The gap effect: discontinuities of preferences around dead

Peep F. M. Stalmeier\textsuperscript{a,b,*}, Jan J. V. Busschbach\textsuperscript{c}, Leida M. Lamers\textsuperscript{d} and Paul F. M. Krabbe\textsuperscript{a}

\textsuperscript{a} Department of Medical Technology Assessment, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
\textsuperscript{b} Department of Radiation Oncology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands
\textsuperscript{c} Department of Medical Psychology and Psychotherapy, Erasmus Medical Centre, Rotterdam, The Netherlands
\textsuperscript{d} Medical Technology Assessment, Erasmus Medical Centre, Rotterdam, The Netherlands

Summary

Background: The assessment of health states considered to be worse than dead is a controversial issue.

Objective: To investigate how health states are valued when they are close to dead. Differences between adjacent states are compared with the differences between the first positive/first negative state with death.

Methods: A secondary analysis of the EuroQol EQ-5D data of the measurement of valuation and health (MVH) study was made. Visual analog scale (VAS) and time trade-off (TTO) preferences for 43 health states were obtained. Various subsets of 13 states were valued by 3395 respondents. States were rank ordered by their VAS and TTO values. Differences between adjacent states were calculated for the VAS and the positive and negative TTO values.

Results: Complete data were obtained in 2997 respondents. The differences between the ordered VAS values were equally large. In contrast, significant gaps around dead were found for the positive as well as the negative TTO values.

Discussion: These results are interpreted in light of a descriptive QALY model. This model was expanded to include utilities worse than dead. The VAS task does not pick up that bad states become intolerable, i.e. worse than dead, when they last too long, but the TTO task does. The current QALY model seems to lack descriptive validity for states valued worse than dead and for states with a maximal endurable time. Copyright © 2005 John Wiley & Sons, Ltd.

Keywords utility assessment; states worse than dead; decision analysis

Introduction

In medical health care evaluations, quality of life is a frequently assessed patient outcome. Within economical evaluations of health care, quality of life is assessed with valuation tasks. In these tasks, preferences for health states with a specified duration are assessed. One of the research questions is how to incorporate preferences for states worse than dead. It is generally assumed that there are health states worse than dead. Valuation tasks to assess preferences for states worse than dead have been proposed [1], and the resulting negative values have been incorporated in social utility valuation systems for health states. However, when preferences are assessed in patients, most investigators have chosen to forego assessments for states worse than dead, and instead assign such states a value of zero. Therefore, there seems to be a lack of agreement on how to handle states worse than dead.

To our knowledge, there have been two empirical studies dealing with the validity of preferences for states worse than dead. Patrick
[2] studied preferences for states worse than dead and recommended systematic inclusion of states worse than dead in utility assessments to provide a more complete range of health states. Dolan [3] found it difficult to predict negative utilities, whereas prediction for positive utilities was excellent.

The purpose of this paper is to examine how health states close to dead are evaluated, keeping in mind that individuals disagree about the position of dead located within a set of health states. Value differences between health states adjacent in rank order were analysed, and compared with value differences between the first positive/first negative state and dead. Two assessment techniques were analysed: the visual analog scale (VAS) and the time trade-off (TTO) method. In the Discussion section, the results will be interpreted with a descriptive quality adjusted life years model.

**Methods**

**Health states**

We used data from the MVH study [4,5]. In this study, preferences were produced for health states based on the standard EQ-5D classification system. The EQ-5D classification describes health status according to five attributes: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each attribute has three levels, i.e. ‘no problems’, ‘some problems’, and ‘severe problems’. Health state descriptions are constructed by taking one level for each attribute (e.g. 11111 represents the best health state).

Preferences were collected during a face-to-face interview. The interview comprises several sections. In one of the sections the respondents valued a subset of 13 health states out of 43 EQ-5D health states with the VAS and TTO method. Because the respondents valued different subsets, a total of 43 EQ-5D states were evaluated.

**Valuation tasks**

For both the VAS and TTO tasks, the respondents had to assume that the duration of the health state was 10 years, followed by dead. In the VAS task, the EQ-5D state 11111 and dead were also valued. In the MVH protocol, respondents were asked to rank order the states, then insert ‘immediate death’ into the ranking, and finally assess the ranked states on the VAS. The end points of this scale are ‘best imaginable health’ and ‘worst imaginable health.’ This VAS exercise employed a bisection method. First the best state was located on the VAS, then the worst ranked state, then ‘the state which came closest to being half-way on the scale’ between the already located two extreme states [6].

In the MVH study, the VAS values were transformed to a scale ranging from minus 1 to plus 1, where 0 represents the value of dead. In the analyses below, however, the raw untransformed VAS values were judged to be more informative to investigate valuations around dead, because we wanted to stay as close as possible to the original data.

In the TTO task, the 13 health states were valued in a random order. For positive states, the time trade-off with Perfect Health (PH) is used to arrive at utilities. If a respondent is indifferent between \( t \) years in Perfect Health and 10 years in a positive health state \( Q^+ \), then by the multiplicative QALY model \( V = L \times Q \), where \( L \) specifies duration, and \( Q \) the quality weight, the quality weight for \( Q^+ \) equals \( t/10 \).

Health states could also be valued as worse than dead. For such states, a negative TTO board was used. The choice was between dying immediately and spending \( s \) in the target state, followed by \( t \) (or 10 – \( s \)) years in the state 11111. If a respondent is indifferent between these two options, and again assuming the QALY model, then \( s \times U(Q^-) + t \times U(PH) = U(Dead) \). Thus \( U(Q^-) = -t/s \). To avoid large negative preferences (in the case of the MVH protocol –39.0 was the minimum) and skewed distributions, \( U \) was transformed [7] to \( U' = U/(1 - U) \) to arrive at a scale between 0 and minus 1. As a result \( U'(Q^-) = -t/10 \).

**Analyses**

Value differences between health states adjacent in rank order were analysed, and compared with value differences between health states and dead. For each respondent, we rank ordered the states by their VAS and TTO values, respectively. When ties existed, these were broken in random order. Note that the rank ordered states are not the same
for each respondent because different sets of states were used and also because individuals differ in their rank order. For each respondent, VAS difference values were calculated between all adjacent values. In addition, we determined the difference between the VAS value for dead and the first positive value, and the difference between the value for dead and the first negative value. The same procedure was followed for the TTO values. By definition, the TTO value of dead equals 0. For the TTO, differences are presented separately for positive and negative states because two different methods are used for positive and negative states.

Results

A total of 3395 respondents were interviewed, 2997 of whom yielded a set of complete data [5]. This data was used for our analyses. Figure 1 shows the percentage of states valued as better than dead, worse than dead, and equal to dead. In the VAS task (Figure 1(a)), less than 2% respondents had worse than dead preferences for the states ranked first, second, and third best. For the worst state, 45% of the respondents had worse than dead preferences. In the TTO task (Figure 1(b)), less than 2% respondents had worse than dead preferences for the states ranked first and second best. For the worst state, 95% of the respondents had worse than dead preferences. Worse than dead preferences occur more often with the TTO task.

Figure 2 shows the differences between adjacent states for the VAS task. The 12 differences between the 13 adjacent states are represented by positions 1–12. The differences between the VAS value for dead and the value of the first positive and the first negative state are represented by the last two positions, respectively. All differences are about equally large. Figure 3 shows the 12 differences between adjacent states for the VAS task (between dead and the first positive value, position 14; between dead and the first negative value, position 15).
between the 13 adjacent states for the TTO task. Again, the bars at the last two position give the difference between the TTO value for dead and the first positive and first negative value. The last two bars are much larger than the other bars.

Discussion

We analysed value differences between adjacent health states, and focused on the differences between dead and the first positive or negative health state. For the VAS scale, differences around dead are of equal size compared to the differences between the remaining states. However, for the TTO values, the differences around dead are at least twice as large compared to differences between remaining states. There appears to be a relatively large gap in the TTO scale around dead, the gap effect. This gap is in qualitative agreement with the results of Busschbach [8] who found that the means of positive TTO values are between 1 and 0.4, and that the means of negative TTO values are between −0.5 and −0.6. The gap effect is also supported by the scarcity of TTO values around zero (see Reference [3, Figure 2]).

The question is what do these gaps around dead signify? One may infer that the TTO fails to produce a continuous scale, whereas this is not the case for the VAS scale. Such an interpretation seems to suggest that differences around dead are not handled properly by the TTO task, as defined in the MVH study.

Another interpretation is that the MVH procedure is correct, but that the VAS and TTO tasks differ in other respects. A difference between the VAS and TTO tasks is the much larger rate of Worse than Dead (WTD) preferences in the TTO task. This finding was previously explored by Robinson [9]. WTD preferences occur more often in the TTO task despite the fact that in both tasks health states are being presented as lasting 10 years and as being followed by dead. One plausible explanation for the larger WTD rate in the TTO task seems to be that, compared to the VAS task, the duration of ten years is more salient in the TTO task [9]. Another perhaps even more important factor is that the TTO task, when compared to the VAS, makes more salient that states are followed by dead [10]. As a result the states are overestimated in the VAS task [11]. However, a general overestimation of the VAS values still does not explain the emergence of gaps on the TTO scale.

A more plausible interpretation is a general measurement bias [12] namely a tendency to avoid using end of scales. With the VAS, dead has no privileged position on the scale, but with the TTO, dead has a special position. The TTO for states better than dead and for states worse than dead are obtained in two separate valuation tasks. So if the very ends of the scales are not used, there will be a gap on the TTO, but not for the VAS. On the VAS scale, the best ranked state had a mean value of 0.88 and the worst ranked state a mean value of 0.02. Similarly, for the positive TTO scale, the best ranked state had a mean value of 0.92 and the worst ranked state a mean value of 0.18. These data suggest that the tendency to avoid the end of scale bias plays a role, but is not sufficiently strong to explain the large gap of 0.35 on the positive TTO scale.

Several aspects that we observed above may be explained by results from earlier studies, describing a descriptive QALY model. Whereas the normative QALY model (see section on Valuation Tasks) describes how decisions ought to be taken, the descriptive model QALY describes how people deviate from the normative model when choosing between health profiles. The descriptive model includes more than one time horizon, whereas the MVH study used a single fixed 10-years horizon. The justification for this approach is that preferences using a single time frame cannot be interpreted without considering the relation between preferences for various time frames. The descriptive QALY model is based on simple preference questions such as ‘which would you prefer: (5 years in health state Q) or (10 years in health state Q)?’ or ‘which would you prefer (5 years in health state Q) or ‘immediate death’.

The descriptive QALY model is presented in Figure 4 [13]. Utilities derived from these simple preference questions are plotted as a function of the duration of a health state. Note that in Figure 4 and 5, utility represents the cumulated benefit of survival across time. Utility in this sense is not the same as the term utility sometimes used in the literature to represent the quality weight. The quality weight is equivalent to the slope of the lines in Figure 4. Figure 5 depicts how the utilities from Figure 4 are (mis)represented when the TTO method is used to assess utilities.

For mild states (region A in Figure 4), longer durations are preferred to shorter durations. Thus, utilities are strictly positive. For these states we
assume that the slopes are constant and the QALY model \( V = L \times Q \) applies. Region A is drawn for quality weights down to 0.70, because above 0.7, states lasting 20 years are generally preferred to states lasting 10 years [14]. For these mild states, utility is plotted in region A of Figure 5 as derived from the TTO task for various durations. For mild states, utilities derived from the TTO task and the descriptive QALY model agree.

For states \( B_i \) in Figure 4, a shorter survival is preferred over a longer survival when utility drops below a certain level, for instance, ‘(6 years in health state \( B_3 \)) is preferred to (12 years in health state \( B_3 \)).’ Such a preference is called a maximal endurable time (MET) preference [15]. This phenomenon has been observed in a wide variety of populations including students [16,17], patients [14], and physicians [15]. When MET preferences apply, independence between duration and quality (as implied by the QALY model \( V = L \times Q \)) does not hold because the slope or quality weight for state \( B_i \) decreases from positive to negative with increasing duration. For health states with quality weights below 0.70, MET preferences occur in 50% of the respondents, and even more frequently if states become worse [14]. Unfortunately, the TTO task does not detect MET preferences due to the proportional heuristic that is used in the TTO task [14,16,17]. The proportional heuristic is a rule of thumb whereby respondents produce the number of years in full health in proportion with the fixed duration of the target health state [14,16,17]. Thus, for utilities derived from the TTO task in Figure 5, the curvature for the states \( B_i \) is absent and replaced by straight fanning lines.

For even longer durations, utilities may drop below 0, for instance, in Figure 4, ‘immediate death’ is preferred to (14 years in health state \( B_4 \)). Such a preference is called a worse than dead (WTD) preference. The TTO method is able to pick up that states become worse than dead, because respondents are able to switch to the negative TTO board when asked ‘what do you prefer: (10 years in health state \( Q \)) or (0 years in Perfect Health)?’ This question is in practice identical to the WTD question. As a consequence, in Figure 5, when states last so long that WTD preferences occur, the utilities drop below 0. This sudden drop causes the gap effect. Figure 3 shows that the gap size on the positive TTO board equals 0.35. In other words, states with an initial utility below 0.35 are considered worse than dead when they last longer than 10 years. This gap of 0.35 at a 10 years duration, is represented by the dotted vertical line \( B_4 \) in Figure 5. More precisely, because utilities are represented by slopes, the size of the gap on the utility scale of 0.35 is represented by the slope of the slanted full line at \( B_4 \). Figure 5 predicts that the size of the gap increases for longer durations, that is, the slopes of the slanted full lines at \( B_5, B_4, \) and \( B_3 \), increase. This prediction can be tested empirically. The question marks in Figure 5 indicate that the descriptive TTO model is unknown for states worse than dead.
In Figure 4, the MET is assumed to become shorter as states become worse. A further restriction has been imposed on Figure 4, namely, that utility curves do not cross. For a moment, let us suppose the opposite, namely that, that below death, the utility for \( B4 \) follows the white striped line in Figure 4. At the intersection point \( X \), crossing would indicate that \((12 \text{ years}, B4) = (12 \text{ years}, B6)\) and this is unlikely as \( B4 \) is preferred to \( B6 \) at shorter durations. Thus curves were not allowed to cross. As a consequence, the slopes below zero must become steeper moving from right to left in Figure 4. This result may appear counter-intuitive but if a state is already negative after even a short duration, it must be really bad, and hence its slope must be strongly negative.

From this descriptive model for the TTO, one would expect a gap for the VAS values as well. However, a gap was not observed in the VAS scale. As discussed above, this may be explained by the fact that (1) the VAS task makes durations of health states less obvious, and (2) that the VAS task makes less obvious that a health state is being followed by dead. Another reason may be found in the bisection procedure used for the MVH VAS method that is described in the Methods. This procedure is likely to make it easier, if not induce respondents, to place the states evenly across the VAS scale and, more importantly, to equalise the distance between the first positive/first negative states with dead with the distance between any other adjacent states.

Another question is what these results mean for the assessment of negative TTO utilities. First, and foremost, for negative utilities, the valuation technique should enable measurement of the slopes below zero in Figure 4. However, the formulas used to calculate negative utilities, that is \( U'(Q^-) = -t/10, \) or \( U(Q^-) = -(t)/s, \) do not measure these slopes. Second, although the gap on the positive TTO scale can be understood within the descriptive model for the TTO task outlined above in Figure 5, there is no explanation for the gap on the negative TTO scale. The unexpected gap on the negative scale could indicate that utilities derived from negative TTO boards are too negative. Third, the problem remains that it is harder to predict negative utilities than positive utilities [3,8]. This suggests that current methods to assess negative utilities need improvement. Reliable methods to assess utilities for states worse than dead are needed as, at least from the societal perspective, such states do exist in some clinical populations, for instance, in patients with chronic, refractory complex Regional Pain Syndrome Type I. Cost–utility analyses of interventions aimed at such clinical populations should not ignore effects on negative utilities.

In conclusion, descriptive QALY and TTO models were presented. These models explain why gaps occur above dead when utilities are assessed with the TTO method. For states worse than dead, present assessment methods need improvement.

Acknowledgements

This paper was presented at the EuroQol Plenary Meeting in Bled, Slovenia, 11–13 September, 2003. Financial support for this study was provided by a grant from the ZonMw ‘Netherlands Organisation for Health Research and Development’ (Project No. 95399). The funding agreement ensured the authors’ independence in designing the study, interpreting the data, writing and publishing the report. The following authors were employed by the sponsor: P.F.M.S., L.M.L.

Financial support for this study was provided by a grant from ZonMw ‘Netherlands Organisation for Health Research and Development’, Project No. 95399.

References

8. Busschbach J, Weijnen T, Nieuwenhuizen M et al. A comparison of EQ-5D time-trade-off values obtained in Germany, The United Kingdom and


