This is a repository copy of Protocols for TTO valuations of health states worse than dead: A literature review and framework for systematic analysis.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/10905/

Monograph:

HEDS Discussion Paper 08/09

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
HEDS Discussion Paper 08/09

Disclaimer:
This is a Discussion Paper produced and published by the Health Economics and Decision Science (HEDS) Section at the School of Health and Related Research (ScHARR), University of Sheffield. HEDS Discussion Papers are intended to provide information and encourage discussion on a topic in advance of formal publication. They represent only the views of the authors, and do not necessarily reflect the views or approval of the sponsors.

White Rose Repository URL for this paper:
http://eprints.whiterose.ac.uk/10905/

Once a version of Discussion Paper content is published in a peer-reviewed journal, this typically supersedes the Discussion Paper and readers are invited to cite the published version in preference to the original version.

Published paper
None.

White Rose Research Online
eprints@whiterose.ac.uk
Health Economics and Decision Science Discussion Paper Series

No. 08/09

Protocols for TTO Valuations of Health States Worse than Dead:
A literature review and framework for systematic analysis

Carl Tilling
School of Health and Related Research, University of Sheffield

Nancy Devlin
Economics Department, City University, London

Aki Tsuchiya
Department of Economics and the School of Health and Related Research,
University of Sheffield

Ken Buckingham
Department of Preventive and Social Medicine, University of Otago, New Zealand

Corresponding author:
Aki Tsuchiya
School of Health and Related Research, University of Sheffield,
Regent Court, 30 Regent Street
Sheffield S1 4DA, UK
Telephone: 0114 222 0710
Fax: 0114 272 4095
Email: a.tsuchiya@Sheffield.ac.uk

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.
Acknowledgements

We would like to thank Dr. Louise Longworth for helpful comments on an earlier version of this paper. This study is funded by a grant from the EuroQoL group, as part of a wider project piloting a new protocol for states worse than dead. The usual disclaimer applies.
1. Introduction

Resource allocation decisions in the health care sector are under ever increasing levels of scrutiny. Cost-Utility Analysis (CUA) is the most widely used method of economic evaluation used to inform such decisions. This involves estimating, for any potential new treatment, the incremental cost per Quality Adjusted Life Year (QALY) gained, in comparison with the existing treatment. QALYs are calculated by multiplying the duration in any given health state by the corresponding values for the health-related quality of life in each period. These values are anchored at 1 for ‘full health’ and 0 for ‘dead or equivalent’; values < 0 represent states considered to be ‘worse than dead’ (SWD). The terms ‘utility’, ‘value’ and ‘preference’ are often used interchangeably in the literature on CUA and QALYs – although the distinction between the terms and underlying concepts continues to be a matter of academic debate (for example, see Drummond et al. 2005, p.140-147, and Parkin and Devlin 2006).

There are three main methods for eliciting preferences to yield a value for the quality adjustment in the QALY: time trade-off (TTO), visual analogue scale and standard gamble. The TTO method, developed by Torrance et al (1972), is the focus of this review, and consists of a trade-off between length and quality of life. Participants in TTO exercises are presented with a choice between two scenarios. In the case of a TTO valuation for a chronic state better than dead (SBD) one of these scenarios consists of living for a given fixed period of time ($t_j$) (e.g. 10 years) in the health state to be valued, followed by death. The other scenario (the ‘trading scenario’) consists of a shorter period of time ($t_i$) in full health, followed by death. The value of $t_i$ is varied in the trading scenario until a point of indifference between the two scenarios is found and then (assuming zero temporal discounting) the utility value for that health state is given by the value $t_i/t_j$. TTO exercises can be undertaken by medical experts,
members of the general population or patients. Furthermore, people can be asked to value either their own current health or health states which are hypothetical to them.

In the case of SWD, there are two commonly used TTO protocols, both of which differ from the protocol for SBD. The method developed by Torrance (1982) gives respondents a choice between a scenario of living in full health for \( t_i \) years followed by the state to be valued for \( t_j \) years \((t_i + t_j = T)\), followed by death; and an alternative scenario which is to die immediately. The value \( T \) is fixed e.g. 10 years. The value of \( t_i \) (and therefore also the value of \( t_j \)) is varied until a point of indifference is found between the two scenarios. The utility value for that health state is then given by \(-t_i/t_j\).

The protocol used in the Measurement and Valuation of Health (MVH) study (Dolan 1997), which generated the UK’s TTO value set for the EQ-5D (Brooks, 1996) used a method similar to this, but the first scenario is to live in the health state to be valued for \( t_j \) years followed by full health for \( t_i \) years (i.e. the ordering of the two states is reversed). The second scenario is still to die immediately, and the formula used to obtain the value of the state is the same as that used in the Torrance method. The reversing of the states in the first scenario (the state to be valued then full health vs. full health then the state to be valued) was based on the argument that this ordering of states forced the respondent “to think more directly about being in that state for \( y \) number of years and thus to recognise the sacrifice that must be made before being in full health for \( x \) number of years” (Gudex 1994 p.3). One way in which the participant may not consider the sacrifice appropriately in the Torrance approach is if they think they can commit suicide once they have enjoyed the time in full health and before the state to be valued begins. If participants do take such a view then their responses would be meaningless.

However, the MVH protocol, although widely used and replicated in subsequent studies, is by no means a perfect alternative to the Torrance method. For example, some respondents may not consider the scenario of going from a very poor health state to full health to be realistic (some extremely poor states may be most realistically associated with terminal illnesses; see for example Robinson et al, 1997). This is problematic if participants are not fully engaging with the task and hence not providing values that are a meaningful representation of their preferences. Unrealistic
scenarios can also be problematic if participants refuse to respond to the questions, although this may be considered preferable to a meaningless and inaccurate response.

The ordering of full health to poor health (as in Torrance) or poor health to full health (in the MVH) may also be important where additive separability (Broome, 1993) does not hold. That is, the same state may be valued differently depending on whether it is part of a declining health profile (i.e. followed by death) or part of an improving health profile (i.e. followed by full health). If so, one may argue that where SBD are followed by dead so should SWD (as is the case in the Torrance design).

There are a number of other important problems associated with both the Torrance and MVH approaches to TTO valuation of SWD. In both cases, valuations for SBD are elicited using fundamentally different trade-off procedures to those used for SWD. This implies that the aggregation of participants’ results for a given state across SBD and SWD (e.g. in order to calculate means or in the estimation of value sets) is of questionable validity. Secondly, the denominator of the utility value calculation \((T-t_i)\) is no longer a fixed number as it is in the protocol for SBD \((t_i)\). One consequence of this is that the method is problematic if people have very different valuations for health experienced for different periods of time. Another consequence is that negative values can be extremely negative. A participant who would not accept any amount of time, however short, in a poor state of health, is implying that such a state is infinitely bad. Given the mathematical intractability of dealing with negative infinity (a single value of negative infinity in a sample of respondents would give a mean value of negative infinity) researchers usually censor such responses. Under such censoring the lower bound is determined by the (relatively arbitrary) choice of units of time that are used in the study. For example, the lowest value that the MVH study elicited was -39, which was a result of performing the TTO in units of 3 months. If respondents stated that 9 years and 9 months in full health, followed by 3 months in the severe state, was equivalent to death their value would be given by, \(-9.75/0.25 = -39\). If the protocol had instead restricted trades to whole years, the minimum value would have been: \(-9/1= -9\). Alternatively, if participants had been allowed to trade in months, the minimum value would have been: \(-9.92/0.08= -119\). A third problem is a direct consequence of these extreme negative values. The conventional use of QALYs in CUA views an improvement from -0.2 to 0 as identical to an improvement from 0 to
However, if the negative values do not fit the range 0 to -1, some individuals may have a disproportional influence upon the calculation of means and tariffs. One established response of researchers has been to transform negative values to fit the scale 0 to -1 (Torrance et al. 1982; Patrick et al. 1994). Lamers (2007) reports and compares three methods of achieving this transformation: monotonic (as used in the MVH study), linear and truncation.

However, it is not obvious why there should be no states worse than -1. While it makes data analysis easier to transform values in this fashion, arguably one year of extreme pain and discomfort might provide as much disutility as two years of full health provides in utility. Richardson and Hawthorne (2001) do not share this view and argue, based on the results from a VAS survey and their belief that it is not possible for individuals to experience a state as bad as -1, for an even higher lower limit of -0.25. Patrick et al. (1994) point out that, once transformed, the negative numbers for SWD can no longer be interpreted as “utility scores”, measured on the same scale as those for SBD. However, if positive and negative values are not measured on the same scale then comparison and aggregation becomes problematic.

An additional problem associated with the use of different TTO valuation procedures for SBD and SWD is the existence of a ‘gap effect’ in the MVH TTO values around dead, as observed by Stalmeier et al. (2005). In an analysis of differences in the values for adjacent states, they find the differences in TTO values for states either just above or below 0 are at least twice as large compared to the differences between other adjacent states. Further, while there are a large number of states valued at exactly 0 in the MVH TTO value set, there are relatively few states that are close to 0 (> 0 or < 0). This means that SWD tend to be substantially worse than dead. While this may reflect the genuine preference of the participants in TTO exercises, on the other hand, it may be caused by the separate valuation procedure employed for SWD. It may also be partially related to a further problem with the MVH protocol: when a SWD is encountered, participants are explicitly asked whether a state is worse than dead early on in the valuation process, by making a choice between immediate death and 10 years in the health state to be valued. This may result in a ‘focusing effect’: by asking participants to consider a stark trade-off between life in a given poor state of health
and death, this may result in an exaggeration of their expressed dislike of the state. A uniform procedure for SWD and SBD could avoid this potential problem.

The above problems with existing protocols for SWD have potentially important implications for the conduct of CUA. The aim of this review is to provide an in-depth account of the different ways of presenting the TTO for SWD, in order to identify any alternatives to the MVH and Torrance approaches, and to consider the relative merits of the approaches identified. This is achieved firstly by examining all health state valuation studies that have employed the TTO, to record, categorise and report the way that SWD are (or are not) handled. Secondly, in order to ensure that there are no unpublished but potentially feasible TTO variants we set out a framework for identifying all potential variants and assess the feasibility of each. While there have been two recent reviews of TTO methodologies (Arnesen and Trommald, 2004; Arnesen and Trommald 2005), the focus of these reviews are on different issues – for example, the latter aimed to identify instances in which the TTO method had been used to value own (current) health in order to see if QALY values from different TTO studies are comparable. In contrast, the present review has its emphasis on identifying different valuation procedures for SWD.

2. Methods and results of the Literature Search

Studies were retrieved from a literature search in the database Medline (via Ovid), from 1950 to September 2007. The search was of titles and abstracts. The initial search term was ‘TTO or Time Trade Off’, yielding 493 results. These results were cross-referenced against the most recent systematic review of TTO methodologies by Arnesen and Trommald (2005; henceforth A&T). Cross referencing showed that five studies identified by A&T did not appear in our search results. A&T do not provide fine detail of their search terms and simply state, “The search criterion was ‘Time trade-off,’ in different spellings” (p.40). Including the search term ‘Time Tradeoff’ gave an additional 98 results but cross referencing showed that 1 study was still missing. Finally, including the term ‘Time Tradeoffs’ gave a further 2 results, 1 of which was the elusive article in A&T. This brought the total number of results to 593. Since the chosen method was a title and abstract search, we do not claim to have
identified all existing studies that employ TTO elicitation of values - it is possible that some studies may do so but not explicitly mention this in either the title or the abstract. When searching for and reviewing studies, no distinction was made between the terms mentioned earlier – ‘utility’, ‘value’ and ‘preference’.

The above process highlights the importance of attention to detail in the terminology used in this area. There is a need for awareness of potential pitfalls when conducting literature searches of this nature and also to encourage uniformity in terminology in future studies. In this vein, it is worth noting that the use of hyphens within all Ovid databases makes no difference to the results obtained. However, within other databases the user may need to specify any hyphens to be included in the search terms.

Any study that had administered a TTO first hand was included, and logged in an Excel spreadsheet. Some studies could be excluded on the basis of the title, some could be excluded on the basis of the abstract and in some cases it was necessary to read the article. Commonly excluded articles were studies on ‘tea tree oil’, reviews (e.g. Torrance, 2006), theoretical comparisons of utility measurement techniques (e.g. Rittenhouse, 1997) and comments on other studies (e.g. Hawthorne et al. 2003). Theoretical discussions of TTO were only included if they specifically focused on valuation of SWD (e.g. Richardson et al. 2001), and were used to assess possible frameworks for TTO but were not included in the review of empirical data. For the studies included in the review (see Appendix), the variables used to record and categorise these are provided in Table 1.

As well as the articles identified through the database search, a number of articles were included that the authors had prior knowledge of and which would not appear in a database search owing to their source. These included papers from the Health Economists’ Study Group archive, University of Monash Centre for Health Economics Working Papers, and an unpublished manuscript.
## Table 1. The data extraction form

<table>
<thead>
<tr>
<th>Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Author</td>
<td></td>
</tr>
<tr>
<td>Year of publication</td>
<td></td>
</tr>
<tr>
<td>Journal Name</td>
<td></td>
</tr>
<tr>
<td>Title of Article</td>
<td></td>
</tr>
<tr>
<td>Valuing own health or hypothetical health state?</td>
<td>Own, Hypothetical, Both</td>
</tr>
<tr>
<td>Is the wording of the TTO given?</td>
<td>No, Partly, Yes</td>
</tr>
<tr>
<td>Is there a protocol for SWD?</td>
<td>Separate from SBD, same as SBD, no, NA, unclear</td>
</tr>
<tr>
<td>If yes, then which protocol?</td>
<td>MVH, Torrance, Modified MVH, Other</td>
</tr>
<tr>
<td>Time horizon (T) for SWD?</td>
<td>10 years, 20 years, Unclear, NA</td>
</tr>
<tr>
<td>Smallest unit of TTO?</td>
<td>1 hour, 1 day, 0.5 weeks, 1 month, 2 months, 3 months, 6 months, 1 year, unclear</td>
</tr>
<tr>
<td>Mode of administration?</td>
<td>Interview, internet questionnaire, postal questionnaire, self-completed computer questionnaire, self-completed paper questionnaire, telephone, interview, other</td>
</tr>
<tr>
<td>Are props used?</td>
<td>Yes, No, With and without, Unclear,</td>
</tr>
<tr>
<td>Was the study part of the valuation of an existing health state ...?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>If yes, then which instrument?</td>
<td></td>
</tr>
<tr>
<td>Are the SWD values transformed?</td>
<td>Yes, No, Unclear, NA</td>
</tr>
<tr>
<td>If yes, what is the range?</td>
<td>-1 to 0, -0.25 to 0</td>
</tr>
</tbody>
</table>
3. Results of the Review

A total of 369 articles were included in the review. Table 2 shows the distribution of the key variables.

Table 2. Distribution of key variables (n=369)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothetical State only</td>
<td>121</td>
<td>32.8%</td>
</tr>
<tr>
<td>Own State only</td>
<td>199</td>
<td>53.9%</td>
</tr>
<tr>
<td>Both</td>
<td>49</td>
<td>13.3%</td>
</tr>
<tr>
<td>Wording of TTO given</td>
<td>75</td>
<td>20.3%</td>
</tr>
<tr>
<td>Wording not given</td>
<td>246</td>
<td>66.7%</td>
</tr>
<tr>
<td>Wording partly given</td>
<td>48</td>
<td>13.0%</td>
</tr>
<tr>
<td>Protocol for SWD separate from SBD</td>
<td>27</td>
<td>7.3%</td>
</tr>
<tr>
<td>Protocol for SWD same as for SBD</td>
<td>2</td>
<td>0.5%</td>
</tr>
<tr>
<td>All states assumed better than dead</td>
<td>333</td>
<td>90.3%</td>
</tr>
<tr>
<td>Unclear or NA</td>
<td>7</td>
<td>1.9%</td>
</tr>
<tr>
<td>Interview</td>
<td>282</td>
<td>76.4%</td>
</tr>
<tr>
<td>Self-completed computer questionnaire</td>
<td>25</td>
<td>6.8%</td>
</tr>
<tr>
<td>Self-completed paper questionnaire</td>
<td>24</td>
<td>6.5%</td>
</tr>
<tr>
<td>Telephone Interview</td>
<td>12</td>
<td>3.3%</td>
</tr>
<tr>
<td>Postal questionnaire</td>
<td>13</td>
<td>3.5%</td>
</tr>
<tr>
<td>Other or Unclear</td>
<td>13</td>
<td>3.5%</td>
</tr>
<tr>
<td>Use Props</td>
<td>127</td>
<td>34.4%</td>
</tr>
<tr>
<td>Don't use Props</td>
<td>68</td>
<td>18.4%</td>
</tr>
<tr>
<td>With and Without</td>
<td>2</td>
<td>0.5%</td>
</tr>
<tr>
<td>Unclear</td>
<td>172</td>
<td>46.7%</td>
</tr>
</tbody>
</table>

The majority of TTO valuation studies asked respondents to value their own health (53.9%), largely accounted for by studies within clinical trials. Only 32.8% of the studies asked respondents to value hypothetical health states. There is also a large number of studies that include TTO valuation tasks on both hypothetical health states and own health. These can predominantly be accounted for by studies that compare patient and population values for a specific condition (e.g. arthritis - see Soucek et al. 2005), and studies (often clinical trials) that obtain health state valuations from both

---

1 We were unable to access 8 articles. The abstracts suggest that although time trade-offs may have been administered they were likely to be valuing own health in clinical trials and hence are highly unlikely to be of importance in identifying protocols for SWD.
patients and clinical experts (e.g. Wells, *et al.* 2004). Of the 369 studies, 11 valued an existing health state classification instrument. Eight of these were generating EQ-5D value sets in different countries. The other three consisted of two condition specific measures and a valuation of the Assessment of Quality of Life (AQoL) Instrument (Richardson *et al.* 2004). All but the two condition specific measures included a protocol for states worse than dead (MVH in all cases).

If a study included the wording of the TTO elicitation procedure in quotation marks within the text, in a separate box (as seems to be the most common and clearest method of reporting), or even if a thorough explanation had been given in the text, then we deemed that the wording of the TTO had been given. If a less detailed and more general explanation was given we deemed that the wording had been partly given. Despite these rather generous criteria, 66.7% of the studies fell into the category of ‘wording not given’. This is alarming given the extent to which health state valuations can be affected by the framing of valuation exercises (see for example Nord, 1992), and the widely acknowledged need for transparency in the reporting of economic evaluations and clinical studies. The poorest reporting seems to be amongst medical studies (clinical trials and RCTs) which attached less importance to all aspects of the methods section.

A further indication of poor reporting is in the case of the use of props and visual aids. In 46.7% of studies it was unclear whether or not props had been used. Among the studies in which it was possible to determine whether or not props had been used, more studies (34.4%) used props than did not (18.4%). It is important to know whether props have been used as they are likely to affect the way participants respond to the valuation task. Therefore, if researchers wanted to replicate the study using a different sample and compare results they would need to know whether or not props had been used, and if so, what kind. Ideally, studies that do not use any props should explicitly report that none were used.

The majority of studies used interviews to administer the TTO exercise (76.4%). The second most common method was self-completed computer questionnaire (6.8%), predominantly through the software U-Titer (Summer *et al.* 1991). Self-completed paper questionnaire was equally prevalent (6.5%). The ‘other’ category included two
internet based questionnaires and five group sessions. The internet and computer based modes of administration were used in more recent studies: the earliest internet based questionnaire was published in 2001 (Chang et al. 2001) and the earliest computer based questionnaire was published in 1995 (Nease et al. 1995).

The review only identified 29 studies – just 7.9% of those in our review – that had allowed for the possibility that a state to be valued by participants might be considered to have a value less than zero, and the characteristics of these studies can be found in Table 3. Twenty seven of these protocols employed a TTO procedure for SWD that was different to the protocol for SBD, of which 23 used the MVH protocol and 4 used the Torrance protocol.

Table 3. Distribution of key variables for studies employing a protocol for states worse than dead (n=29)

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothetical State only</td>
<td>27</td>
<td>93.1%</td>
</tr>
<tr>
<td>Own State only</td>
<td>2</td>
<td>6.9%</td>
</tr>
<tr>
<td>Both</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Wording of TTO given</td>
<td>5</td>
<td>17.3%</td>
</tr>
<tr>
<td>Wording not given</td>
<td>17</td>
<td>58.6%</td>
</tr>
<tr>
<td>Wording partly given</td>
<td>7</td>
<td>24.1%</td>
</tr>
<tr>
<td>Interview</td>
<td>27</td>
<td>93.2%</td>
</tr>
<tr>
<td>Self-complete computer questionnaire</td>
<td>1</td>
<td>3.4%</td>
</tr>
<tr>
<td>Group Session</td>
<td>1</td>
<td>3.4%</td>
</tr>
<tr>
<td>Use Props</td>
<td>21</td>
<td>72.4%</td>
</tr>
<tr>
<td>Don’t use Props</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>With and Without</td>
<td>1</td>
<td>3.5%</td>
</tr>
<tr>
<td>Unclear</td>
<td>7</td>
<td>24.1%</td>
</tr>
<tr>
<td>Smallest unit of TTO - 1 year</td>
<td>12</td>
<td>41.4%</td>
</tr>
<tr>
<td>1 day</td>
<td>3</td>
<td>10.3%</td>
</tr>
<tr>
<td>6 months</td>
<td>2</td>
<td>6.9%</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>6.9%</td>
</tr>
<tr>
<td>Unclear</td>
<td>10</td>
<td>34.5%</td>
</tr>
<tr>
<td>SWD values transformed</td>
<td>19</td>
<td>65.5%</td>
</tr>
<tr>
<td>SWD values untransformed</td>
<td>9</td>
<td>31.0%</td>
</tr>
<tr>
<td>Unclear</td>
<td>1</td>
<td>3.5%</td>
</tr>
</tbody>
</table>
A notable exception of the 29 is the study by Robinson and Spencer (2006), which reported a TTO procedure that was the same for SWD and SBD. This study imitated a TTO by asking respondents to rank different life profiles consisting of combinations of good health, poor health and death. The principal innovation of the method is its introduction of a ‘lead time’ in full health in each of the alternatives presented. The approach avoids the need to have different valuation procedures for SBD and SWD by allowing participants to trade their lead time to avoid the poor health state.

The other study that used the same protocol for SWD and SBD, by Detsky et al. (1986), used a two stage TTO procedure with chaining. Respondents first valued state 1 (SBD) in the normal TTO fashion. They were then asked to choose between one month in health state 1 and one month comprising a number of days in health state 2 (potentially worse than dead) and the remainder in full health. The number of days in full health is varied until the respondent is indifferent between the two options. The point at which indifference is achieved is chained via the previously obtained value for health state 1 to produce a value for health state 2, which can potentially be worse than dead. This approach has several disadvantages. Firstly, it is time consuming and complex and does not enable researchers to calculate a value for a SWD unless a SBD has already been valued. Secondly, the second stage of the TTO implies variable denominators when calculating the value, and as a result, values for SWD may be 'extremely' negative, as in the Torrance or MVH designs. However, the paper does not discuss whether there were observations less than -1, and if there were, what was done about them. As a result, whilst this approach uses a unified protocol for SWD and SBD, it does not solve the key problems of the current protocols for SWD. A citation search of Detsky et al. (1996) suggests that there are no publications that used this approach since.

A further study, by Stalmeier et al (2007), is not a TTO (and hence not included in the tables above, or considered in the following section on possible TTO variants) but is arguably relevant in this context. The approach asks respondents to make a binary choice between whether they think a health state is worse than dead. The responses are only useful for health state values at the aggregate level. If state X is preferred to dead (by an individual) it receives a score of 1, if state X is considered equal to dead it receives a score of 0, and if state X is considered worse than dead it receives a score
of -1. Scores can then be aggregated across respondents. For example, if 70%, 20% and 10% of the respondents had respectively responded with ‘better than dead,’ ‘equal to dead’ and ‘worse than dead’, then the health state would receive the following value: \((0.7 \times 1) + (0.2 \times 0) + (0.1 \times -1) = 0.6\). This method is dependent upon the existence of large variations in responses between individuals. For example, for a state with a value in the MVH value set of 0.5 to gain the same value through the Stalmeier method, half of the respondents would have to indicate that this state was worse than dead, while the other half would have to indicate that it was better than dead.

All but two of the studies using a protocol for SWD ask respondents about a hypothetical health state – this is probably a product of the ethical considerations involved in asking people to contemplate the possibility that their own health state may be worse than dead. Furthermore, individuals who are likely to value their health state as worse than dead are arguably less likely to be able to complete a TTO exercise. The first of the two studies that used a protocol for SWD and valued own health was by König, Roick and Angermeyer (2007) and elicited values for patients with schizophrenic, schizotypal or delusional disorders. The negative values were unbounded (and ranged from -1.9 to 0). The study is unclear about how many respondents gave negative values but given that the mean TTO score is 0.75 one would suspect only a few negative values were given. The authors acknowledge that the reliability of TTO results elicited from patients with mental disorders is questionable. The second study, by Detsky et al. (1986), asked respondents to value the state of health they had experienced two months prior to the interview (generally a state of chronic malnutrition). The mean TTO score obtained was -0.30. The standard deviation was very high (0.99) which may be a result of the two stage procedure described earlier.

Given the complexity of using two different protocols, one for SBD and one for SWD, this means participants are likely to require a substantial amount of explanation and guidance. This explains why none of these studies was administered via postal survey. All but two of the 29 studies use the interview method of administration. The two other studies used a group session which allows for explanation and guidance, and a self-complete computer questionnaire which uses graphical props. As with the full
sample of studies, the standards of reporting are poor. Of the 29 studies, 17 do not give the wording of the TTO explicitly, and it is unclear in 7 of them whether props have been used. An interesting observation is that the use of props appears to be more prevalent amongst studies employing a protocol for SWD (72.4% compared with 34.3%). This is most probably due to the more complex nature of using multiple procedures which necessitates the improved clarity introduced by using props.

The most common unit of TTO for SWD was 1 year which, given that 21 of the 29 studies had a time horizon of 10 years for SWD, means that these SWD values, if left untransformed, would have a lower bound of -9. However, in 19 of the 29 studies the SWD values were transformed. All but one of these transformations were to fit to the range -1 to 0, with the other being to the range of -0.25 to 0 (Richardson et al. 2004, as noted earlier).

4. Conclusions from the literature review

The most striking findings of our review of the published literature are as follows:

(i) TTO valuation research appears to be characterised by attempts to elicit values from patients, rather than from members of the general public. Partly for this reason,

(ii) A relatively small number of studies incorporate valuation procedures that allow for states to be worse than dead in TTO.

(iii) All of those TTO studies that do address SWD, with two notable exceptions, employ separate elicitation procedures for SWD than SBD.

(iv) Reporting standards regarding TTO valuation exercises are generally poor, with widespread under-reporting of interview techniques, the nature of visual aids and props and how they were used, and other key aspects of methods known to affect the values.

Reporting of studies using TTO need to be improved and standardised. This stretches from the specific term used for TTO in the title and abstract, to accurate and precise descriptions of the TTO protocol used, whether props were included, what the
The smallest unit of trade-off was, and so on. The first point here is highlighted by the problems encountered when trying to replicate the search terms used in the review by A&T. Furthermore, the need to improve reporting standards for comparability across studies is becoming recognised and for example within the U.K the National Institute for Health and Clinical Excellence (NICE) has published guidelines for performing and reporting economic evaluations submitted to the Institute (NICE, 2008).

The interview mode of administration is the most prevalent in TTO studies, especially amongst studies with protocols for SWD. This is encouraging because the interview approach is generally considered to be the most accurate and reliable, given that the interviewer is able to ensure that the participant fully understands the exercise and can also encourage the participant to give as much thought as possible to the key decisions. However, the interview method is both time consuming and expensive. The internet now provides an inexpensive and easy alternative that enables researchers to gain large sample sizes. One emerging research topic in this context is the extent to which face-to-face interview surveys and on-line surveys generate similar data.

The majority of studies identified in our review asked people to value their own health, while amongst the smaller number of studies using protocols for SWD, the majority ask people about health states which are hypothetical to them. Further comparison of the two sets of studies shows that the use of props is more common amongst the latter.

Almost all the studies identified that had a protocol for SWD used either the Torrance or MVH protocols, so are subject to the issues noted in our Introduction. Further, 19 of the 29 studies with protocols for SWD transformed the resulting negative values, which is also problematic. This review reinforces the need for a protocol that is able to elicit values for SBD and SWD in a ‘one size fits all’ fashion. If SWD are valued in a uniform protocol with SBD then it is anticipated that there will be no need – or justification – to transform, since the values for SWD will be on the same scale as the values for SBD, and therefore will need to be taken at face value. Furthermore, a ‘one size fits all’ protocol would produce consistent results irrespective of the choice of

---

2 A new set of NICE guidelines are currently under consideration and are likely to be confirmed and published by the end of 2008
units and would not make it obvious to respondents at which point a state becomes worse than dead.

In summary, the literature review identified five TTO protocols for SWD of note. Two of these – the MVH and Torrance designs – are widely used but, as discussed earlier, give rise to problems because of their use of different valuation procedures for SBD and SWD. Of the three further protocols identified, that used by Detsky et al. (1986) which relies on a chaining method, and Stalmeier et al. (2007), which is not a TTO, both overcome the problem of separate elicitation procedures for SBD and SWD, but are each associated with other problems. The method adopted by Robinson and Spencer (2006) appears promising as a way potentially to overcome the problems with existing protocols.

In order to assess this method further, and to consider whether there may be other, related variants of the TTO that are theoretically possible but are not discussed or tested in the literature, we developed a framework for systematically identifying the component parts of any TTO and how these may feasibly be combined to create alternative elicitation procedures. This is reported in the following section.

5. A framework for the systematic identification of all feasible TTO variants

The aim of this discussion paper is to identify the complete set of methodologies that have been used in research to elicit values for SWD. However, we also wanted to explore whether there exist other, theoretically possible TTO methodologies that had not yet been discussed in the literature or empirically tested. To this end we set out to identify, in general terms, the components of TTO methodologies and systematically to investigate whether it would be possible to construct other, untried TTO methodologies for the valuation of SWD using those components.

We were particularly interested to discover whether the approach of Robinson and Spencer (2006), noted in the previous section, may have opened a door to a wider set of other related, feasible TTO methodologies. As already noted, their approach
operates by adding a ‘lead time’, spent in full health, to both of the scenarios in a conventional TTO. This lead time enables trades to take place that would not otherwise be feasible. Box 1 illustrates the lead time technique when used to value a SBD. The area outlined by the dotted line in Box 1 comprises the ‘core’ states within a conventional TTO. The area outlined by the dashed lines shows the addition of lead times. In the case of SBD, trading into the lead time is not necessary. Box 2 shows the valuation of a SWD. Some of the lead time in full health in the trading scenario is ‘given up’ to equalise the utility of this scenario with that of the alternative scenario - effectively balancing out the negative utility associated with time spent in a SWD.

The Robinson and Spencer (2006) technique might be thought of as a form of equivalent variation for a health loss to a SWD, in which a reduction of time in full health produces the same decrement in utility as experiencing poor health. The method was also suggested in a theoretical paper by Buckingham and Devlin (2006). Depending on the precise manner in which the TTO question is framed, the conventional TTO can be similarly regarded, but for poor states of health better than being dead.

It is perhaps worth observing that the ‘conventional’ approach to TTO valuation of SBD (i.e. the Torrance and MVH protocols outlined in the previous section) is not feasible for SWD because the state of health being valued, if worse than dead, is dominated by both of the two states (full health and dead) being used in the tradeoff. *No* combination of time spent in full health and time dead can provide the same negative utility as a SWD. The usefulness of lead times therefore lies in their ability to alter the dominance between the ‘trading’ scenario (full health and dead) and the ‘alternative’ scenario, comprising time in the state to be valued. Adding a lead time in full health to the conventional TTO means that the trading scenario no longer inevitably dominates the alternative scenario. (It is perfectly feasible for full health and being dead in the trading scenario to be dispreferred to full health and a SWD in the alternative scenario, given sufficient lead time).
Box 1. The lead time TTO for a state that happens to be better than dead.
Box 2. The lead time TTO for a state that happens to be worse than dead

Given that lead times represent one possibility, symmetry suggests that lag times are also possible - comparable additions of times available for trading, but added at the end of the scenarios rather than at the start. An example of such an approach is illustrated in Box 3.

Box 3. A lag time TTO for a state that happens to be better than dead
We do not consider the multiple additional possibilities that ‘chaining’ (in which valuations of one state are used in the valuation of other states) might offer. Such chainings might be applied to any valuation methodology and provide further possibilities in the valuation of states worse than dead. The TTO methodologies that we do consider have the following general components:

1. The use of three ‘health’ states, namely: being in full health (or some equivalent); being dead; and being in the health state to be valued (in our context, the state that is valued as worse than being dead).
2. Two alternative scenarios of equal total duration, each being followed by dead, in which the ‘trading scenario’ contains at least two component states whose proportions within the fixed total duration are varied to equalise the utility of the two alternatives.
3. The use of additional time common to both alternatives (lead or lag times), to increase the amount of time subject to trading.

The intention in this section is to assess the viability of the various potential TTO methodologies that might be constructed from combinations of these characteristics. Viability is considered in the light of the criteria described below.

Most importantly, it is essential that trading between time spent in the various health states within the trading scenario is capable of achieving the same utility as the alternative scenario. If the states within the trading scenario are both superior to (i.e. dominate) the state(s) in the alternative scenario (or vice versa) no equilibrium can be achieved; and the potential methodology would be ‘non-feasible’.

In addition to the question of feasibility, other matters influence the viability of potential methods. Incorporating all three health states (the anchor states of full health and dead, as well as the state being valued) in the trading scenario can be shown to be unhelpful. This possibility arises when lead times or lag times are used and when the lead or lag time differs from the core health states in the trading scenario to which it is attached. For example if the lag time shown in Box 3 were spent in poor health rather
than in full health, the trading scenario would comprise the two ‘core states’ (time spent in full health, time spent dead) as well as time spent in poor health. It follows that the alternative scenario would now contain poor health plus more time spent in poor health. This lag time would be redundant; it would not alter the existence or otherwise of dominance between the scenarios. Redundancy can occur in other circumstances where, under normal preference structures (good health preferred to both being dead and bad health) equilibria would be achieved without trading into lead or lag times.

There are three further considerations. Firstly, it is advantageous if the same method is generalisable for use in the valuation of both SWD and SBD as, if different methods were used within a study, the valuations might be incommensurable. Secondly, it is potentially disadvantageous if the method confounds duration and health state. This may be the case if equilibrium is achieved by varying the duration of time spent in the health state being valued. Thirdly, it is also disadvantageous if the method relies on the assumption of a ‘Lazarus effect’ in which living states follow dead states. To be even remotely realistic, we would have to provide a believable temporary alternative to being dead, such as a coma (however, this may not have the same value as dead – Dolan 1997).

The full set of variants that we consider is shown in the cells of Table 4. We consider whether feasibility, redundancy, generalisability, confounding, or Lazarus effects apply to the combinations described by these cells.

Cell A1. This variant is used in the MVH valuation of SWD. The method is feasible for SWD, but is not generalisable to SBD. It does not require a Lazarus effect, however confounding exists which may be a problem.

Cell A2. This variant is used in the MVH valuation for SBD. It is feasible, does not result in confounding, does not give rise to temporal inconsistency, however it is not feasible for SWD.

Cell A3. This variant is not feasible for any state of health as full health dominates any combination of poor health and dead.
Cell B1. This variant has several problems. It is feasible for SWD, however the lead time contributes nothing, it is not generalisable to SBD, it confounds duration with the value of the health states, and it would rely on a Lazarus effect.

Cell B2. This combination is not feasible for SWD. No combination of being in poor health and dead could render the trading scenario equivalent to full health and dead in the alternative scenario.

Cell B3. This combination is not feasible. No combination of being in full health and dead could render the trading scenario equivalent to poor health and dead in the alternative scenario.

Cell C1. This combination is feasible. The trading scenario contains both full health and poor health. Increasing the time in full health can be used to equalise the utility of the trading scenario and the alternative scenario. However, the technique is unsatisfactory in that the time spent in the state of health being valued is not fixed.

Cell C2. This combination is not feasible as no combination of full health, dead and poor health can equal poor health worse than dead. Furthermore, the addition of poor health to an alternative scenario that already contains poor health renders the lead time redundant.

Cell C3. This combination is feasible (with a shorter time in poor health balancing the combination of time in poor health and time in full health), however, it cannot be generalised to SBD where the alternative scenario would dominate. Moreover, confounding would occur.

Cell D1. This combination is feasible, however confounding could be a problem.
Table 4. Full set of possible TTO Variants

<table>
<thead>
<tr>
<th>States in the trading scenario</th>
<th>full and poor</th>
<th>full and dead</th>
<th>poor and dead</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Neither lead time nor lag time

A Feasible, not generalisable, suffers from confounding. This technique was used by the MVH group for SWD
Not feasible for SWD. No other disadvantages. This technique was used by the MVH group for SBD
Not feasible

Lead time = being dead

B Feasible, suffers from redundancy, confounding and a Lazarus effect, not generalisable
Not feasible
Not feasible

Lead time = poor health

C Feasible, suffers from confounding
Not feasible
Feasible, not generalisable, suffers from confounding

Lead time = full health

D Feasible, suffers from confounding
Feasible, without redundancy, is generalisable, no confounding, no Lazarus effect. This is the method suggested by Robinson and Spencer (2006)
Not feasible

Lag time = being dead

E Feasible, not generalisable, suffers from confounding and redundancy
Not feasible
Not feasible

Lag time = poor health

F Feasible, but suffers from redundancy and a Lazarus effect
Not feasible
Feasible, but not generalisable and suffers from a Lazarus effect

Lag time = full health

G Feasible, but suffers from redundancy, is not generalisable, and suffers from a Lazarus effect
Feasible, without redundancy, it is generalisable, no confounding, suffers from a Lazarus effect
Not feasible
Cell D2. This is the method suggested by Robinson and Spencer (2006). It is feasible, does not suffer from redundancy, is generalisable to SBD, does not give rise to confounding, and does not rely on a Lazarus effect.

Cell D3. This combination is not feasible. The alternative scenario of full health for the full duration, dominates the trading scenario. Furthermore, there is redundancy.

Cell E1. This combination is feasible but not generalisable to SBD. Confounding could be a problem and the lag time is redundant.

Cell E2. This combination is not feasible. The trading scenario dominates the alternative scenario.

Cell E3. This combination is not feasible. The alternative scenario dominates the trading scenario.

Cell F1. This combination is feasible but no trading would take place into the lag time for SWD. Furthermore, it relies on a Lazarus effect.

Cell F2. This combination is not feasible as the trading scenario dominates the alternative scenario.

Cell F3. This combination is feasible, but would have to rely on a Lazarus effect. Moreover it could not be generalised to SBD (in such cases the alternative scenario would dominate the trading scenario).

Cell G1. This combination is feasible, but, for SWD, no trading would take place into the lag time, while for SBD the trading scenario would dominate. Moreover, the technique would require a Lazarus effect.

Cell G2. This combination is feasible. In fact it is the lag equivalent to Robinson and Spencer’s technique, with similar advantages but with the exception that it would require a Lazarus effect.
Cell G3. Again this combination is not feasible as the alternative scenario dominates the trading scenario. Moreover, the lag time is redundant.

After the completion of our literature search, our attention was drawn to an unpublished University of Utrecht PhD thesis by Verschuuren (2006), supervised by Professors Ben van Hout and Gouke Bonsel. The thesis reports a TTO method, employed in the valuation of episodes of pain, which might be thought of as a combination of lead and lag times. A lead time in good health is added to the trading scenario of good health (followed by dead), while a lag time in good health is added after the time spent in poor health in the alternative scenario (followed by dead). The method is similar to that of Robinson and Spencer (2006) except for the temporal re-positioning of time spent in poor health in the alternative scenario. If we assume away problems of time preference (the usual assumption in relation to the construction of TTO valuations) and framing effects, the valuations should be the same as for the Robinson and Spencer (2006) approach.

In fact, any sub-variant in which the health states take different temporal positions between the trading and alternative scenarios will have similar feasibility properties as the variants in which additional time is added in the same position in the trading and alternative scenarios, with the exception of time preference and framing effects. Most importantly, if dominance exists between the trading and alternative scenarios, altering the position of the added time will not alter that dominance. In variants F1 and G1, temporal re-positioning can overcome Lazarus effects, however as we observe in Table 4, these variants do have other serious problems associated with them.

If we were to extend our validity criteria to incorporate the impact of time preference on valuations, then sub-variants in which the temporal positioning of the added time is altered seem unlikely to be helpful. When added time is in the same temporal position in both the trading and alternative scenarios, the impact of time preference is likely to be less than when the added time is in different temporal positions within the scenarios.
6. Conclusions from the application of a systematic framework

Breaking TTO into its component parts and considering alternative ways these might be combined provides a way of categorising existing approaches, as well as identifying remaining possibilities and their feasibility. Robinson and Spencer (2006) is one of a wider class of lead time and counterpart lag time approaches, although in terms of the criteria of feasibility, validity, redundancy and the avoidance of implausibility (full health following death), none of these alternatives are obviously superior to it. Further, it remains superior \textit{a priori} to the class of non-lead time (and non-lag time) approaches, including MVH and Torrance, on the grounds of its employing a uniform elicitation procedure for SBD and SWD.

7. Conclusions

The way in which valuations of SWD are elicited is crucial – it affects not just the valuations for states which, on mean values, are deemed worse than dead (< 0) but also potentially affects the mean values of \textit{all} states including those which, in valuation ‘tariffs’, are better than dead (> 0). This in turn means that the way SWD are valued will have important implications for the estimation of QALYs in economic evaluation. Despite this, and non-trivial problems associated with widely used methods for eliciting TTO values for these states, little attention has been given to these issues in the literature.

Our review of the literature reveals that relatively few TTO studies incorporate protocols for dealing with SWD – and still fewer are explicit about the way valuations for these states have been handled. However, in large part this is related to the predominance of TTO studies examining patients’ valuation of their own health where, for a variety of reasons, SWD may be less evident. In valuation studies where members of the general public have been asked to value health states which are hypothetical to them – for example, to facilitate the estimation of the social value sets widely used in economic evaluation – the protocols for valuation of SWD tend to be more clearly described. However, nearly all such studies employ either the Torrance
or MVH procedures – both of which rely on different elicitation procedures for SBD and SWD, with all the attendant problems noted in the Introduction.

Out review of the empirical literature identified one study, Robinson and Spencer (2006), which introduced a lead time in full health into both the ‘trading’ and ‘alternative’ scenarios, allowing a uniform TTO elicitation procedure to be used for both SWD and SBD. Further investigation of the components of the TTO, using a framework for systematically analysing all feasible variants thereof, suggests that Robinson and Spencer’s lead time approach is one of a wider class of approaches that might employ either lead times or lag times in the TTO scenarios. Further, the temporal positioning of the lead/lag times might also vary between the scenarios being compared.

We conclude that alternative protocols exist that can, a priori, overcome the problems with current (Torrance and MVH) approaches to the TTO valuation of SWD. Further research is required to design, test and compare the properties of the valuation data generated by these alternative approaches.
References


[accessed 16 Oct 208]

APPENDIX: Articles included in the literature review


Gold, L., Rankin, B., Riley, T., Shiel, A. 2007. Australian health state values for EQ-5D: findings from VAS and TTO. Unpublished Manuscript


