. Macran S, Kind P. 'Death' and the valuation of health-related quality of life. Med Care 2001;39:217–22. [PubMed: 11242317]

[26]. Craig BM. The effect of duration: a link between TTO and VAS values. Health Econ. 2008Epub ahead of print; PMID: 18351621

[27]. 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;15:835–48. [PubMed: 16532509]

[28]. Badia X, Roset M, Herdman M, Kind P. 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]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

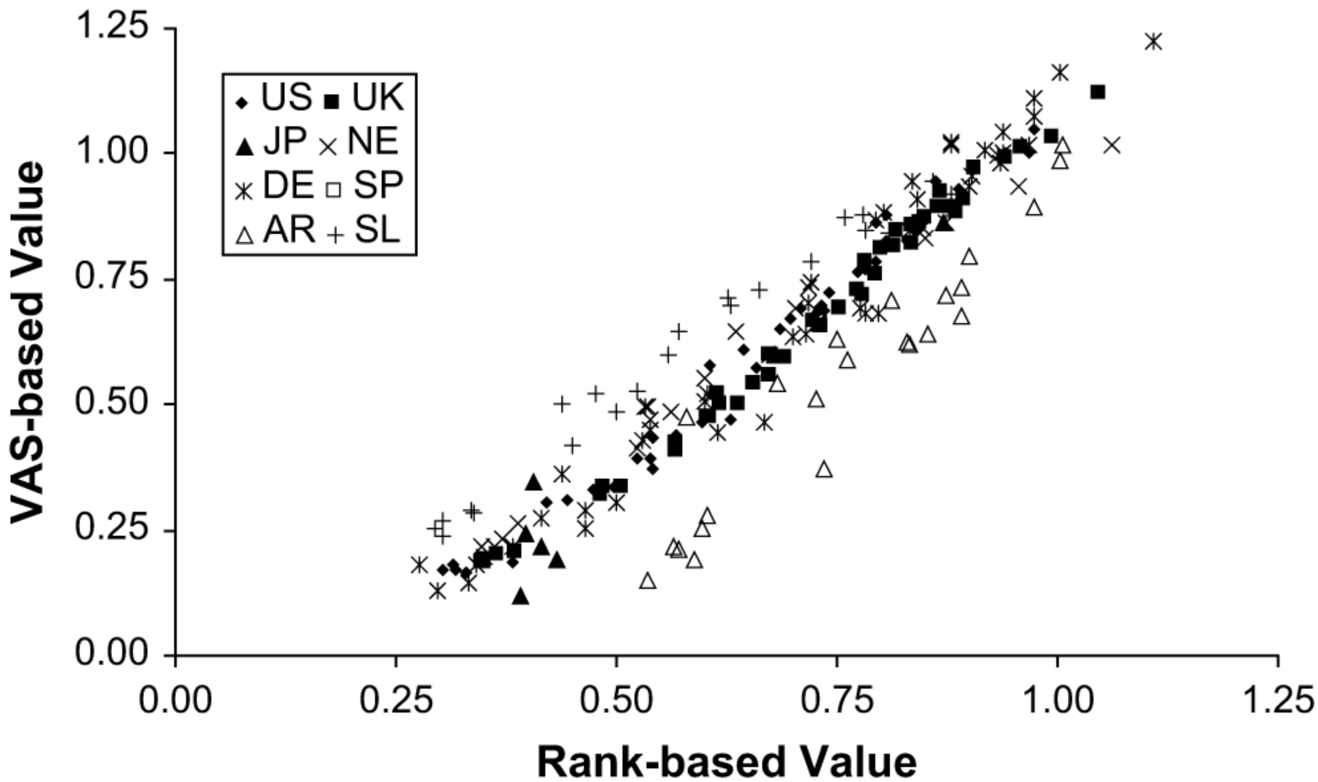

Fig. 1.

VAS and rank expected values by country. On both axes, zero represents optimal health and one represents “Immediate Death.” Abbreviations: US, United States; UK, United Kingdom; JP, Japan; NE, Netherlands; DE, Denmark; SP, Spain; AR, Argentina; SL, Slovenia.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

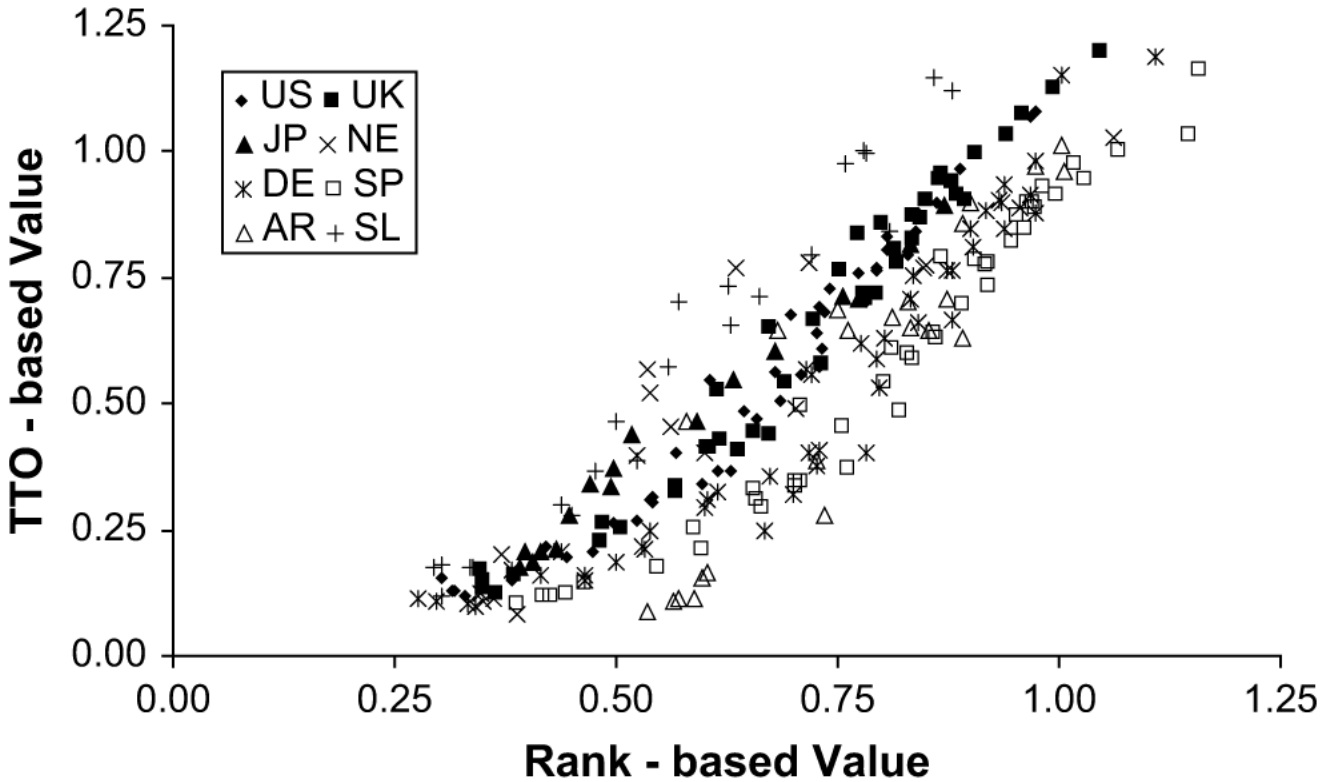

Fig. 2.

TTO and rank expected values by country. On both axes, zero represents optimal health and one represents “Immediate Death.” Abbreviations: US, United States; UK, United Kingdom; JP, Japan; NE, Netherlands; DE, Denmark; SP, Spain; AR, Argentina; SL, Slovenia.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Number of respondents and responses by method and country

	US	UK	JP	NE	DE	SP	AR	SL
Original sample	4025	3395	607	309	1332	979	611	225
Analytical sample								
Rank	3969	3337	553	205	1196	978	609	622
VAS	3853	3289	607	204	1251		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		14.03	14.65
TTO	14.83	14.81	18.80	19.00	16.00	14.89	13.97	14.67

Abbreviations: US, United States; UK, United Kingdom; JP, Japan; NE, Netherlands; DE, Denmark; SP, Spain; AR, Argentina; SL, Slovenia.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Comparison of responses by method and country

	US	UK	JP	NE	DE	SP	AR	SL
Tied pairs, %								
Rank	0.01	0.01	0.00	0.01	0.05	0.01	0.01	0.02
VAS	0.01	0.02	0.08	0.01	0.05		0.02	0.02
TTO	0.16	0.11	0.19	0.14	0.17	0.11	0.03	0.08
Worse than death, %								
Rank	0.07	0.11	0.02	0.07	0.11	0.21	0.08	0.03
VAS	0.07	0.10	0.03	0.06	0.09		0.07	0.02
TTO	0.24	0.32	0.08	0.20	0.23	0.36	0.17	0.23
VAS and rank								
Spearman's rho	0.94	0.96	0.76	0.96	0.93		0.94	1.00
Kendall's tau	0.88	0.91	0.64	0.92	0.88		0.86	1.00
TTO and rank								
Spearman's rho	0.75	0.79	0.84	0.76	0.78	0.78	0.90	0.81
Kendall's tau	0.61	0.65	0.73	0.61	0.65	0.63	0.80	0.67

Abbreviations: US, United States; UK, United Kingdom; JP, Japan; NE, Netherlands; DE, Denmark; SP, Spain; AR, Argentina; SL, Slovenia.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Comparison of expected values by method and country

	US	UK	JP	NE	DE	SP	AR	SL
VAS and rank values								
Pearson s rho	0.98	0.99	0.93	0.98	0.96		0.91	0.98
Rank value	0.88	0.89	1.23	2.07	0.81		1.07	1.92
Rank value squared	0.40	0.39	0.13	-0.65	0.46		0.36	-0.60
Constant	-0.17	-0.19	-0.30	-0.45	-0.15		-0.47	-0.28
TTO and rank values								
Pearson s rho	0.94	0.96	0.99	0.88	0.90	0.95	0.85	0.96
Rank value	-0.12	0.38	1.94	2.58	-0.91	0.17	2.77	1.17
Rank value squared	1.32	0.93	-0.41	-0.98	1.71	0.86	-0.57	0.49
Constant	0.01	-0.13	-0.52	-0.67	0.23	-0.14	-1.22	-0.26

Abbreviations: US, United States; UK, United Kingdom; JP, Japan; NE, Netherlands; DE, Denmark; SP, Spain; AR, Argentina; SL, Slovenia.