

## Time trade-off: one methodology, different methods

Arthur E. Attema · Yvette Edelaar-Peeters ·  
Matthijs M. Versteegh · Elly A. Stolk

© The Author(s) 2013. This article is published with open access at Springerlink.com

**Abstract** There is no scientific consensus on the optimal specification of the time trade-off (TTO) task. As a consequence, studies using TTO to value health states may share the core element of trading length of life for quality of life, but can differ considerably on many other elements. While this pluriformity in specifications advances the understanding of TTO from a methodological point of view, it also results in incomparable health state values. Health state values are applied in health technology assessments, and in that context comparability of information is desired. In this article, we discuss several alternative specifications of TTO presented in the literature. The defining elements of these specifications are identified as being either methodological, procedural or analytical in nature. Where possible, it is indicated how these elements affect health state values (i.e., upward or downward). Finally, a checklist for TTO studies is presented, which incorporates a list of choices to be made by researchers who wish to perform a TTO task. Such a checklist enables other researchers to align methodologies in order to enhance the comparability of health state values.

**Keywords** Time trade-off · Design · Methodology · Health state valuation

**JEL Classification** B40 · I10

---

A. E. Attema (✉) · M. M. Versteegh · E. A. Stolk  
iBMG/iMTA, Erasmus University, P.O. Box 1738,  
3000 DR Rotterdam, The Netherlands  
e-mail: attema@bmg.eur.nl

Y. Edelaar-Peeters  
Department of Medical Decision Making, Leiden University  
Medical Centre, Leiden, The Netherlands

### Introduction

A cornerstone of economic evaluations is the quality-adjusted life year (QALY), a measure of quantity and quality of life. The QALY is designed to allow for comparison of treatments across conditions. However, empirical research shows that different valuation methods, used to generate the quality-adjustment part of the QALY, produce different results. Even when researchers have used the same technique to elicit values for the same health states, such as time trade-off (TTO), large differences in produced values have been found [1]. While some variation may be expected due to differences between respondents, large differences probably reflect that TTO tasks are conducted very differently so that methodological differences affect the outcome. Incomparability of information is a key problem for the legitimacy of decisions based on economic evaluations. Therefore, the comparability of TTO studies needs to be improved. The usual way to achieve this is by harmonizing data collection efforts. Unfortunately, there have been serious discussions about several aspects of the TTO methodology in recent years, leading to an increased variety of methodological approaches. In order to permit harmonized methods for data collection, we review central elements of the TTO with the aim to understand if, and how, results of TTO studies may be influenced by the specifications of the TTO.

The TTO method elicits preferences for health states by letting a subject imagine living a defined number of years in an imperfect health state. The subject then has to indicate the number of remaining life years in full health at which the respondent is indifferent between the longer period of impaired health and the shorter period of full health. Normalizing the value of full health to 1.0 leaves us the value of the impaired health state being represented by

the ratio of the two periods, i.e., the number of years in full health divided by the number of years in the impaired health state. There is little guidance on how researchers should proceed to determine this indifference point, and as a result they do it very differently. A recent comparison of the different variants that may be subsumed under the name of ‘time trade-off’ has revealed how incomparable these approaches can become. This led the authors to conclude that TTO studies may have little more in common than their objective to quantify a trade-off between length of life and quality of life [2].

Apart from the possibility that the TTO may be not an accurate measurement technique for health states after all, another point of view is to regard the co-existence of different TTO variants as a natural state in incremental method development. The TTO method may be considered finalized when it satisfies all of its requirements. This still seems not to be the case, which we illustrate by two prominent methodological deficits. A first problem in TTO concerns its feasibility. TTO has been introduced as an interviewer-based method, but researchers, who lack time or budget to collect TTO values using time-consuming interviews continue to look for alternatives. Furthermore, the TTO procedure is criticized for problems associated with valuation of health states that are considered to be worse than dead [3, 4]. Not surprisingly, such perceived problems with TTO gave rise to a number of efforts to improve this method. Innovations in TTO are found in all elements related to the methods: new tasks, different procedures for data collection and different analytical strategies. While decision makers require standardization, the lack of scientific agreement on the best methodology sparked innovation and, as a result, comparing TTO studies is harder than ever.

While the proliferation of different TTO specifications may be seen as a valuable and integral part of the scientific endeavor to achieve optimal valuation methods, methodological variation in TTO studies is problematic from a policy point of view. TTO studies are typically conducted in the context of health technology assessments (HTA), which aim to inform policy makers in making resource allocation decisions in health care. As variants of TTO are known to produce different values for the same states [5–7], this reduces comparability of the HTA submissions. Hence, it is relevant to strive for more standardization. Standards have been developed for cost studies. In contrast, little guidance is provided for health state valuation studies. To our knowledge, NICE is the only institute that has issued guidance on how TTO studies should be conducted [8]. Their guidance indicates that those who conduct a TTO study should follow the measurement and valuation of health (MVH) protocol developed in the UK EQ-5D

valuation studies, based on the desire to enhance comparability with studies that derive utilities from the EQ-5D [9]. While other governmental institutions, health insurers and other organizations are equally dependent on comparable information, they have not issued similar guidance. A lack of scientific agreement combined with little attention for the need for standardization may be the reason for this.

Against this background, the current article has two aims. First, we aim to present the state of play in TTO and achieve an understanding of the merit of various ways in which the TTO method is applied. Second, ideas will be distilled about characteristics of TTO studies that ought to be reported in order to account for consequences of particular methodological choices in comparing TTO results. We bring these elements together in a checklist.

## Methods

To identify characteristics of TTO exercises that affect outcomes, we followed a convenient approach. Our starting point was the list proposed by Stalmeier et al. [10]. This list was expanded on the basis of expert opinion, literature and knowledge of the recent developments in TTO explored in the EQ group research program. As a result, we came up with a list of factors that investigators on TTO have to decide on before conducting a TTO elicitation (Tables 1, 2, 3). This list is not comprehensive as some aspects were deliberately excluded. For instance, paper-based TTOs were omitted, because the development of computerized designs has made this very uncommon in recent TTO elicitations.

In order to add structure, the identified characteristics have been subsumed under three headings: methodological issues, procedural issues and analytical issues. All these issues can be expected to influence results, based on either theoretical predictions or on empirical findings. Issues are classified as methodological if they are of a substantive nature, i.e., these issues specify how the objective of measuring the trade-off between quantity and quality of life is attained. They are classified as procedural if they are more related to the appearance and structure of the experiment. Finally, issues related to analyses of the results are classified as analytical issues.

Information about how TTO study characteristics affect values was extracted from the literature by AA and YEP. We have mainly drawn upon our extensive knowledge of this literature and previous review articles, such as Arnesen and Trommald [2] and Attema and Brouwer [11]. However, we do not claim to have covered all existing literature, since no systematic literature review was performed.

**Table 1** Methodological aspects

Methodological question	Specification	Not applicable/ not specified
1. What value range was assessed?	<input type="checkbox"/> Both states BTM and WTD <input type="checkbox"/> Only states BTM	<input type="checkbox"/>
2. What method was used for valuation of worse than dead states?	<input type="checkbox"/> Classical TTM <input type="checkbox"/> MVH protocol <input type="checkbox"/> Lead-time TTM <input type="checkbox"/> Lag-time TTM <input type="checkbox"/> Composite TTM <input type="checkbox"/> Other	<input type="checkbox"/>
3. What was the disease duration?	<input type="checkbox"/> 10 years <input type="checkbox"/> Life time <input type="checkbox"/> Other	<input type="checkbox"/>
4. Was the smallest tradable unit listed?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
5. Was the lead or lag time listed?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/>
6. What iteration procedure was used?	<input type="checkbox"/> Bisection <input type="checkbox"/> MVH fixed sequence <input type="checkbox"/> Other	<input type="checkbox"/>
7. What was the response scale?	<input type="checkbox"/> Years lived in full health <input type="checkbox"/> Years lived in impaired health	<input type="checkbox"/>

## Results

In total, 25 characteristics of TTM were derived that may influence the valuations obtained by the TTM method (Tables 1, 2, 3). The greatest variation across TTM studies has been observed in methodological and procedural aspects. The effect of procedural aspects on the obtained values is often not clear. For methodological aspects, more information is available. Below, we discuss each of these factors in turn, including the predicted effects of these factors on the results, based on both theory and empirical studies.

### Methodological aspects

Table 1 summarizes methodological aspects of a TTM study, which are discussed below.

#### *Value range spanned*

One of the most debated characteristics of TTM methodology is the strategy for evaluation of worse than dead (WTD) states. This issue arises because the method outlined above is only able to elicit positive values, i.e., values of better than dead (BTM) states, since negative life years are obviously not possible.<sup>1</sup> Hence, a separate method is

needed for health states that bring negative utility, if researchers wish to estimate both BTM and WTD states.<sup>2</sup>

Tilling et al. [12] reviewed the literature to see how this issue was addressed. They found that valuation of WTD states is often not pursued, but when it is, most often the MVH protocol is followed. The MVH procedure uses a sorting question to find out if a respondent considers a disease state BTM or WTD: a respondent chooses between immediate death and living ( $t$ ) years in the disease state. In case the respondent would opt for immediate death, this would indicate the respondent considered the health state [at least when lasting for ( $t$ ) years] to be WTD. Since dead is usually assigned a value of 0, this health state should have a negative utility. Subsequently the *WTD procedure* is started, where immediate death is compared to a health profile consisting of both a period in full health and a period of equal length in the disease state to be valued [13]. If the respondent still prefers immediate death, the period in full health is lengthened and the period in the disease state shortened, and vice versa. In this way, it is possible to elicit any negative utility between zero and minus infinity; in practice, the lowest attainable value depends on the *smallest tradable unit*. When the unit is small, very negative numbers, possibly even approaching minus infinity may be observed. However, because of the conceptual and

<sup>1</sup> We will term this method “classical TTM” from now on.

<sup>2</sup> The MVH WTD procedure has never been labeled and hence will just be termed ‘MVH WTD procedure’ in this article.

**Table 2** Procedural aspects

Procedural question	Specification	Not applicable/ not specified
1. What was the mode of administration?	<input type="checkbox"/> Face-to-face interviews <input type="checkbox"/> Group interviews <input type="checkbox"/> Self-administered questionnaire <input type="checkbox"/> Internet experiment <input type="checkbox"/> Other	<input type="checkbox"/>
2. Were visual aids used?	<input type="checkbox"/> Yes, TTO board and health state cards <input type="checkbox"/> Yes, computer assisted <input type="checkbox"/> No	<input type="checkbox"/>
3. What context effects were considered?	<input type="checkbox"/> Warm-up task: TTO <input type="checkbox"/> Warm-up task: other utility valuations <input type="checkbox"/> Starting point bias <input type="checkbox"/> Ordering of states <input type="checkbox"/> Other	<input type="checkbox"/>
4. What was the sampling frame?	<input type="checkbox"/> General population <input type="checkbox"/> Patients <input type="checkbox"/> Health-care providers <input type="checkbox"/> Significant others <input type="checkbox"/> Other	<input type="checkbox"/>
5. Were all health state values observed?	<input type="checkbox"/> Yes (direct valuations) <input type="checkbox"/> No (indirect valuations)	<input type="checkbox"/>
<i>If indirect valuations were used</i>		
6. Which modeling approach was used?	<input type="checkbox"/> Multi-attribute utility theory <input type="checkbox"/> Statistical inference	<input type="checkbox"/>
7. How many respondents were included?	<input type="checkbox"/> <200 <input type="checkbox"/> ≥200 <input type="checkbox"/> ≥400 <input type="checkbox"/> ≥800	<input type="checkbox"/>
8. Number of health states to be valued	<input type="checkbox"/> <50 <input type="checkbox"/> >50	<input type="checkbox"/>
9. How were health states selected? (multiple answers possible)	<input type="checkbox"/> Covering severity range <input type="checkbox"/> Orthogonality <input type="checkbox"/> Health state plausibility <input type="checkbox"/> Other	<input type="checkbox"/>
10. Lowest health state in the valuation task	<input type="checkbox"/> PITS <input type="checkbox"/> Other	<input type="checkbox"/>
11. Model fit criteria	<input type="checkbox"/> Mean absolute error (MAE) <input type="checkbox"/> Root mean square error (RMSE) <input type="checkbox"/> Other	
<i>If direct valuations were used</i>		
12. How was the health state described?	<input type="checkbox"/> Own health <input type="checkbox"/> In generic terms <input type="checkbox"/> In disease-specific terms	<input type="checkbox"/>
13. How were the health state descriptions generated? (multiple answers possible)	<input type="checkbox"/> Extracted from an existing questionnaire <input type="checkbox"/> Expert experience <input type="checkbox"/> Literature <input type="checkbox"/> Other	<input type="checkbox"/>

**Table 2** continued

Procedural question	Specification	Not applicable/ not specified
14. How was the text of the description structured?	<input type="checkbox"/> Narrated with label <input type="checkbox"/> Narrated without label <input type="checkbox"/> Bulleted with label <input type="checkbox"/> Bulleted without label	<input type="checkbox"/>

**Table 3** Analytical aspects

Analytical question	Specification	Not applicable/ not specified
1. What exclusion criteria were used?	<input type="checkbox"/> Feasibility questionnaire <input type="checkbox"/> Indifference between health states <input type="checkbox"/> Number of iterations <input type="checkbox"/> Trade-off, non-traders <input type="checkbox"/> Logical errors, ordering errors	<input type="checkbox"/>
2. How was the best possible health defined?	<input type="checkbox"/> Absence of a disease <input type="checkbox"/> Perfect health	<input type="checkbox"/>
3. How were WTD values analyzed?	<input type="checkbox"/> Transformation $x/(1-x)$ <input type="checkbox"/> Analyses of the medians <input type="checkbox"/> Other	<input type="checkbox"/>
4. Was the TTO adjusted for time preference?	<input type="checkbox"/> Yes, by separately eliciting time preferences <input type="checkbox"/> Yes, by including multiple time frames <input type="checkbox"/> Yes, by using one fixed discount rate <input type="checkbox"/> Yes, by performing both a lead- and lag-time TTO <input type="checkbox"/> No	<input type="checkbox"/>

practical problems this evoked, especially the need for two different kinds of questions for BTD and WTD states, a better alternative has been called for [12]. Tilling et al. identified three alternatives: lead-time TTO, lag-time TTO and chained TTO. The *lead-time and lag-time* approaches [3] were identified as the most promising.

Adding lead (lag) time in full health to both options of the TTO allows for a uniform procedure to value BTD and WTD states, which is the key theoretical advantage of the lead- and lag-time approaches. It introduces new methodological questions, such as those related to the *length of the lead (lag) time*, the *length of the disease time*, and the ratio between the lead (lag) time and disease time. Pilot studies suggested that lead-time TTO is susceptible to a framing effect that also affects BTD values (dragging them down in comparison to values observed in classical TTO) [14–16]. This effect worsens with lengthening of the lead time relative to the disease time.

#### Time frame

The total time frame, composed of disease duration and lead or lag time, if applicable, is important because TTO

values are often found to vary with it [11, 17, 18], notwithstanding the fact that the QALY model predicts them to be independent of the time frame, as implied by the condition of constant proportional trade-offs (CPTO). However, no systematic effect appears from empirical studies. A tendency for TTO values to decrease with time frame is found, but a lot of mixed evidence and even studies reporting an increasing relationship also exist, making it hard to reach definitive conclusions [17]. Most TTO studies use a 10- or 20-year time frame [2], less often actuarial life expectancy is used [18, 19], and sometimes respondents' own life expectation is used [20–22]. A clear-cut answer to the question of which of these strategies is most appropriate does not exist, because a trade-off must be made between the desire to use a realistic time frame for a condition and minimization of distorting effects, such as loss aversion and time preference [23]. One explanation for violations of the CPTO is a positive, non-exponential, *time preference*. This brings us to the issue of whether or not, and, if so, how, to deal with time preferences, which is addressed below (see analytical issues).

### Iteration procedure

The general opinion is that some kind of choice-based iteration process works better than directly asking for an indifference value [24]. Therefore, we only focus on this possibility in this article. Within this approach, there are several variants. One is to use a bisection procedure. A second is to add or detract one unit at each consecutive question, also known as top-down titration, as was done in the MVH study [9]. The main difference between these two variants is that the number of iterations is fixed for the former, whereas it depends on the respondent's answers in the latter. Respondents may be aware of this property of top-down titration and try to answer strategically in order to finish the experiment sooner [12]. Next to bisection and top-down titration [25], a ping-pong approach can be used. Health state valuations are between 0.10 and 0.15 higher with titration compared to ping-pong [26]. The highest attainable value is 1.0 if non-trading is allowed, and some smaller number if non-trading is not allowed or if the iteration procedure does not allow for an indifference value of exactly 1 [16].

The iteration procedure is further characterized by its *first questions*. While variables such as these starting points may be considered irrelevant to people's perception of health states, Samuelsen et al. [27] reported that TTO values are influenced by anchoring. Specifically, this study reported an upward shift of values with higher starting points. It is common to start with a comparison of ( $t$ ) years in the disease state to ( $t$ ) years in full health. Because the latter option dominates the former, it would then be rational to choose the latter. If the respondent instead chooses the former option, it may indicate s/he has not (yet) understood the question, and hence it may be a natural way to test for reliability. If the respondent chooses the dominant option, a logical follow-up question is to let the respondent choose between immediate death and ( $t$ ) years in the disease state. This allows for a division into BTD and WTD states, although a problem with this approach is that it does not take into account the possibility of maximum endurable time states, i.e., states that give a positive utility for some period  $x < t$ , but a negative utility afterwards, such that the overall utility after ( $t$ ) years is negative [28–31].

### Response scale

The response scale that is applied in virtually all TTO studies is duration, because duration is a quantitative variable, and quality of life is by definition qualitative (i.e., it needs some kind of description), making it difficult to let respondents give an answer in terms of quality of life. A remaining issue is which duration to use as the response

variable. Most researchers use the duration in full health for this, but it may also be the duration in the disease state. In the latter case, one may fix the duration in full health and ask for the duration in the disease state that renders indifference [6, 7, 32–35]. This generally causes TTO values to become much lower, so this is an important decision to be made. Furthermore, it is possible to choose another health state than full health as the anchor state against which to value a particular health state [36]. However, proper valuation then first requires the assignment of a value to this alternative anchor state. This can only be achieved if eventually full health is used as the anchor state, implying the use of full health as anchor state cannot be avoided. In order to prevent the additional effort of having to perform multiple tasks, as well as the expected biases resulting from chaining [37], it would therefore be advisable to directly anchor on full health.

### Procedural aspects

We now turn to a discussion of the procedural aspects that are illustrated in Table 2.

### Mode of administration

The mode of administration will also influence the results [38]. The most preferable mode is to use personal interviews, which are the most expensive as well. The advantage is that interaction between interviewer and respondent promotes good data quality. Disadvantages are the high cost and possible interviewer effects. TTO studies have increased in size over time, often necessitating the participation of multiple interviewers. The effort that is made to minimize interviewer differences is therefore relevant (e.g., training, availability of an interview script and intervision). Moreover, the interviewer help may lead to interviewer bias such as socially desirable bias and acquiescence bias. For example, respondents find it easier to agree than to disagree [39]. Even small verbal reinforcements have been shown to lead to different reactions of respondents [40].

Internet experiments have emerged as a way to obtain large representative data sets against relatively low costs [41, 42]. However, Internet experiments do not allow the researcher to monitor the effort put forward by the respondent, nor do they give the respondent the opportunity to ask questions for clarification or feedback. Versteegh et al. [41] in this issue report that Internet studies can be problematic for eliciting TTO tariffs. In-between these two is the group experiment, where sessions with small groups are run, with one experimenter present for about each 4–10 respondents. After a plenary description of the purpose of the experiment, the respondents can then answer with the experimenters walking around and answering questions if

needed. Although these studies have been shown to be feasible [43, 44], this method also seems less favorable than personal interviews.

### *Visual aids*

Investigators of TTO tend to have a preference for the use of graphs/illustrations to present the choice situation, since it appears that respondents find this easier than a numerical description [45, 46]. In the old days, TTO boards were commonly used. Today, the norm is computer-assisted personal interviews, because they promote correct implementation of the iteration procedure as well as a graphical illustration of the tasks. The *visual presentation* still varies between studies, which may influence results [47]. Often, a screen-shot of valuation software or applied visual presentation is requested during the peer-review procedure for the publication of results of TTO studies.

### *Context effects*

People tend to learn during a TTO experiment [48]. This is typically dealt with by inclusion of a warm-up task. TTO applications differ in the efforts put in to familiarize respondents with both the tasks and the health problems under consideration. Common *warm-up tasks* are TTO questions using different health states or valuation of the same health states using *different* valuation techniques (such as the visual analog scale, a ranking task, discrete choices or best-worse scaling) [41, 49, 50]. A further concern is the *order of health states* that a respondent has to value. Randomization is common practice, but more research into the most appropriate strategy may be warranted. Pinto Prades found in a recent study that the precision of health state values is contingent on *ordering of the states* [51]; more precise values are obtained when a TTO sequence begins with a mild state rather than a severe state.

### *Sampling frame*

It is generally recognized that the value of a health state varies with the sampling frame. Economic evaluations in the setting of health care are recommended to be made from the social perspective. Organizations involved in developing guidelines on the use of new and existing treatments, such as the National Institute for Health and Clinical Excellence (NICE), the panel of the US Public Health Service and the Dutch Health Care Insurance Board (CvZ), prefer health state values elicited from a fully informed representative sample of members of the public [8, 52, 53]. It might be challenging to fully inform members of the public. Instead, there are good arguments to use a patient sample, because these people are more familiar

with the symptoms of the disease than non-patients. The panel of the US Public Health Service already suggested that in economic evaluations in which alternative interventions are compared patients' preferences might be the better choice [52]. However, when investigating a patient sample, one should be aware that adaptation and/or strategic misrepresentation may influence valuation estimates [54]. Values shaped by adaptation typically lead to smaller effect sizes in the valuation of quality of life-enhancing treatments [55]. On the other hand, the influence of adaptation will differ between health states, and it provides valid information about the perceived severity of a health state. For instance, people might better adapt to physical impairments compared to mental diseases such as depression or skin diseases such as eczema.

### *Indirect valuation*

Health state valuation methods such as TTO may be used to value the health states of a health state classification system, such as the EQ-5D, the SF-6D or disease-specific questionnaires. Most classification systems contain too many health states to value all of them, and so values are elicited only for a subset. A modeling approach is used to estimate values for all health states. Modeling may be based on multi-attribute utility theory (such as with the Health Utility Index, HUI) or statistical inference. Both approaches are built on different assumptions and come with different requirements with regard to the subset of states that have to be valued directly. In the comparison of TTO values elicited from different experiments, comparing the health state selection and modeling efforts may be relevant.

When modeling is based on statistical inference, regression analysis is applied to estimate values for all health states on the basis of the subset of state observed. The impact of regression assumptions on the predicted values is greater in the case of extrapolation (outside the range of values in the data set) than in the case of interpolation (within the range of values in the data set). Therefore, it is relevant to report the *worst health state* offered to respondents in the valuation study. Furthermore, prediction intervals and goodness of fit criteria ought to be reported.

There has been little guidance to researchers about state selection, resulting in an unclear state of play. Researchers have considered covering the severity range, orthogonality and health state plausibility, but practice varies. A further issue is how many states need to be valued. Based on theory and observations, Lamers et al. [56] suggest that a minimum number of respondents per health state is required (for TTO approximately  $n = 100$ ) and that in principle adding more states (each assessed by 100 respondents) leads to more

information, hence more precise regression estimates, than increasing the number of respondents per health state. But good results have also been obtained valuing many health states with few observations per state [57]. Bagust [58] has recently argued that state selection may be improved by adopting more criteria for state selection, such as health state relevance and direct coverage of simple increments in health. Versteegh et al. [44] argue that the statistically most interesting set of health states may not be the set of health states that occurs most often in patients and show that the inclusion of the states that occur most in patients affects modeled health state values. Whatever the selection method, the selection may still result in a number of health states that is too high to value for an individual. A common solution then is to use a *blocked design*, including only a part of the subset of health states in each individual's questionnaire, while making sure all health states of the subset are valued by a sufficient number of respondents.

In blocking the design, a concern may be obtaining a low anchor. The worst possible health state in a classification scheme is the health state where all dimensions are at their worst possible level, in other words, having severe problems on all dimensions. This state is called the PITS state. This is state 55555 in the EQ-5D-5L system. It would be advisable to include this health state in the valuation task in order to have a lower anchor, but also in this regard practice has varied. Moreover, it is essential to list the *number of health states* that were valued (overall and per respondent) and sample size, as these characteristics may also affect the predictive quality of the regression model.

When only directly observed TTO values are used rather than modeled ones, the above concerns do not apply. Direct health-state valuations could be used when a limited number of health states have to be valued, e.g., to obtain health state values to health states presented in a Markov model. This approach generates another set of methodological concerns, e.g., related to how the disease state is described (generic or disease specific terms), narrated or bulleted, labeled or unlabeled [59]. Health state descriptions are developed based on literature, on expert experience or using classification systems such as the EQ-5D, SF-6D and HUI. Health state descriptions need to be specific to ensure respondents are fully informed, but also restrictive to avoid information overload. Evidence of the impact of such choices is limited. Two studies found that the exact labeling and framing of the health description did not seem to affect respondents' valuations [60], nor did the sparseness of an EQ-5D health state description [61].

Part of direct health-state valuations are health-state valuations of the own health. This avoids the need to describe health, since the person experiencing the health problem is also the one valuing it [59]. However, health state valuations of the own health are difficult to interpret because

of the lack of clarity about the health problem, e.g., respondents tend to value their whole life including minor positive [62] and minor negative events [63]. Direct health state valuations of the own health are preferred when researchers want to incorporate the effect of adaptation, for instance, cost-effectiveness analyses of psychological interventions. Direct health state valuations of the own health are also preferred for psychological illnesses [64, 65].

#### Analytical aspects

This section ends with a consideration of several analytical aspects, as shown in Table 3.

#### Exclusion criteria

Several criteria can be used to exclude respondents from the analysis. One can exclude respondents who: (1) indicated they did not understand the task on a feasibility questionnaire, (2) did not differentiate between any of the different health states, (3) used only a limited number of iterations for all health states [41], (4) did not trade off any time at all (non-traders) and (5) who rated mild health states lower than severe health states [66]. Some researchers apply criterion 4 and exclude non-traders in their analysis [67–69], but this is not common practice. An average proportion of 57 % of non-traders has been reported by Arnesen and Trommald [1]. In general, cluster effects like non-trading behavior are a direct result of the desire to derive the value of a QALY by means of a trade of life years, since for some people the value of life approaches infinity. This point of view makes exclusion of non-traders inappropriate. Also criterion 5 is argued against, as its use may result in the exclusion of up to half of the respondents, and preference reversals may just indicate uncertainty. Consequently, researchers should beware of selection bias. For instance, Arnesen and Norheim [70] report that aspects of life such as having children, friends and social esteem in many cases has a higher impact than the health problem being studied. Moreover, people with lower education levels have a higher propensity to be non-traders [21, 49]. It might be questioned if researchers should exclude non-traders and respondents who misorder health states, although some researchers argue that non-traders need to be challenged by additional questions involving smaller trade units [71].

#### Definition of anchor points

In applications of TTO, the best possible health state that gets a utility of 1 is not always explicitly defined. The state that receives a utility of 1 typically is the state where the health problems for which the value is sought are absent.



When using a classification scheme, such as the EQ-5D-5L, this is the health state where all dimensions are at the best possible level (i.e., no problems on any dimension; 11111 in case of the EQ-5D-5L). To avoid lengthy health state descriptions, this health state is often termed ‘full health’ or ‘perfect health.’ Care should be taken that having no problems on any of the dimensions of the description system is not necessarily the same as being perfectly healthy. For instance, the five dimensions of the EQ-5D do not capture all possible health impairments. Hence, health state 11111 does not by definition have a utility of 1 in the sense of living without any health problems. Using absence of disease instead of perfect health in cost-utility analyses seems to make health interventions appear less costly and more effective [72], although the effect on the TTO is inconclusive [73, 74].

#### *Analysis of WTD values*

Whenever a TTO is used, it is important to know how the analyst handles values of states considered WTD. Because those values can theoretically become minus infinity, one of them can already heavily influence the average TTO value [75]. Applications of TTO differ in what the lowest value is that can be achieved and in how extremely negative values are handled. Where lead-time TTO explicitly defines the observed value range, the MVH approach to estimating WTD values implied that the lowest achievable value was defined by the selection of the smallest unit of time that could be traded off. In most valuation studies for EQ-5D-3L, this unit was 3 months; correspondingly, the lowest achievable value was  $-39$ . Researchers have proposed and adopted a broad range of strategies to deal with extremely negative values. Negative values in TTO are often transformed in one way or another. A common transformation is  $x/(1-x)$ , constraining WTD values to  $-1$  [76]. Alternative strategies could be to report medians instead of means or to model the data differently, i.e., on the basis of a different economic or mathematical model [77, 78].

Lead-time TTO represents an alternative way to handle negative values: by setting the ratio between lead-time TTO and disease time, the scale of observed values is explicitly defined. For example, when both the lead time and disease time are 10 years, the ratio is 1:1. The lowest possible response, i.e., declaring immediate death to give as much utility (0) as 10 years in full health followed by 10 years in the disease state then indicates  $10 \times 1 + 10 \times x = 0$ , i.e.,  $x = -1$ . This does resolve the issue of very negative values; however, it comes at the cost of WTD values being censured to  $-1$ . One solution to this constraint is to extend the lead time and thus to modify the ratio of lead time to disease time, enabling lower

minimum WTD values. But this strategy is not expected to remove the problem. The piloting studies of lead- and lag-time TTO indicate that a significant fraction of the sample expresses preferences at the very bottom of the scale, even for high lead-time to disease-time ratios. Devlin et al. [14] therefore attempt to tackle this problem analytically by applying survival analysis to model their values. Given the variety of approaches that can be adopted, researchers should report the range of values that is explored and the analytical methods that were adopted to deal with extremely negative values.

#### *Time preference*

As TTO values are affected by time preferences, adjustment of observed TTO values can be considered. Investigators of TTO often just neglect time preferences, but this causes an underestimation of TTO scores [79]. Adjusting TTO valuation for the influence of time preference can be done by: (1) separately eliciting time preferences and using these estimates to correct the initial TTO values [80–85], (2) including multiple time frames in the TTO [86, 87], (3) correcting all TTO values using one fixed discount rate [88] or (4) performing both a lead-time and lag-time TTO [89]. Concerning (1), several methods exist to elicit time preferences, including riskless and risky (often certainty equivalence) methods. The issue of time preference is particularly important for the lead- and lag-time TTOS since these involve a longer horizon and hence are more susceptible to discounting.

Many researchers consider the measurement of time preference as rather problematic [15]. The required methods are often not up to the task. Therefore, although correcting for discounting is theoretically attractive, it is not very practical to do so. This stresses the importance of developing time preference elicitation methods that are more feasible [81] and to adopt a standardized time preference elicitation protocol alongside a standardized TTO protocol, at least for TTO studies that lie at the heart of HTA submissions.

## **Discussion**

This article has investigated and explored differences in TTO studies with the aim to increase understanding of how differences in methodology between studies may affect comparability. The overview makes clear that for most characteristics of TTO, best practices cannot be defined unambiguously. Our aim is not to produce guidance on how TTO studies ought to be conducted. Instead, our goal is to raise increased awareness and understanding of the effects of different TTO factors that imply a need for

standardization. In addition, this overview may facilitate explorations into which factors are most likely to receive the broadest acceptance.

Although drafting of guidelines was not the aim of this article, exploration of differences in how studies are conducted such as presented here may be at the heart of future developments in the area of harmonization, because we may learn what works and what does not work from existing differences in TTO studies. Sometimes we have been able to identify a best practice; on other occasions we have highlighted areas of TTO where ambiguity remains about best practice. This can be used to put together an agenda for methodological research in the area of TTO.

This study has been conducted against the background of the development of a TTO health-state valuation protocol for EQ-5D-5L valuation studies. Developing a protocol serves two goals: reducing method variation across valuation studies and dealing with perceived shortcomings in previous valuation studies. Recognizing that TTO methodology is far from standardized [2] and that none of the adopted TTO approaches may count on general acceptance to be considered a standard, the aim of the research program has been to compare the benefits of innovative solutions to existing shortcomings. The developmental process of the valuation protocol for EQ-5D-5L studies reported in this issue of the journal comprised of a series of methodologically oriented studies, all with a slightly different objective. Key identified issues for TTO are the WTD estimation approach and the effect of mode of administration on data quality. While the data quality concerns are currently dealt with by offering a mix of services (interviewer training, protocols, logistic support, data quality control tools), it appeared impossible to find an unambiguous solution for assessing the values of states that are considered worse than dead. Lead-time TTO may be theoretically sound but in practice suffers from a framing effect, which makes it necessary to shape this approach on the basis of arbitrary grounds. The current protocol therefore develops a status quo that serves to promote comparability of studies for the forthcoming years, although it should not stop evaluation of alternatives.

Since Stalmeier et al. [10] published their checklist on TTO, much progress has been made in the area of health-state valuation. However, none of the methods available for health-state valuation can claim to be the widely or universally accepted method. As such, the search for alternative methods continues. One innovative approach is the use of discrete choice experiments to collect response data that can be used to derive health-state values, such as proposed by Bansback et al. [90]. This approach resembles TTO in the respect that health states are valued in a trade-off between length of life and quality of life, but iteration is avoided. Instead, choice models are applied to responses

derived from discrete choices about trade-offs between length and quality of life. We support experimentation with this method and are keen to learn to what extent it can resolve problems in TTO.

This article highlighted how factors in the TTO method may affect the elicited values and therefore restrain the comparability of results from different studies. We agree with Arnesen and Trommald that the current use of the TTO should not be regarded as the use of one specific method [2], and values need to be discussed in relation to how they were assessed, as previously emphasized by Stalmeier et al. [10]. Researchers using the TTO need to be aware of these effects when comparing their results with related literature using the TTO. This is not only a task of researchers but also of peers reviewing papers and editors. However, we feel that the responsibility of the research community stretches beyond that: our conviction is that efforts need to be made to reduce practice variation in TTO studies. As this article revealed that for most characteristics of TTO best practices cannot be defined unambiguously, guidance must be developed in such a way that a balance is found between the pros and cons of the different TTO approaches.

## Conclusion

The presented literature overview highlights the need for harmonization. By listing characteristics of TTO studies that affect the obtained values, our checklist offers support to those who might eventually attempt to bring convergence into TTO study practices.

**Acknowledgments** This research was supported by the EuroQol Group. Elly Stolk and Matthijs Versteegh disclose that they are members of the EuroQol Group, a not-for-profit group that develops and distributes instruments to assess and value health. We thank Paul Krabbe and Nancy Devlin for comments on a previous version of this article.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

1. Arnesen, T., Trommald, M.: Roughly right or precisely wrong? Systematic review of quality-of-life weights elicited with the time trade-off method. *J. Health Serv. Res. Policy* **9**, 43–50 (2004)
2. Arnesen, T., Trommald, M.: Are QALYs based on time trade-off comparable?—A systematic review of TTO methodologies. *Health Econ.* **14**, 39–53 (2005)
3. Robinson, A., Spencer, A.: Exploring challenges to TTO utilities: valuing states worse than dead. *Health Econ.* **15**, 393–402 (2006)

4. Macran, S., Kind, P.: "Death" and the valuation of health-related quality of life. *Med. Care* **39**, 217–227 (2001)
5. Brazier, J., Ratcliffe, J., Salomon, J.A., Tsuchiya, A.: *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford University Press (2007)
6. Bleichrodt, H., Pinto, J.L., Abellán-Perpinán, J.M.: A consistency test of the time trade-off. *J. Health Econ.* **22**, 1037–1052 (2003)
7. Attema, A.E., Brouwer, W.B.F.: Can we fix it? Yes we can! But what? A new test of procedural invariance in TTO-measurement. *Health Econ.* **17**, 877–885 (2008)
8. NICE: Guide to the methods of technology appraisal. <http://www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/guidetothemethodsoftechnologyappraisal.jsp> (2008). Accessed 26 Apr 2013
9. Dolan, P.: Modeling valuations for EuroQol health states. *Med. Care* **35**, 1095–1108 (1997)
10. Stalmeier, P.F.M., Goldstein, M.K., Holmes, A.M., Lenert, L., Miyamoto, J., Stiggelbout, A.M., Torrance, G.W., Tsevat, J.: What should be reported in a methods section on utility assessment? *Med. Decis. Making* **21**, 200–207 (2001)
11. Attema, A.E., Brouwer, W.B.F.: On the (not so) constant proportional tradeoff in TTO. *Qual. Life Res.* **19**, 489–497 (2010)
12. Tilling, C., Devlin, N., Tsuchiya, A., Buckingham, K.: Protocols for time trade off valuations of health states worse than dead: a literature review. *Med. Decis. Making* **30**, 610–619 (2010)
13. Torrance, G.W.: Measurement of health state utilities for economic appraisal. *J. Health Econ.* **5**, 1–30 (1986)
14. Devlin, N.J., Buckingham, K., Tsuchiya, A., Shah, K., Tilling, C., Wilkinson, G., van Hout, B.A.: A comparison of alternative variants of the lead and lag time TTO. *Health Econ.* **22**, 517–532 (2013)
15. Devlin, N., 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.* **20**, 348–361 (2011)
16. Attema, A.E., Versteegh, M.M., Oppe, M., Brouwer, W.B.F., Stolk, E.A.: Lead time TTO: leading to better health state valuations? *Health Econ.* **22**, 376–392 (2013)
17. Attema, A.E., Brouwer, W.B.F.: Constantly proving the opposite? A test of CPTO using a broad horizon and correcting for discounting. *Qual. Life Res.* **21**, 25–34 (2012)
18. Stiggelbout, A.M., Kiebert, G.M., Kievit, J., Leer, J.W., Habbema, J.D., De Haes, J.C.: The "utility" of the time trade-off method in cancer patients: feasibility and proportional trade-off. *J. Clin. Epidemiol.* **48**, 1207–1214 (1995)
19. Essink-Bot, M.L., Stuijbergen, M.C., Meerding, W.J., Looman, C.W., Bonsel, G.J.: Individual differences in the use of the response scale determine valuations of hypothetical health states: an empirical study. *BMC Health Serv. Res.* **7**, 62 (2007)
20. Heintz, E., Krol, M., Levin, L.A.: The impact of patients' subjective life expectancy on time trade-off valuations. *Med. Decis. Making* **33**, 261–270 (2013)
21. van Nooten, F.E., Koolman, X., Brouwer, W.B.F.: The influence of subjective life expectancy on health state valuations using a 10 year TTO. *Health Econ.* **18**, 549–558 (2009)
22. Kattan, M.W., Fearn, P.A., Miles, B.J.: Time trade-off utility modified to accommodate degenerative and life-threatening conditions. In: *Proceedings of AMIA. Annual symposium*, 304–308 (2001)
23. Attema, A.E., Brouwer, W.B.F.: In search of a preferred preference elicitation method: a test of the internal consistency of choice and matching tasks. Erasmus University Rotterdam. [http://www.bmg.eur.nl/personal/attema/PrefRev\\_2013.pdf](http://www.bmg.eur.nl/personal/attema/PrefRev_2013.pdf) (2013). Accessed 8 April 2013
24. Bostic, R., Herrnstein, R.J., Luce, R.D.: The effect on the preference-reversal phenomenon of using choice indifferences. *J. Econ. Behav. Organ.* **13**, 193–212 (1990)
25. Delquié, P.: "Bi-matching": a new preference assessment method to reduce compatibility effects. *Manage. Sci.* **43**, 640–658 (1997)
26. Lenert, L.A., Cher, D.J., Goldstein, M.K., Bergen, M.R., Garber, A.: The effect of search procedures on utility elicitation. *Med. Decis. Making* **18**, 76–83 (1998)
27. Samuelsen, C.H., Augestad, L.A., Stavem, K., Kristiansen, I.S.: Rand-Hendriksen. Anchoring bias in the lead-time time trade-off (2012)
28. Stalmeier, P.F.M., Chapman, G.B., de Boer, A.G.E.M., van Lanschot, J.J.B.: A fallacy of the multiplicative QALY model for low-quality weights in students and patients judging hypothetical health states. *Int. J. Technol. Assess. Health Care* **17**, 488–496 (2001)
29. Stalmeier, P.F., Bezembinder, T.G., Unic, I.J.: Proportional heuristics in time tradeoff and conjoint measurement. *Med. Decis. Making* **16**, 36–44 (1996)
30. Unic, I., Stalmeier, P.F., Verhoef, L.C., van Daal, W.A.: Assessment of the time-tradeoff values for prophylactic mastectomy of women with a suspected genetic predisposition to breast cancer. *Med. Decis. Making* **18**, 268–277 (1998)
31. Dolan, P., Stalmeier, P.F.M.: The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. *J. Health Econ.* **22**, 445–458 (2003)
32. Attema, A.E., Brouwer, W.B.F.: The way that you do it? An elaborate test of procedural invariance of TTO, using a choice-based design. *Eur J Health Econ* **13**, 491–500 (2012)
33. Spencer, A.: The TTO method and procedural invariance. *Health Econ.* **12**, 655–668 (2003)
34. Krabbe, P.F., Bonsel, G.J.: Sequence effects, health profiles, and the QALY model: in search of realistic modeling. *Med. Decis. Making* **18**, 178–186 (1998)
35. Krabbe, P.F., Essink-Bot, M.L., Bonsel, G.J.: On the equivalence of collectively and individually collected responses: standard-gamble and time-tradeoff judgments of health states. *Med. Decis. Making* **16**, 120–132 (1996)
36. Jansen, S.J., Stiggelbout, A.M., Wakker, P.P., Nooij, M.A., Noordijk, E.M., Kievit, J.: Unstable preferences: a shift in valuation or an effect of the elicitation procedure? *Med. Decis. Making* **20**, 62–71 (2000)
37. Spencer, A.: The implications of linking questions within the SG and TTO methods. *Health Econ.* **13**, 807–818 (2004)
38. Norman, R., King, M., Clarke, D., Viney, R., Cronin, P., Street, D.: Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. *Qual. Life Res.* **19**, 499–508 (2010)
39. Bowling, A.: Mode of questionnaire administration can have serious effects on data quality. *J. Public Health* **27**, 281–291 (2005)
40. Hildum, D.C., Brown, R.W.: Verbal reinforcement and interviewer bias. *J. Abnorm. Soc. Psychol.* **53**, 108–111 (1956)
41. Versteegh, M.M., Attema, A.E., Oppe, M., Devlin, N.J., Stolk, E.A.: Time to tweak the TTO: results from a comparison of alternative specifications of the TTO. *Eur. J. Health Econ.* (Forthcoming)
42. Bansback, N., Tsuchiya, A., Brazier, J., Anis, A.: Canadian valuation of EQ-5D health states: preliminary value set and considerations for future valuation studies. *PLoS One* **7**, e31115 (2012)
43. Stolk, E.A., Busschbach, J.J.: Validity and feasibility of the use of condition-specific outcome measures in economic evaluation. *Qual. Life Res.* **12**, 363–371 (2003)
44. Versteegh, M.M., Leunis, A., Uyl-de Groot, C.A., Stolk, E.A.: Condition-specific preference-based measures: benefit or burden? *Value Health* **15**, 504–513 (2012)
45. Zikmund-Fisher, B.J., Fagerlin, A., Ubel, P.A.: A demonstration of "less can be more" in risk graphics. *Med. Decis. Making* **30**, 661–671 (2010)

46. Zikmund-Fisher, B.J., Ubel, P.A., Smith, D.M., Derry, H.A., McClure, J.B., Stark, A., Pitsch, R.K., Fagerlin, A.: Communicating side effect risks in a tamoxifen prophylaxis decision aid: the debiasing influence of pictographs. *Patient Educ. Couns.* **73**, 209–214 (2008)
47. Swinburn, P.: EQ-5D orientation study report produced for the EuroQol Group (2010)
48. Augestad, L.A., Rand-Hendriksen, K., Kristiansen, I.S., Stavem, K.: Learning effects in time trade-off based valuation of EQ-5D health states. *Value Health* **15**, 340–345 (2012)
49. Dolan, P., Gudex, C., Kind, P., Williams, A.: The time trade-off method: results from a general population study. *Health Econ.* **5**, 141–154 (1996)
50. Furlong, W., Feeny, D., Torrance, G., Barr, R., Horsman, J.: Guide to design and development of health-state utility instrumentation. (1992)
51. Pinto-Prades, J.L.: Imprecise preferences and order effects in the time trade-off. (2013)
52. Gold, M.R., Siegel, J.E., Russell, L.B., Weinstein, M.C. (eds.): *Cost-Effectiveness in Health and Medicine*. Oxford University Press, New York (1996)
53. Riteco, J.A., de Heij, L.J.M., van Luijn, J.C.F., Wolff, I.F.: *Richtlijnen voor Farmaco-Economisch Onderzoek*. College voor Zorgverzekeringen, Amstelveen (1999)
54. Dolan, P., Kahneman, D.: Interpretations of utility and their implications for the valuation of health. *Econ. J.* **118**, 215–234 (2008)
55. Menzel, P., Dolan, P., Richardson, J., Olsen, J.A.: The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Soc. Sci. Med.* **55**, 2149–2158 (2002)
56. Lamers, L.M., McDonnell, J., Stalmeier, P.F.M., Krabbe, P.F.M., Busschbach, J.J.V.: The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ.* **15**, 1121–1132 (2006)
57. Viney, R., Norman, R., King, M., Cronin, P., Street, D.J., Knox, S., Ratcliffe, J.: Time trade-off derived EQ-5D weights for Australia. *Value Health* **14**, 928–936 (2011)
58. Bagust, A.: Improving valuation sampling of EQ-5D health states. *Health Qual. Life Outcomes* **11**, 14 (2013)
59. Brazier, J., Rowen, D.: NICE DSU technical support document 11: alternatives to EQ-5D for generating health state utility values (2011)
60. Gerard, K., Dobson, M., Hall, J.: Framing and labelling effects in health descriptions: quality adjusted life years for treatment of breast cancer. *J. Clin. Epidemiol.* **46**, 77–84 (1993)
61. Peeters, Y., Stiggelbout, A.M.: Valuing health: does enriching a scenario lead to higher utilities? *Med. Decis. Making* **29**, 334–342 (2009)
62. Peeters, Y., Vliet Vlieland, T.P.M., Stiggelbout, A.M.: Focusing illusion, adaptation and EQ-5D health state descriptions: the difference between patients and public. *Health Expect.* **15**, 367–378 (2012)
63. Insinga, R.P., Fryback, D.G.: Understanding differences between self-ratings and population ratings for health in the EuroQOL. *Qual. Life Res.* **12**, 611–619 (2003)
64. Pyne, J.M., Fortney, J.C., Tripathi, S., Feeny, D., Ubel, P., Brazier, J.: How bad is depression? Preference score estimates from depressed patients and the general population. *Health Serv. Res.* **44**, 1406–1423 (2009)
65. Smith, D., Damschroder, L., Kim, S., Ubel, P.: What's it worth? Public willingness to pay to avoid mental illnesses compared with general medical illnesses. *Psychiatr. Serv.* **63**, 319–324 (2012)
66. Schiffman, R.M., Walt, J.G., Jacobsen, G., Doyle, J.J., Lebovics, G., Sumner, W.: Utility assessment among patients with dry eye disease. *Ophthalmology* **110**, 1412–1419 (2003)
67. Handler, R.M., Hynes, L.M., Nease Jr, R.F.: Effect of locus of control and consideration of future consequences on time tradeoff utilities for current health. *Qual. Life Res.* **6**, 54–60 (1997)
68. Churchill, D.N., Torrance, G.W., Taylor, D.W., Barnes, C.C., Ludwin, D., Shimizu, A., Smith, E.K.: Measurement of quality of life in end-stage renal disease: the time trade-off approach. *Clin. Invest. Med.* **10**, 14–20 (1987)
69. Churchill, D.N., Morgan, J., Torrance, G.W.: Quality of life in end-stage renal disease. *Perit. Dial. Bull.* **4**, 20–23 (1984)
70. Arnesen, T.M., Norheim, O.F.: Quantifying quality of life for economic analysis: time out for time tradeoff. *Medical Humanit.* **29**, 81–86 (2003)
71. Iezzi, A., Richardson, J.: *Measuring Quality of Life at the Centre for Health Economics*. Melbourne. (2009)
72. Fryback, D.G., Lawrence, W.F.: Dollars may not buy as many QALYs as we think: a problem with defining quality-of-life adjustments. *Med. Decis. Making* **17**, 276–284 (1997)
73. King Jr, J.T., Styn, M.A., Tsevat, J., Roberts, M.S.: “Perfect health” versus “disease free”: the impact of anchor point choice on the measurement of preferences and the calculation of disease-specific disutilities. *Med. Decis. Making* **23**, 212–225 (2003)
74. King, J.T., Tsevat, J., Roberts, M.S.: Impact of the scale upper anchor on health state preferences. *Med. Decis. Making* **29**, 257–266 (2009)
75. Lamers, L.M.: The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. *Med. Care* **45**, 238–244 (2007)
76. Patrick, D.L., Starks, H.E., Cain, K.C., Uhlmann, R.F., Pearlman, R.A.: Measuring preferences for health states worse than death. *Med. Decis. Making* **14**, 9–18 (1994)
77. Craig, B.M., Oppe, M.: From a different angle: a novel approach to health valuation. *Soc. Sci. Med.* **70**, 169–174 (2010)
78. Craig, B.M., Busschbach, J.: The episodic random utility model unifies time trade-off and discrete choice approaches in health state valuation. *Popul. Health Metr.* **7**, 3 (2009)
79. Attema, A.E., Brouwer, W.B.F.: The value of correcting values: influence and importance of correcting TTO scores for time preference. *Value Health* **13**, 879–884 (2010)
80. Attema, A.E., Brouwer, W.B.F.: The correction of TTO-scores for utility curvature using a risk-free utility elicitation method. *J. Health Econ.* **28**, 234–243 (2009)
81. Attema, A.E., Brouwer, W.B.F.: Deriving time discounting correction factors for TTO tariffs. *Health Econ.* doi:10.1002/hec.2921
82. Stiggelbout, A.M., Kiebert, G.M., Kievit, J., Leer, J.W., Stoter, G., de Haes, J.C.: Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. *Med. Decis. Making* **14**, 82–90 (1994)
83. Martin, A.J., Glasziou, P.P., Simes, R.J., Lumley, T.: A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *Int. J. Technol. Assess. Health Care* **16**, 137–147 (2000)
84. van der Pol, M., Roux, L.: Time preference bias in time trade-off. *Eur. J. Health Econ.* **6**, 107–111 (2005)
85. van Osch, S.M., Wakker, P.P., van den Hout, W.B., Stiggelbout, A.M.: Correcting biases in standard gamble and time tradeoff utilities. *Med. Decis. Making* **24**, 511–517 (2004)
86. Olsen, J.A.: Persons versus years: two ways of eliciting implicit weights. *Health Econ.* **3**, 39–46 (1994)
87. Gyrd-Hansen, D.: Comparing the results of applying different methods of eliciting time preferences for health. *Eur. J. Health Econ.* **3**, 10–16 (2002)
88. Bleichrodt, H., Johannesson, M.: Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *J. Health Econ.* **16**, 155–175 (1997)
89. Attema, A.E., Versteegh, M.M.: Would you rather be ill now, or later?. *Health Econ.* doi: 10.1002/hec.2894 (forthcoming)
90. Bansback, N., Brazier, J., Tsuchiya, A., Anis, A.: Using a discrete choice experiment to estimate health state utility values. *J. Health Econ.* **31**, 306–318 (2012)