



VALUING HEALTH STATUS USING VAS AND TTO: WHAT LIES BEHIND THE NUMBERS?

ANGELA ROBINSON,^{1*} PAUL DOLAN¹ and ALAN WILLIAMS²

¹Department of Economics, University of Newcastle upon Tyne, Newcastle upon Tyne, U.K. and

²Department of Economics, University of York, York, U.K.

Abstract—It is well known that different methods of eliciting the valuations attached to various health states, such as the Visual Analogue Scale (VAS) and the Time Trade Off (TTO), yield different results. This study gathers qualitative data from a group of 43 respondents who had previously taken part in a large scale national study which set out to elicit the values attached by individuals to various health states using both the VAS and the TTO techniques. The findings of this study raised three questions which are of particular interest here: (1) Why are some states that are rated better than dead on the VAS often rated as worse than dead in TTO? (2) Why are some respondents unwilling to trade off any time at all in order to avoid a health state that they place below full health on the VAS? (3) Why are TTO valuations of older respondents for the more severe health states lower than those of the younger age groups? This study has uncovered qualitative evidence on each of these three key issues. Regarding the first question, many respondents did not appear to interpret a better than dead VAS score as a strict preference for spending 10 years in a health state over immediate death. Several different factors appeared to contribute towards this, an important one being the tendency of respondents to ignore the *duration* of the health state during the VAS task. Regarding the second question, there is evidence of the existence of a “threshold of tolerability” below which states would have to fall before some respondents would be willing to give up any time at all on the TTO. Regarding the last question, it appears that older respondents are less likely to find the worse than dead TTO scenario plausible than those in the younger age groups. However, whilst this may explain why older respondents attach lower *worse than dead* valuations to health states, it does not appear to account for the entire difference in TTO valuations between the two age groups. In addition, it appears that older respondents may be less prepared to live for the next 10 years in a diminished health state. © 1997 Elsevier Science Ltd

Key words—time trade off, visual analogue scale, health state valuation inconsistency

INTRODUCTION

In evaluating the outcomes of health care, one important element is health-related quality of life (HRQOL). It is desirable when assessing such benefits that the views of the potential beneficiaries should play a significant role. Health-related quality of life assessment methods generally involve the direct elicitation of the values attached by individuals to various states of health. However, it has been well documented that different valuation methods yield different results for identical health states (Bombardier *et al.*, 1982; Torrance, 1976; Read *et al.*, 1984; Llewellyn-Thomas *et al.*, 1984). Whilst numerous studies have explored *how* valuations differ, there has been relatively little research into *why* they differ. This study attempts to redress this imbalance, the motivation being provided, in particular, by the findings of a large scale general population study.

In 1993, the Measurement and Valuation of Health (MVH) group at the University of York conducted a study which set out to establish the relative valuations attached to different states of health (described in HRQOL terms) by members of the general public. The study used the EuroQol classification system (see Appendix A) which describes HRQOL in five dimensions—mobility, self-care, usual activities, pain and anxiety—with no disease specificity (see EuroQol Group, 1990). A health state description consists of one statement from each of these dimensions. Three levels are possible within each dimension—no problems, some problems and severe problems—denoted by 1, 2 and 3, respectively. Under this system the state 11111 is essentially full health whilst the state 33333 involves severe restrictions on all dimensions. Whilst the EuroQol classification generates 243 theoretically possible health states, direct valuations were obtained for only 43 states which were subsequently used to interpolate values for the remainder (see Dolan *et al.*, 1996).

The two principal measurement methods used were the Visual Analogue Scale (VAS) and the Time Trade Off (TTO). The VAS exercise required

*Author for correspondence. Department of Epidemiology and Public Health, School of Health Sciences, the Medical School, University of Newcastle, Newcastle upon Tyne NE2 4HH, U.K.

respondents to indicate where on a scale with end-points of 100 (best imaginable health state) and 0 (worst imaginable health state) they would rate each of the states. For states rated as better than dead, the TTO exercise required respondents to select the length of time in the 11111 state that they regarded as equivalent to 10 years in the target state (H); the shorter the equivalent length of time, the worse the health state. In the case of states rated as worse than dead, the choice was between dying immediately and spending a length of time (X) in the target state followed by $(10 - X)$ years in full health: the more time required in the 11111 state to compensate for a shorter time in the target state, the worse the target state. Both methods are described in detail below. A total of 3395 interviews were carried out. For full details of the main study see Gudex *et al.* (1996).

FINDINGS TO BE INVESTIGATED FURTHER

A variety of relationships between values from the VAS and those from the TTO has been reported in the literature. For example, Torrance *et al.* (1982), Bombardier *et al.* (1982), and Read *et al.* (1984) report TTO scores which are all higher than VAS scores. Conversely, Churchill *et al.* (1987) report TTO scores which are all lower than VAS scores, whilst Richardson *et al.* (1989) report TTO scores which are higher than VAS scores for milder states but lower than VAS scores for the more severe states. This latter pattern is the one found in the data from the MVH study.

Comparisons of this nature are inevitably problematic as the methods used vary across studies as well as the mechanisms for transforming "raw" scores into valuation indices.* Indeed, there is a

*In the MVH study raw VAS scores were adjusted in order to set 11111 equal to 1 and "immediate death" equal to 0. For states that are rated better than dead in the TTO the method implicitly assigns a score of 1 to full health and 0 to death. The score (assuming no discounting) is therefore simply the number of years (X) in state 11111 the respondent considers equivalent to 10 years in the target state divided by 10. For states that are rated worse than dead, the score is given by the following formula: $T(h) = -X/10 - X$. Where $T(h)$ = the TTO score for state h and X = the number of years spent in good health. However, as this produces scores with a lower bound of -39 ($X = 9.75$), these valuations were transformed by a process that sets this lower bound at -1 , so that the overall scores work within a range that has an equal distance from death in both the positive and negative directions. This transformation makes the worse than dead score simply $-X/10$.

†Citing the psychometric scaling literature, in which the difference between a category scale and a magnitude scale can be described by a concave power function. Torrance (1976) postulates that VAS and TTO scores can be related to one another in such a manner. However, the results from other studies, and certainly the MVH study, suggest a different relationship.

considerable amount of controversy in the literature regarding this latter issue, particularly in relation to states rated as worse than dead. Clearly, the particular transformations used will affect the cardinal properties of the indices generated and hence the nature of the relationship between methods. However, the findings of the MVH study which are of particular interest here may be investigated at the ordinal level.

Whilst it is difficult to point to any theoretical link between the scores generated by VAS and those generated by TTO at the cardinal level,† we would expect the two methods to produce the same ordinal index. For example, we would expect that a state rated as better than dead in one exercise would also be rated as better than dead in the other. In the MVH study however 83.7% of respondents rated at least one state worse than death in the TTO, whilst rating it better than dead in the VAS. Of course, such violations of ordinality are entirely independent of the way in which VAS and TTO scores are subsequently transformed.

Differences in preference were also occurring at the other end of the valuation space, around the anchor of full health. For the less severe states, many respondents were unwilling to trade off any length of life at all in order to avoid a health state which they placed below 11111 on the VAS. All together 52.9% of respondents refused to trade off any time at all for at least one state that they had placed below 11111 in the VAS.

With regard to differences in valuations *within* methods, the main finding to emerge from the MVH study was that the TTO values for the severe health states were lower for older respondents than for the younger ones (see Dolan *et al.*, 1996). For 20 of the 43 health states the TTO values of the 60+ age group were significantly (Mann-Whitney $P < 0.01$) lower than those of the under 60 s. This indicates that respondents in the 60+ group were trading off more years to avoid a state rated better than dead and/or needed more years in the 11111 state in order to compensate them for time spent in a state rated worse than dead. As with the findings discussed above, this observation will clearly be robust to alternative transformations of the raw scores. The health states valued lower by the 60+ age group were predominantly the more severe states and are listed in Table 6 in Appendix B.

Thus, three key questions seemed to emerge from the MVH study:

1. Why are some states rated better than dead on the VAS often rated as worse than dead in TTO?

2. Why are some respondents unwilling to trade off any time at all in order to avoid a health state that they place below 11111 on VAS?
3. Why are TTO valuations of older respondents for the more severe health states lower than those of the younger age groups?

HYPOTHESES GENERATED

Each of the three findings listed above generated a number of alternative hypotheses by way of explanation. Regarding the first, one possible interpretation is that respondents were simply making some sort of "mistake" in one or other of the exercises. Another possibility is that the nature of the TTO exercise focuses the respondent's attention on the *duration* of the health state to a greater extent than the VAS does. A better than dead score on the VAS may be indicating that there is *some* period of time respondents would be prepared to spend in a health state they would not tolerate for 10 years. Another feature of the health state scenario which may be more salient in the TTO is the certainty of death at the end of the 10 years. Thus, a worse than dead score in this exercise may encompass an element of "dread" which is not being picked up in the VAS. If this were the case then we might expect fewer worse than dead TTO responses were the health state to last for a lifetime.

One possible explanation of the second finding is that respondents were simply refusing "to play the game" in the TTO and that these were protest responses. Again it was hypothesised that the somewhat artificial nature of the "10 years followed by death" scenario may have been responsible for certain such protest responses, if indeed that is what they were. Another possibility is that respondents were giving "no time at all" responses to approximate some very short, but nevertheless positive, length of time they would be prepared to give up in order to avoid the health state. If this were the case then we might expect fewer such responses if more emphasis were placed on the possibility of them trading off even very short periods of time.

A number of explanations were considered with regard to the differential effect age appeared to have on TTO valuations. Broadly speaking, it could be that the valuations of older respondents are *genuinely* lower than those of the younger age groups or some *artefact* was responsible for the pattern uncovered. One possible genuine reason for the lower valuations of the 60+ age group may be that

they are more acutely aware of the fact that they may become a burden to their family and they wish to take this "external effect" into account. It is hypothesised that one source of potential artefact may be that many older respondents did not expect to live for another 10 years and hence were giving up years they did not expect to reach anyway. Whilst such explanations are discussed in Dolan *et al.* (1996), without qualitative data it is difficult to discriminate between them.

Against this background it is clear that a better understanding of the cognitive processes at work when respondents complete VAS and TTO exercises is required. To this end a subsample of respondents underwent a (tape recorded) re-interview.

METHODS

Eighty-three respondents, residing in the Northeast of England, who had taken part in the MVH study and had indicated a willingness to be re-interviewed, were selected for this follow-up study. One of the 83 had died and 10 had moved leaving 72 respondents available for follow-up. Forty-five interviews were achieved, giving a response rate of 62.5%.* The age-sex breakdown of these 45 respondents is shown in Table 1. Of these 45 interviews, two were aborted as the respondents could not understand what was required of them. One was a female in the 60+ age group, the other a male in the 18-39 age group.

The interviews were carried out by one of the research team (AR) and as far as possible the structure of the interview matched the one in the MVH study. However, as the main purpose of this study was qualitative, rather than quantitative, each respondent was shown only a subset of the 15 states they had been asked to consider in the first interview. Each respondent was first asked to describe their own health using the EuroQol descriptive system. They were then asked to rank a set of seven health states including death and 11111. The other five health states were a subsample of the set of the respondents had seen in the first interview and were chosen in order to maximise the likelihood of reproducing the disparities outlined above. Each respondent was given the state which they had given the highest score to in the VAS (apart from 11111), two states with TTO scores on either side of 0 and the two "core" (i.e. valued by all respondents) states, 33333 and unconscious.

Table 1. Breakdown of respondents by age and sex

	18-39	40-59	60+	Total
Male	8	7	5	20
Female	8	7	10	25
Total	16	14	15	45

*The remaining respondents said that they no longer wished to be involved in the study. The length of time which had elapsed (over six months) since the first interview and the fact that they were not contacted by the interviewer they had met previously may explain this.

It was explained that each state was to be regarded as lasting for 10 years without any change, followed by death. Respondents were then asked to indicate where on a VAS with endpoints of 100 (best imaginable health state) and 0 (worst imaginable health state) they would rate each of the states. They were first asked to place the card they had ranked as best on the scale and then to do likewise with the card they had ranked worst. They were then asked if they felt any of the remaining cards would come roughly half way between the best and the worst and, if so, to locate this state on the scale. They were asked to locate the remaining health states somewhere on the scale such that states they felt were almost the same would have scores that were close together whilst those they felt were very different would have scores which were further apart. They were told that ties were allowed and that they could change the order of the health states from the ranking exercise if they so wished. They then rated their own health on a similar VAS.

Five health states (7–11111 and “immediate death” which are used as anchors) were then valued by the TTO method using a specially designed double-sided board. One side was relevant for states which were regarded as being better than dead, and the other side for states that were regarded as worse than dead. First, respondents were asked whether they preferred 10 years in a particular health state (*H*) to immediate death.

For states which were preferred to immediate death respondents were led by an iterative process to select the length of time (*X*) in the 11111 state that they regarded as equivalent to 10 years in *H*. Respondents were first asked to consider whether they would prefer five years in 11111 to 10 years in *H*. Whenever respondents indicated that they preferred the five years in 11111 (or the 10 years in *H*) the question was repeated with the number of years in 11111 being reduced (or increased) by one year. Whenever respondents switched from one option to the other in two adjacent years they were asked whether they preferred “something and six months” in 11111 to 10 years in *H*.

It was made clear to respondents that they could trade off even just a few days if they wished to do so. In the case of states worse than dead, the choice was between dying immediately and spending a length of time (*X*) in the target state followed by (10 – *X*) years in the 11111 state. The same iterative process was used as for states rated better than dead.

Whilst the procedures described above replicate those used in the MVH study, albeit with fewer health states, several additional features were built into the interview protocols used in this study. Respondents were asked to “think aloud” as they completed the ranking and scaling exercises. If they decided that one health state was better (or worse) than another the instructions asked them to say

what it was about that state which made it better (or worse) for them personally. If they positioned one state roughly halfway between another two states on the VAS they were asked *what* it was that made it about halfway for them personally.

In addition, respondents were also asked to explain why they made certain decisions during the TTO exercise. It was made clear to respondents that everyone was to be routinely asked the same questions and they did not mean that their preceding answer was wrong or unusual in any way. Respondents were asked “Do you think you could try to tell me why you chose life A (so many years in 11111) or life B (10 years in the target state) just now?” at the following key points in the TTO:

1. where they made the initial worse than/better than death decision;
2. at the “five years” point—i.e. the decision immediately following the better/worse than death one; and
3. when (if) no time was traded off at all.

After completing the ranking, VAS and TTO exercises respondents were asked more specific questions about their responses. Those respondents who had rated the same state as better than dead on the VAS but worse than dead in the TTO were asked whether or not the position of that state on the VAS meant that they personally preferred spending 10 years in that state to immediate death. They were then asked whether or not they would stick with their worse than dead TTO decision if the state lasted for a lifetime, rather than for 10 years. In addition, such respondents were asked specifically whether they took more notice of the 10 year time span in one or other of the exercises. Those respondents who had not traded any time at all in order to avoid a health state which they had below state 11111 on the VAS were asked whether or not that state was as good as full health for them personally. They were then asked whether they would be willing to trade off any time if the state lasted for a lifetime, rather than for 10 years. The interview protocol used in the study is available from the authors on request.

Such a highly structured protocol was developed in order to provide more systematic evidence than it was considered the “think aloud” exercise would generate. However, a verbal protocol analysis was carried out on all the taped material and responses were classified into broad groupings. Whilst the classification system was largely based upon *a priori* expectations, certain categories emerged from the data itself (see below).

RESULTS

Within-subject disparities in rating a state better than/worse than dead

Twenty-nine of the 43 respondents (67.4%) ranked and scored at least one state above death on the VAS but subsequently rated that state as worse than dead in the TTO. One possible interpretation is that these respondents simply made the "wrong" choice at the initial worse/better than dead decision node in the TTO. However, respondents were asked to explain their decision at this point and all 29 confirmed that they would rather die immediately than spend 10 years in the health state in question. When asked whether their VAS answer meant that they personally preferred 10 years in that health state to immediate death, 12 of the 29 said that it did, and 14 said that it did not, whilst three did not know.

It did appear as if certain respondents were ignoring the duration of the state when completing their VAS as the following comment indicates:

Yes it is better than dead, you are still alive, but not for 10 years.

Thirteen of the 29 with a different ordering of dead between the VAS and TTO said that the 10 year time scale had been more salient in the TTO than in the VAS. Further, less than a third of respondents mentioned time at all whilst explaining their VAS orderings, apparently providing further evidence that duration was not a prominent factor in this exercise.

Whilst this appears to offer a partial explanation for the disparity, it also emerged that certain respondents were interpreting their VAS responses rather differently than had been expected. In order to gain a clearer picture of what was going on during the VAS, it is necessary to look more closely at how respondents treated the "immediate death" state in this exercise. Eighteen of the 29 had death at the bottom of the scale. Of this subgroup, 14 respondents made comments which seemed to imply that for them death "automatically" went to the bottom of the scale. These fell into one of two categories, five making references to death as "having to be" worst or bottom, a typical statement being "death has got to be the worst for everybody doesn't it?", and nine making references to death in the chronological order of events, i.e. being last or final. One comment which made this particularly explicit was the following:

then the last one where you are confined to bed...and then you come to "unconscious" and I suppose death follows doesn't it after that?

Only three of the 29 changed a worse than dead TTO decision to better than dead when the scenario was for the rest of their life, rather than for 10 years. Thus, it would appear that any "dread" concerning the time of death was not a significant fac-

tor in these states receiving worse than dead TTO valuations.

Taking the sample as a whole, two more general issues emerged which appear to shed some light onto how respondents dealt with the two different valuation tasks. First, twelve respondents talked about their VAS values in terms of percentages, offering some support for the finding of Morris and Durand (1989) that VAS scales can be interpreted as percentages of fitness scores. Second, only seven respondents mentioned the effect the health state would have on their family during the Ranking and VAS, whilst 27 did so during the TTO, appearing to indicate that the TTO exercise brought about more considered responses. Consider the following comments:

The (TTO) board just made it more stark—I had not thought about it as much on the (VAS) scale

On the (VAS) scale, frankly "unconscious" wasn't a problem, I simply wouldn't be there. The (TTO) board was a more emotional decision.

Respondents unwilling to trade off any time at all to avoid a health state

Fifteen of the 43 respondents (34.8%) refused to trade off even a few days or weeks in order to avoid a health state which they had placed below 11111 on the VAS. All 15 confirmed that their VAS response *did* mean that they considered 10 years in that state to be worse than 10 years in 11111. They did not, however, seem to translate this into a willingness to trade off time to avoid that state. Only one of the 15 did not trade off any time at all *throughout* the TTO exercise, apparently objecting to the task on religious grounds. Thus, it does not appear that refusing to "play the game" in the TTO would, in itself, account for this disparity. Rather, the predominant message was that, as long as they could cope with the state in question, they would not consider giving up any of the 10 years to avoid it. Consider two examples of comments made by such respondents:

This (11112) can be controlled...I mean it's not a show stopper. It has to be less than AP (11111) but only slightly less...it's not worth giving up time for

Of course I would rather have AP (11111), anybody would, but I could cope with this (11121). If I could have AP I would have it but I can put up with this—I would not lose my life because of a bit pain.

Only two of the 15 respondents (both under 40) who were unwilling to give up any of the 10 years did indicate that they would be willing to give up some time *at the end of their life* instead of at the end of 10 years. Again it would appear that it is *not* problems associated with the duration of the health state which is driving the results.

Table 2. Time trade off scores for "core" states by age

		18-39	40-59	60+
Unconscious 33333	Median	-0.10	-0.50	-0.75
	33333 Median	-0.38	-0.58	-0.65

TTO valuations and age

As in the main study, respondents in the 60+ age group gave lower TTO valuations to the severe states than the other age groups: Table 2 gives the scores* for the two core states unconscious and 33333. One possible real explanation for the lower TTO scores for severe health states is that older respondents are more worried about becoming a burden to their families. Table 3 shows the number of respondents in each age group mentioning this in the TTO and that, although a large number of respondents mentioned this as a factor, there is no greater tendency for older respondents to mention becoming a burden than younger respondents.

Whilst the life expectancy argument seemed to offer an artefact explanation for this, no respondent said directly that they did not expect to live for 10 years. Table 4 shows that there were five cases out of a possible 40 where respondents switched a "for 10 years" worse than dead decision to a "for rest of life" better than dead decision. In only two of these five cases (both involving respondents aged under 60) was age or life expectancy used as an explanation of the "for rest of your life" choice.

Whilst there was no strong evidence to support an artefact explanation on the grounds of life expectancy, it did emerge that older respondents are less likely than younger ones to find the worse than dead scenario plausible. Whilst no respondent in the 18-39 age group questioned whether they would return to full health after a number of years in the target health state, half of the respondents in the 60+ age group said that they thought this was impossible. Consider one comment from a 78-year-old:

This just does not happen, you have more problems as you get older, you do not expect to get better

or another from an 80-year-old:

After this (33333) you could not possibly get better... you would never get to the pink state (11111).

Performing a chi-squared test on the results in Table 5 shows that finding the scenario implausible

*For reasons of comparability with the results from the main study, raw TTO scores have been transformed in the manner outlined in the first footnote. The transformation, for example, indicates that, on average, respondents in the 18-39 age group considered nine years in the unconscious state followed by one year in the 11111 state to be equivalent to dying immediately, whereas those in the 60+ age group considered 2.5 years in the unconscious state followed by 7.5 years in full health equivalent to dying immediately.

Table 3. Number of respondents who mentioned being a burden in TTO by age

	18-39	40-59	60+	Total
Yes	11	8	8	27
No	4	6	6	16
Total	15	14	14	

is not independent of age ($P < 0.05$). Of those respondents saying the scenario was implausible the median response for state 33333 of those 10 respondents saying the scenario was implausible is -0.71 as opposed to -0.36 for the remainder, indicating that this particular artefact may have a considerable downward bias on valuations.

DISCUSSION

The results presented in this paper suggest that the finding in the MVH study, that more states are considered to be worse than dead in the TTO method than on the VAS, is a robust one. With respect to either spending 10 years in a health state or dying immediately we would expect the two assessment methods to "reveal" the same ordering of preferences. However, Behavioural Decision Theory suggests that preferences are often constructed, rather than revealed, and that simplifying strategies are often adopted in this construction process (see Payne and Bettman, 1992, for an overview).

One way in which a complex choice can be made to appear simpler is by the "editing out" of information common to all alternatives. As respondents were required to consider seven alternatives at once (15 in the main study) during the VAS, this may help explain why the duration of the health states, which was constant at 10 years, did not appear to play a prominent role in this exercise. Such insensitivity to duration was also uncovered in Jones-Lee *et al.* (1995) where VAS scores did not appear to differentiate adequately between temporary states and those involving permanent disability. In addition, it appears that respondents in this study took into consideration a wider range of issues (i.e. their family circumstances) in the TTO exercise than they did in the VAS.

There may also be explanations rooted in Behavioural Decision Theory for those respondents who were unwilling to trade off any time at all in order to avoid a health state they clearly stated they

Table 4. Respondents switching from worse than to better than dead when TTO scenario was for rest of life

	18-39	40-59	60+	Total
Yes	1	2	2	5
No	14	10	11	35
Total	15	12	13	

Table 5. Number of respondents who said that the WD TTO scenario was implausible by age

	18-39	40-59	60+	Total
Yes	0	3	7	10
No	15	11	7	33
Total	15	14	14	43

did not think was as good as full health. In the TTO scenario respondents are given 10 years in the target health state and asked if they will give up any of this time in order to achieve state 11111. Thus, they are asked to weigh up the prospect of a loss (in terms of life expectancy) in order to receive a gain (in terms of improved quality of life). In Prospect Theory, losses are weighted more heavily than gains so an asymmetry arises in the value function around the reference point (see Kahneman and Tversky, 1979). This would result in a disproportionately large gain in improved health status being required as compensation for the loss of life expectancy. This may explain why the health state had to be below some "tolerance level" before they would be willing to give up even a few days to avoid it.

Such a "threshold of tolerability" effect has also been found in studies using the Standard Gamble (SG) technique where respondents are often unwilling to accept any risk of death whatsoever in order to avoid the certainty of a mild health state (see Jones-Lee *et al.*, 1995). Whilst the SG is fundamentally different technique from the TTO, the two methods are similar in one important respect: they both require respondents to sacrifice something else to which they attach value in order to return to full health. In contrast respondents are required to make no sacrifice whatsoever in the VAS exercise.

Although the possibility had not been considered that some respondents may find the worse than

dead TTO scenario implausible prior to conducting the present study, it is not difficult to see why this may be the case. For certain states the worse than dead scenario asks respondents to believe that they will be confined to bed for five years after which time they will return to full health. Whilst younger respondents may be able to imagine a lengthy period of illness from which they will eventually recover fully, this may be much harder for older respondents. Even more than this however, the scenario is asking some older respondents, who are in less than full health now, to believe that there will be an *improvement* in their health status following the period of illness.

Interestingly, of the 10 respondents who said they found the worse than dead TTO scenario implausible, all but one *were* willing to offer some response at this stage. Thus, it seems likely that these respondents are somehow re-interpreting the scenario in order to enable them to provide a response. A similar phenomenon has been uncovered in Contingent Valuation (CV) studies where respondents often report answering a different question than the one posed (i.e. see Fischhoff *et al.*, 1993). Wherever such disparities exist, the resulting valuations will clearly be subject to some degree of bias.

One way to avoid this problem is to place the years in good health *before* the years in the target health state in the worse than dead scenario, as formulated by Torrance (1986). The potential problem with this worse than dead scenario is that respondents might adopt an "I'll have the good years and then jump in front of a train" attitude. This would result in biasing the valuations of worse than dead states upwards. It would seem that valuations derived using either of the two formulations of the worse than dead scenario are, at least potentially, subject to the influence of artefact.

CONCLUSION

The broad pattern of responses uncovered in the main study have more or less been replicated here. Thus, at least some evidence has been uncovered on each of the three key issues this study set out to examine. Regarding the first question, the majority of respondents who rated a state better than dead on VAS but worse than dead on TTO indicated that it was their VAS response which did not truly reflect their preference for spending 10 years in that state compared with dying immediately. A number of factors arose which each seem to offer a partial explanation for this. A minority of respondents were clearly interpreting their VAS as a sort of percentage of functioning scale. In addition, respondents did not appear to consider either the duration of the state or its likely affect on them and their family. That a number of health states are being

Table 6. Median scores of 60+ group different from 18-59 group

	18-59	60+	BD scores	% WD	WD scores
13311	0.53	0.38		✓	
12223	0.38	0.25			✓
32211	0.30	0.00		✓	✓
23321	0.38	0.03		✓	
22323	0.13	-0.03			✓
33212	0.11	-0.40		✓	✓
32313	-0.08	-0.43		✓	✓
21133	0.00	-0.28		✓	✓
23232	0.00	-0.38		✓	✓
23313	0.00	-0.28		✓	✓
33321	-0.13	-0.38		✓	
22233	-0.19	-0.36			
32223	-0.18	-0.48		✓	✓
32232	-0.30	-0.50		✓	✓
13332	-0.28	-0.48		✓	✓
32331	-0.30	-0.58		✓	✓
33232	-0.38	-0.60		✓	✓
33323	-0.38	-0.73		✓	✓
UN	-0.30	-0.59		✓	
33333	-0.63	-0.76			✓

assessed simultaneously during this exercise offers a plausible explanation of this finding.

The evidence from this study suggests that "no trade off at all" responses were not generally being used to approximate some short length of time respondents would be willing to give up to avoid the health state. Rather, there is evidence of the existence of a "threshold of tolerability" below which states would have to fall before some respondents would be willing to give up even a few days.

There was no evidence uncovered here to suggest that the observation that older respondents give lower TTO scores to severe health states is *predominantly* artefact. Whilst older respondents do appear to be less willing to accept the worse than dead TTO scenario at face value, they also rate a state as worse than dead more often than younger respondents. This appears to be a genuine reflection of the fact that older respondents are less prepared to live for the next 10 years in a diminished health state.

This study has offered some tentative explanations of what might lie behind some of the findings uncovered in the MVH study. Whilst the main focus of this study has been to examine the cognitive processes which underlie the VAS and TTO methods, many of the findings will be of relevance to other value elicitation procedures.

Acknowledgements—The support of the ESRC (award number R000 234987) and the Department of Health is gratefully acknowledged. The authors would like to thank Professor Graham Loomes for his considerable input into the study design and for his comments on earlier drafts of this paper.

REFERENCES

- Bombardier, C., Wolfson, A. D., Sinclair, A. J. and McGreer, A. (1982) Comparison of three preference measurement methodologies in the evaluation of a functional status index. In *Choices in Health Care: Decision Making and Evaluation of Effectiveness*, eds R. Deber and G. Thompson. University of Toronto, Toronto.
- Churchill, D. N., Torrance, G. W., Taylor, D. W., Barnes, C. C., Ludwin, D., Shimizu, A. and Smith, E. K. M. (1987) Measurement of quality of life in end stage renal disease: the time trade-off approach. *Clinical and Investigative Medicine* 10(1), 14–20.
- Dolan, P., Gudex, C., Kind, P. and Williams, A. (1996) The time trade-off method: results from a general population study. *Health Economics* 5, 141–154.
- EuroQol Group (1990) EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 16, 199–208.
- Fischhoff, B., Quadrel, M. J., Kamlet, M., Loewenstein, G., Dawes, R., Fishbeck, P., Klepper, S., Lelend, J. and Stroh, P. (1993) Embedding effects: stimulus representation and response mode. *Journal of Risk and Uncertainty* 6(3), 211–234.
- Gudex, C., Dolan, P., Kind, P. and Williams, A. (1996) Valuing health states: interviews with the general public. *European Journal of Public Health*.
- Jones-Lee, M. W., Loomes, G. and Philips, P. (1995) *Valuing the Prevention of Non-fatal Injuries: Contingent Valuation vs Standard Gambles*. Oxford Economic Papers, Oxford.
- Kahneman, D. and Tversky, A. (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–281.
- Llewellyn-Thomas, H., Sutherland, H. J., Tibshirani, R., Ciampi, A., Till, J. E. and Boyd, N. F. (1984) The measurement of patient's values in medicine. *Medical Decision Making* 2, 449–462.
- Morris, J. and Durand, A. (1989) Category rating methods: numerical and verbal scales. Mimeo, Centre for Health Economics, University of York.
- Payne, J. W. and Bettman, J. R. (1992) Behavioral decision research: a constructive processing perspective. *Annual Review of Psychology* 43, 87–131.
- Read, J. L., Quinn, R. J., Berwick, D. M., Fineberg, H. V. and Weinstein, M. C. (1984) Preferences for health outcomes: comparison of assessment methods. *Medical Decision Making* 4(3), 1984.
- Richardson, J., Hall, J. and Salkeld, G. (1989) Cost utility analysis: the comparability of measurement techniques and the measurement of utility through time. In *Economics and Health 1989 Proceedings of the Eleventh Australian Conference of Health Economists*, ed. S. C. Smith. Public Sector Management Institute, Clayton.
- Torrance, G. W. (1976) Social preferences for health states: an empirical evaluation of three measurement techniques. *Socio-Economic Planning Sciences* 10, 129–136.
- Torrance, G. W. (1986) Measurement of health state utilities for economic appraisal: a review. *Journal of Health Economics* 5, 1–30.
- Torrance, G. W., Boyle, M. H. and Horwood, S. P. (1982) Application of multi-attribute theory to measure social preferences for health states. *Operations Research* 30(6), 1982.

APPENDIX A

EuroQol Classification System

<u>Mobility</u>	
	No problems in walking about
	Some problems in walking about
	Confined to bed
<u>Self-Care</u>	
	No problems with self care
	Some problems with washing or dressing self
	Unable to wash or dress self
<u>Usual Activities</u>	
	No problems with performing usual activities
	Some problems with performing usual activities
	Unable to perform usual activities
<u>Pain and Discomfort</u>	
	No pain or discomfort
	Moderate pain or discomfort
	Extreme pain or discomfort
<u>Anxiety and Depression</u>	
	Not anxious or depressed
	Moderately anxious or depressed
	Extremely anxious or depressed

Core	Very mild	Mild	Moderate	Severe
11111	11112	12211	13212	33232
33333	11121	11133	32331	23232
Unconscious	11211	22121	13311	23321
Death	12111	12121	22122	13332
	21111	22112	12222	22233
		11122	21323	22323
		11312	32211	32223
		21312	12223	32232
		21222	22331	33321
		21133	21232	33323
		11113	32313	23313
		11131	22222	33212

APPENDIX B

Re-examination Of The Age Disparity Of TTO Scores In Main Data Set

In light of the evidence that the worse than dead TTO scores of the elderly may be contaminated by some artefact, those 20 states in the main study with significantly lower scores for this age group than for the 18–59 age group were re-examined. For these states some combination of the following factors must be at play:

- (a) states received lower better than dead valuations from the 60+ age group;
- (b) states were rated as being worse than dead more often by the 60+ age group; and
- (c) states received lower worse than dead valuations from the 60+ age group.

Whilst the implausibility argument may offer a partial explanation for (c), there is no evidence in this study to suggest there is anything other than a real explanation for (a) (had life expectancy proved an important factor we would have expected this to contaminate better than dead scores also). If older respondents are more likely to get the better than/worse than dead decision “wrong”, then there may be an artefact explanation for (b). If this were true we would expect the proportions of the elderly rating the *very mild states* as worse than dead to be different from those in other age groups. However, for only one of the five very mild states was this difference significant: 4% of the over 60 s rated state 12111 as worse than dead as compared to 1.6% of under 60 s.

Table 6 shows that for only four of the 20 states with significantly lower 60+ valuations were the worse than dead scores the sole factor accounting for the differences. Further, there is no reason to assume that the *entire* difference in the worse than dead scores of the two age groups is due to artefact, after all only half the respondents in the 60+ group in this study appeared to find the scenario implausible.