Monograph:
Tsuchiya, A. and Ternent, L. (2013) A note on the expected biases in conventional iterative health state valuation protocols. Discussion Paper. HEDS Discussion Paper (12.11).

HEDS Discussion Paper

No.12.11

A note on the expected biases in conventional iterative health state valuation protocols

Laura Ternent PhD[1], Aki Tsuchiya PhD[2]

1. Institute of Health and Society, Newcastle University, UK

2. School of Health and Related Research, University of Sheffield, UK

Disclaimer:

This series is intended to promote discussion and to provide information about work in progress. The views expressed in this series are those of the authors, and should not be quoted without their permission. Comments are welcome, and should be sent to the corresponding author.

A later, substantially shorter version of this paper, without the examples, is accepted for publication from Medical Decision Making, doi: 10.1177/0272989X12475093

A note on the expected biases in conventional iterative health state valuation protocols

Laura Ternent, PhD, Institute of Health and Society, Newcastle University, UK

Aki Tsuchiya, PhD, School of Health and Related Research, University of Sheffield, UK

Abstract

Background: Typical health state valuation exercises use trade off methods, such as the Time Trade Off or the Standard Gamble, involving a series of iterated questions so that a value for each health state by each individual respondent is elicited. This iterative process is a source of potential biases, but this has not received much attention in the health state valuation literature. The issue has been researched widely in the contingent valuation (CV) literature which elicits the monetary value of hypothetical outcomes. **Methods:** The lessons learnt in the CV literature are revisited in the context of the design and administration of health state valuations. The paper introduces the main known biases in the CV literature, and then examines how each might affect conventional iterative health state valuations. **Results**: Of the eight main types of biases, starting point bias, range bias, and incentive incompatibility bias are found to be potentially relevant. Furthermore, the magnitude and direction of the bases are unlikely to be uniform, and depend on the range of the value (e.g. between 0 and 0.5). **Limitation**: this is an overview paper and the conclusions drawn need to be tested empirically. **Conclusions:** health state valuation studies, like CV studies, are susceptible to a number of possible biases that affect the resulting values. Their magnitude and direction are unlikely to be uniform, and thus empirical studies are needed to diagnose the problem and if necessary to address it.

A note on the expected biases in conventional iterative health state valuation protocols

1. Introduction and background

A large proportion of health state valuation exercises are carried out using trade off methods, such as the Time Trade Off (TTO) or the Standard Gamble (SG) (1). These are methods where respondents are asked to trade between on the one hand a long or certain survival in a problematic health state and on the other a shorter or uncertain survival in full health. Most administrations of TTO and SG are 'iterative' in that they ask a series of questions where the health state is fixed, and the degree by which the prospect in full health is shorter or uncertain is varied. This process follows a pre-determined path, or 'routing', and iterated until the respondent either indicates their indifference between the two prospects, or gives an answer that allows researchers to infer the value at which they would be indifferent. The reason behind these iterative designs is that health state valuation studies have typically aimed to elicitate a value for each health state by each individual respondent.

On the other hand, there is increasing interest in the application of non-iterative methods for health state valuations. Where the health state scenario includes a duration dimension, Discrete Choice Experiments (DCE) can be used to generate health state preferences (2). Although DCE only elicits ordinal information from each respondent, it allows the generation of cardinal health state values at the aggregate level, by assuming that the strength of preference is associated with the proportion of people who chose an option over another. DCE tasks are typically presented to respondents in a random order. In other words, each choice task is independent from the previous one, and there is no systematic routing through them.

With the notable exception of Brazier and Dolan (2005) (3) for SG, expected biases caused by systematic routing have not received much attention in the health state valuation literature. However, the issue has been researched in the contingent valuation (CV) literature which elicits the monetary value of hypothetical outcomes. The aim of the CV method is to identify for each individual respondent the maximum amount of money that they would pay in exchange for the hypothetical good and still be as well off as the original situation without the good (4).

The aim of this note is to explore the extent to which the lessons learnt in the CV literature can be transferred to the design and administration of health state valuation studies. The paper is organised as follows. Section 2 will present the main known biases in the CV literature, and section 3 will examine how conventional iterative health state valuation is susceptible to each. Throughout, the valuation protocol recommended by the Measurement and Valuation of Health (MVH) study is used as a reference (5). Section 4 will conclude.

2. Review of the CV literature and expected biases

This section will begin with a brief summary of the different elicitation methods used in CV, and then provide an overview of the biases found in the CV literature.

3.1 Different methods of CV

There are a number of different methods to elicit willingness to pay. For example, the 'open ended' method or the 'payment card' method can be used. These methods are often referred to as direct methods of elicitation. Alternatively, the 'single bounded dichotomous choice method' (SBDC), the 'double bounded dichotomous choice' (DBDC) method, or the 'bidding game' (BG) method can be used. These methods are often referred to as closed ended methods of elicitation.

Using the open ended method, respondents are asked to state their maximum willingness to pay for the good in question. In the payment card method, respondents are presented with a range of values, from which they are asked to pick their maximum willingness to pay. Using the SBDC method respondents are asked a binary yes/no question: e.g. are you willing to pay £10 – 'Yes' or 'No'? No further follow-up questions are asked. The DBDC method is similar to the SBDC method in that it has an initial binary yes/no question ($DBDC_1$), which then leads to a follow-up question ($DBDC_2$). The bid offered at $DBDC_2$ depends on the response to $DBDC_1$. If the respondent answers 'yes' to $DBDC_1$, then $DBDC_2$ is a higher, predetermined amount. If the respondent answers 'no' to $DBDC_1$, then $DBDC_2$ will be a lower predetermined amount. The BG method is an extension of the DBDC method and uses a predetermined algorithm to bid the respondent up or down, conditional upon responses to prompted values (6). Typically, four or five follow up questions are asked, unless they switch from a 'yes' to a 'no', or vice versa. If a respondent says 'yes' or 'no' throughout, then an additional open ended question can be asked, where respondents are asked to state their maximum willingness to pay.

It is known that different methods of elicitation can lead to different types of bias (7). For example the SBDC, DBDC and BG methods may suffer from starting point bias; and payment card method may suffer from range bias. Whilst the use of direct questions such as the open ended method may have advantages in terms of not being prone to some of these biases found in closed ended methods, they may be more likely to suffer from hypothetical bias. Each is explained below.

3.2 The main biases in the CV literature

Let us begin from the more general biases relevant to stated preference designs, to the more specific.

'Hypothetical bias' is the difference between hypothetical (stated) willingness to pay and actual willingness to pay derived from a real market setting. If hypothetical bias exists stated willingness to pay will not be a good indication of actual willingness to pay (8). Evidence suggests that stated willingness to pay overestimates actual willingness to pay (9, 10). Methods have been employed in the literature to mitigate the effects of hypothetical bias, such as 'cheap talk', where the potential for hypothetical bias is discussed with respondents prior to eliciting their willingness to pay (11, 12).

'Scenario misspecification bias' is where the scenario presented to respondents is misunderstood (13). 'Payment vehicle bias' occurs when different payment vehicles, such as out of pocket payment, payment via insurance premium, payment via taxation, influence willingness to pay values. There is limited evidence of payment vehicle bias from health economics: for example, evidence of payment vehicle bias when using the DCE method has been found (14). 'Payment mode bias' is where the mode of payment, such as simply stating raw willingness to pay or willingness to pay as a percentage of income can influence final willingness to pay values.

'Question order bias' can arise when more than one good is valued in a CV survey. It exists when the value of a good is dependent on the order in which that good was presented to respondents in the CV questionnaire, where the first good valued in a sequence of goods tends to be given the highest willingness to pay. There is limited and mixed evidence on question ordering effects in health economics using contingent valuation methods (15). Possible explanations for question order bias include income and substitution effects, survey design, and administration issues (16). In addition, Stewart et al (2002) (15) suggested 'fading glow' where the first programme in the sequence of programmes valued captures the majority of the utility associated with giving.

'Range bias' exists when willingness to pay is influenced by the range of values presented to respondents. This is a particular issue when using the payment card method, where respondents can see the entire range of values to be used. This has been found in previous health economics research (17). Whynes et al (2004) tested for range bias within a CV survey by using two different payment scales in two separate samples of respondents. One sample received a payment card with a maximum willingness to pay value of £100 (short scale) and the other with a maximum willingness to pay value of £1000 (long scale). The authors found that after controlling for differences between samples, the long scale payment card resulted in a 30% increase in mean WTP values.

'Yea saying' occurs if a respondent replies 'yes' when they are asked whether they are willing to pay an amount regardless of whether they are actually willing to pay it (18, 19). Yea saying could occur in all closed ended methods of eliciting willingness to pay and would lead to artificially high estimates of mean willingness to pay.

'Starting point bias' exists when individuals anchor their willingness to pay to the initial bid used to begin one of the closed ended methods of eliciting willingness to pay (such as the BG or DBDC methods). In this instance, individuals who do not have well defined preferences for the good or service, and therefore do not a have a clear idea of their maximum willingness to pay, may think that the initial bid is a clue to the value of the good or service in question (20). Starting point bias will result in a positive correlation between the starting bid and the final monetary amount.

Starting point bias is a well documented problem in the CV literature. For example it has been found in the DBDC method in environmental economics and in health economics (19, 21). Starting point bias has also been found extensively in the BG method (6, 7, 22-24).

'Incentive incompatibility' has been found in iterative elicitation methods where the iterative nature of the elicitation method may provide respondents with the incentive to behave in a strategic fashion in the follow-up questions, such that responses are not a true measure of willingness to pay (6). For example, in a DBDC, if on the one hand a respondent answers 'yes' to an initial bid and is then offered a higher bid, this would not make sense in a real-world market. The respondent may therefore reject the second higher bid even if they are willing to pay the amount, if they expect it to become even higher should they say 'yes'. (And, if this was a BG, the bidding will stop when they switch from 'yes' to 'no'.) If on the other hand a respondent answers 'no' to an initial bid and is then offered a lower bid, the respondent may reject the second lower bid even if they were willing to pay the amount, if they expect it to become even lower should they say 'no'. (If this was a BG, the bidding will continue, since they do not switch from 'no' to 'yes'.) These examples are said to be incentive incompatible, because the final response is biased by strategic behaviour of respondents caused by the iterative nature of the exercise (25). Both will result in the response to a second bid being 'no', where they may have been 'yes' had it been asked as a first bid. Incentive incompatibility bias will depend on the relative size of the starting point (I), of the second bid (II), and of the actual willingness to pay (W). Overall, there are four possibilities. If $I < W$, then $I < II$. So either $I < II < W$ (where the respondent may strategically say 'no' instead of 'yes' due to incentive incompatibility), or $I < W < II$ (where the respondent says 'no' and means it). If $W < I$, then $II < I$, so either $W < II < I$ (where the respondent says 'no' and means it), or $II < W < I$ (where the respondent may strategically say 'no' instead of 'yes' due to incentive incompatibility). If incentive incompatibility takes full effect, then all replies to a second bid (and onwards) will be 'no'. Note that, while BG with $I < W$ will stop at the second bid, BG with W < I will not, so the extent of the bias may not be symmetric around the starting point. In health, incentive incompatibility has been found in the BG method (McNamee et al., 2010) and the DBDC method (26).

3. Potential biases in iterative TTO

This section will give a very brief outline of the routing for TTO tasks recommended by the MVH study (27), and point out the possible biases at each stage. A more detailed description of the protocol is given in the appendix. Note that this is *not* the protocol which produced the EQ-5D population value set for the UK (28). The reason for using the *recommended* MVH TTO protocol (as opposed to the *actual* MVH TTO protocol used in their study) as the reference in this exercise is because this design has been adopted in numerous subsequent studies that estimated country specific value sets for EQ-5D (29), and in a growing number of studies that generated value sets for condition specific preference based instruments (30). However, most of the observations in this section are generalisable beyond any particular TTO or SG protocol.

Health state valuations are a type of stated preference methods, requiring respondents to make choices over hypothetical decisions, and thus are susceptible to hypothetical bias. However, unlike CV studies, there is no obvious way of establishing the extent of this bias in health state valuation.

The recommended MVH TTO protocol values a given health state (H) lasting for 10 years, followed by death. This is done by identifying the number of years to live in full health followed by death that is equivalent to it. While people have everyday experiences of transactions involving money, the kind of questions asked in health state valuation exercises is completely idiosyncratic. They may therefore be more susceptible to scenario misspecification bias and payment vehicle bias than already observed in CV studies.

There are three things to note at this stage. First, since CV bids are presented in terms of amounts of money to be paid to have the good (bids), the parallel concept in TTO would be numbers of years to be given up to live in full health. However, most TTO are presented in terms of numbers of years to be lived in full health (let us refer to these as ‘offers’). There are no studies we are aware of that use the corresponding CV framing where offers are presented in terms of remaining disposable income after the payment, and the effect of payment vehicle bias under such framing is unknown. Second, because the TTO does not usually specify how the individual would come to die in each alternative, payment vehicle/mode biases are unlikely to be relevant. Third, as long as the respondent is asked *which* of the two possible lives they prefer, yea saying is likely to be avoided. However, yea saying may be relevant if the respondent is asked *whether* they would give up years of life in exchange for better health.

Health state valuation studies typically ask respondents to value a number of states, which makes them susceptible to question order bias. It has been observed that the TTO value of a health state is affected by the number of health states preceding it in the valuation exercise (31). The typical way of overcoming question order bias is to randomise the order in which different health states are valued.

The MVH TTO protocol takes three steps for the very first health state valued. Let us refer to the different numbers of years in full health that is varied in the TTO exercise in terms of the health state values they imply, and call them ‘offers’. In the first step, the offer implies that the value of the state is 1.0. Unless the respondent feels that state H is indistinguishable from full health, they will take the offer. In the second step, the offer implies a health state value of 0.0. The answer will depend on whether the respondent thinks state H is better or worse than being dead. If H is better than being dead, then the respondent will reject the offer of 0. Alternatively, if H is worse than being dead, then the respondent will take the offer. If there is a starting point bias, then the response at the second step may be affected by the first step where the offer was 1.0. On the other hand, whether or not a health state is better or worse than being dead may be robust enough so that step two would not be affected by starting point bias. Interestingly, the recommended MVH protocol skips the first step for the second health state onwards and begins with an offer of 0.0. So if there is a starting point bias, its manifestation will depend on the position of the health state in the survey.

The use of visual aids in TTO may invite range bias. This is because the TTO ‘board’ visualises the range of values that are available to the respondent. However, since the upper bound for states better than dead indicated by the TTO board (1.0) is the theoretically determined upper bound, the impact this may have on the TTO values is much less than the range bias in the CV, where the choice of upper bound is (in most cases) entirely arbitrary. The lower bound for states better than dead and the upper bound for states worse than dead (both set to 0) may be more problematic, if the boundaries suggest to the respondent that the value needs to stay within the positive or negative range. To some extent, the impact of this depends on whether or not the protocol allows a respondent to change their mind, and to go back and revise their response at the second step.

Range bias introduced at the lower bound of states worse than dead is the most complex. The MVH protocol involves transforming negative values, so that in effect the value of a one-year duration on the TTO board is mapped to an absolute value of 0.1, with the lowest possible value of -1.0. However, this is an artefact of the arbitrary transformation, of which respondents are not informed and its impact is unclear. On the other hand, untransformed negative values using the TTO board range from -39 to 0, but it is not clear whether respondents fully appreciate this, or how they would react if they did. Furthermore, there is evidence to suggest that the recently developed 'lead-time' version of TTO, which reconciles the inconsistency between the visual aid and implied values in the negative range (32), suffers from its own range bias, where the values are affected by the lowest value that the visual aid accommodates (33).

Going back to the MVH routing protocol, the third step for states better than being dead is in effect a BG that begins with an offer at 0.5, and homes in on the indifference value. Similarly if the state is worse than being dead, then the BG offer begins at -0.5 (see Appendix for the exact presentation). Either way, if the respondent takes the offer, then the next value is reduced linearly by 0.1. If the respondent does not take the offer, then the next value is increased linearly by 0.1. This is clearly susceptible to starting point bias. There is some evidence based on a large scale survey of the general public to suggest that when the initial offer in the third step is varied the final TTO values vary in line with it (34). Most TTO protocols including the recommended MVH protocol use an identical starting point and routing for all respondents and all states, so this bias is not controlled for.

Moving on, if the respondent 'switches' between taking or not taking an offer, this means the last linear offer has overshot, so the offer is adjusted back by 0.05. If at any point in the process the respondent is indifferent between taking the offer and not taking the offer, then that completes the process for state H. If indifference is not established after a 0.05 adjustment, the respondent is deemed to be indifferent at the midpoint. For example, they may take offers of 0.5 and 0.4 but not 0.3; they will then be offered 0.35; if they reject it, then it is assumed they are indifferent at 0.375. This involves three features. First, as can be seen the number of iterations it takes per health state depends on the final value. Assuming the first step explained above is skipped, a value of 0.4 will take three iterations, while a similar value of 0.375 will take five. Second, the number of iterations needed is not correlated with the value (e.g. 0.2 takes the same number of iterations as 0.375: five). Third, all values have a fixed pathway leading to it. For example, the only way to achieve 0.375 is by taking the first two higher offers, and then rejecting two lower offers, 0.3 and 0.35. While it is quite possible that these features bias the final TTO values, to the best of our knowledge, there are no known corresponding biases in the CV literature.

Regarding incentive compatibility bias, this arises in CV surveys when respondents wish to secure a bid that is lower than the amount they are willing to pay. Applying the same logic would suggest that the bias would arise in TTO surveys if respondents wish to secure an offer that is higher than what is equivalent to their valuation of the health state, so that TTO values became upward biased. If so, then the bias may be asymmetric and larger for health states that lie above the starting point than below it. However, because nobody has direct experience of markets where a longer survival in poorer health could be exchanged for a shorter survival in better health through incentive compatible procedures, it is not clear how strategically people would behave when faced with an activity involving trade-offs in duration and quality of life. Furthermore, there may be other possible strategic behaviours. If, for example, people believed (rightly or wrongly) that patients would benefit if health problems were exaggerated, then this may lead to strategic behaviour in the opposite direction.

4. Conclusion

This note has presented an overview of CV methods and known biases, and examined the recommended MVH TTO protocol against them. Of the different types of bias examined, conventional iterative TTO is likely to suffer from a few. Since health state valuations use hypothetical health states and use unusual scenarios involving death, they are susceptible to hypothetical, and scenario misspecification biases, but the direction of these are unclear. However, payment vehicle bias is not likely. It is possible that TTO is free of yea saying, because it asks respondents to choose between two scenarios, rather than whether they would accept to do something. If so, it is probably good practice not to ask TTO questions in terms of whether the respondent would give up years of life in exchange for improved quality of life. Health state valuation exercises may be subject to question order bias, but this can be accounted for by randomising the ordering of health states across different respondents. Starting point bias is possibly the most serious of all the biases. Since there is no 'correct' starting value, this cannot be controlled for within an iterative design by varying the initial offer across respondents and states. It is likely that the direction and magnitude of this bias will be dependent upon the value of each health state. If the TTO for all states were to start at 0.5 the effect of starting point bias for positive states would be that values greater than 0.5 would be downward biased and values less than 0.5 would be upward biased. However, the extent of this bias may diminish as respondents' value more health states. For negative health states with a starting value of -0.5 values greater than -0.5 are likely to be downward biased and values less than -0.5 will be upward biased. Regarding incentive compatibility, it is possible that the recommended MVH TTO protocol is susceptible to this, and if so, all values would be biased upwards, and values above the starting point would be affected more. Range bias is possibly the most complex issue. It is unlikely that TTO values at the top end of the positive states will be affected given that the upper bound is the highest possible valuation. However, range bias may be an issue at the lower end of the positive scale (if respondents believe that values cannot go lower). If this is the case health state values would be biased upwards. For negative health states valuations may be downward biased if respondents believe that values cannot go any lower.

Over all, the extent and impact of these biases on TTO health state values is likely to be dependent on the value of the health state. Values between 0 and 0.5 are likely to be the most affected given that the biases all move in the same direction and may result in higher health state values. Health state valuation using DCE with duration will not solve all the challenges, and it will remain susceptible to hypothetical, and scenario misspecification biases. However, most importantly, DCE will avoid the starting point bias and incentive compatibility, since there are no follow-up questions on the given health state.

This paper has argued that iterative health state valuation exercises such as the TTO or the SG are, like iterative CV studies, susceptible to a series of possible biases. However, currently, extremely little is known about them. Empirical studies that are designed to diagnose the existence and extent of these biases are needed, and/or methods that are inherently less susceptible should be developed.

References

1. Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ. 1986;5(1):1-30. Epub 1986/02/09.

2. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. J Health Econ. 2012;31(1):306-18. Epub 2011/12/27.

3. Brazier J, Dolan P. Evidence of preference construction in a comparison of variants of the standard gamble method. Health Economics and Decision Science Discussion Paper Series 05/04, University of Sheffield 2005.

4. McIntosh E, Clarke PM, Frew E, Louviere J, editors. Applied Methods of Cost-Benefit Analysis in Health care. New York: Oxford University Press; 2010.

5. Group M. The Measurement and Valuation of Health: First Report on the Main Survey. Centre for Health Economics, University of York: 1994.

6. McNamee P, Ternent L, Gbangou A, Newlands D. A game of two halves? Incentive incompatibility, starting point bias and the bidding game contingent valuation method. Health Econ. 2010;19(1):75-87. Epub 2009/02/05.

7. Bhatia MR. From evidence to calibration for starting point bias: willingness to pay for treated mosquito nets in Gujarat, India. Appl Econ. 2005;37(1):1-7.

8. Loomis J, Brown T, Lucero B, Peterson G. Improving validity experiments of contingent valuation methods: Results of efforts to reduce the disparity of hypothetical and actual willingness to pay. Land Econ. 1996;72(4):450-61.

9. Murphy JJ, Stevens TH, Weatherhead D. Is cheap talk effective at eliminating hypothetical bias in a provision point mechanism? Environ Resour Econ. 2005;30(3):327-43.

10. Blumenschein K, Johannesson M, Yokoyama KK, Freeman PR. Hypothetical versus real willingness to pay in the health care sector: results from a field experiment. Journal of Health Economics. 2001;20(3):441-57.

11. Cummings RG, Taylor LO. Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. Am Econ Rev. 1999;89(3):649-65.

12. Ozdemir S, Johnson FR, Hauber AB. Hypothetical bias, cheap talk, and stated willingness to pay for health care. Journal of Health Economics. 2009;28(4):894-901.

13. Klose T. The contingent valuation method in health care. Health Policy. 1999;47(2):97-123. Epub 1999/10/28.

14. Skjoldborg US, Gyrd-Hansen D. Conjoint analysis. The cost variable: an Achilles' heel? Health Economics. 2003;12(6):479-91.

15. Stewart JM, O'Shea E, Donaldson C, Shackley P. Do ordering effects matter in willingness-to-pay studies of health care? J Health Econ. 2002;21(4):585-99. Epub 2002/07/31.

16. Venkatachalam. The contingent valuation method: a review. Environmental Impact Assessment Review. 2004;24:89-124.

17. Whynes DK, Wolstenholme JL, Frew E. Evidence of range bias in contingent valuation payment scales. Health Econ. 2004;13(2):183-90. Epub 2004/01/23.

18. Ready RC, Buzby JC, Hu DY. Differences between continuous and discrete contingent value estimates. Land Econ. 1996;72(3):397-411.

19. Holmes TP, Kramer RA. An Independent Sample Test of Yea-Saying and Starting Point Bias in Dichotomous-Choice Contingent Valuation. J Environ Econ Manag. 1995;29(1):121-32.

20. Whitehead JC. Incentive incompatibility and starting-point bias in iterative valuation questions. Land Econ. 2002;78(2):285-97.

21. Herriges JA, Shogren JF. Starting point bias in dichotomous choice valuation with follow-up questioning. J Environ Econ Manag. 1996;30(1):112-31.

22. Frew EJ, Wolstenholme JL, Whynes DK. Comparing willingness-to-pay: bidding game format versus open-ended and payment scale formats. Health Policy. 2004;68(3):289-98. Epub 2004/04/29.

23. Phillips KA, Homan RK, Luft HS, Hiatt PH, Olson KR, Kearney TE, et al. Willingness to pay for poison control centers. Journal of Health Economics. 1997;16(3):343-57.

24. Stalhammar NO. An empirical note on willingness to pay and starting-point bias. Med Decis Making. 1996;16(3):242-7.

25. Carson RT, Groves T. Incentive and informational properties of preference questions. Environ Resour Econ. 2007;37(1):181-210.

26. Watson V, Ryan M. Exploring preference anomalies in double bounded contingent valuation. J Health Econ. 2007;26(3):463-82. Epub 2006/11/25.

27. Gudex C. Time Trade-Off User Manual: Props and Self-Completion Methods. 1994.

28. Dolan P. Modeling valuations for EuroQol health states. Medical care. 1997;35(11):1095-108. Epub 1997/11/21.

29. Szende A, Oppe M, Devlin N, editors. EQ-5D Value Sets: Inventory, Comparative Review, and User Guide: Spinger; 2007.

30. Brazier J, Tsuchiya A. Preference-based condition-specific measures of health: what happens to cross programme comparability? Health Econ. 2010;19(2):125-9. Epub 2010/01/06.

31. Augestad LA, Rand-Hendriksen K, Kristiansen IS, Stavem K. Learning effects in time trade-off based valuation of EQ-5D health states. Value Health. 2012;15(2):340-5. Epub 2012/03/22.

32. Devlin NJ, Tsuchiya A, Buckingham K, Tilling C. A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. Health Econ. 2011;20(3):348-61. Epub 2011/02/11.

33. Devlin N, Buckingham K, Shah K, Tsuchiya A, Tilling C, Wilkinson G, et al. A Comparison of Alternative Variants of the Lead and Lag Time Tto. Health Econ. 2012. Epub 2012/06/21.

34. Samuelsen C, Augestad LA, Stavem K, Kristiansen IS, Rand-Hendriksen. K. Anchoring bias in the Lead-Time Time Trade-Off: Heuristic influence on the elicitation of preferences for hypothetical health sates. 2012.

35. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. Med Decis Making. 1994;14(1):9-18. Epub 1994/01/01.

36. Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. Medical care. 2007;45(3):238-44. Epub 2007/02/17.

Appendix: Summary of the recommended MVH TTO protocol

The recommended MVH TTO protocol (27) asks the respondent to choose between two alternatives: Life A and Life B. The exercise has a special visual aid called the TTO Board. For states better than being dead, the front of the board is used, where Life A is to live in full health for a duration between 0 years and 10 years, indicated by the sliding pointer the interviewer operates. Life B is fixed: to live in the health state in question (say, state H) for 10 years. On the other hand, for states worse than being dead, the board is flipped over, where Life A now consists of first living in state H in question, and then in full health. These two states have a fixed total duration of 10 years, and the pointer specifies the timing of the transition from state H to full health. Life B is immediate death.

Step 1. For the very first health state only, the TTO exercise begins with the pointer on the front side of the TTO board at 10 years. The accompanying question asks the respondent to choose between living in full health for 10 years (Life A) and living in the health state in question for 10 years (Life B). Unless the health state in question is indistinguishable from full health, the respondent is expected to select Life A, and to proceed to Step 2.

Step 2. For the second health state onwards, the TTO starts here, with the pointer on the front side of the TTO board at 0 years. The interviewer asks to choose between Life A, which is immediate death, and Life B, which is 10 years in state H. This question screens out those who consider state H is worse than being dead.

Step 3 for states better than dead. Those respondents who choose Life B will now see the pointer at 5 years. From here onwards, if the respondent's choice is Life A, the pointer will move to the left by 1 year, and if it is Life B, then the pointer will move to the right by 1 year. If at any point the respondent becomes indifferent between Life A and Life B, the valuation for state H is complete. If at any point the respondent ‘switches’ between Life A and Life B, then the pointer is moved back by 6 months. At this point, the valuation for state H is complete regardless of what the respondent chooses. If indifference is not reached after a 6-month move, indifference is assumed for the mid-point between the shortest duration where Life A was chosen and the longest duration where Life A was rejected.

For all states better than dead, the value of the health state is derived by dividing the duration in full health by 10 years. So if a respondent chooses Life A when the pointer is at 4 years or longer and rejects Life A when the pointer is at 3 years and 6 months or less, then the health state value for H is 0.375. The open ended question for those who rejects Life A at 9 years 6 months asks whether the respondent would give up any time in Life A to avoid Life B, and the answer is recorded in the number of weeks.

Step 3 for states worse than dead. Those respondents who choose Life A at Step 2 will go to the reverse side of the TTO board. The pointer will first be at 5 years, indicating that Life A consists of state H for 5 years followed by full health for 5 years. From here, a procedure similar to that for states better than dead is followed. The pointer will move either to the left or right by 1 year depending on the respondent’s choice, followed by a 6-momth move after switches.

Deriving the value of a health state worse than dead is controversial (35, 36), but with the MVH protocol, the duration in full health is divided by -10. So if a respondent is indifferent between Life A and Life B when the pointer is at 6 years and 6 months, that means 3 years and 6 months in full health, and the value of health state H is calculated as -0.35 (see (28) for more details)).

Once the valuation of state H is complete, the interview will move on to the TTO valuation of the next state. The MVH TTO protocol consisted of each respondent valuing 13 EQ-5D health states of varying severity, in random order. Other studies using the MVH TTO protocol have used 13 to 17 states per respondent.