

# EuroQol Protocols for Time Trade-Off Valuation of Health Outcomes

Mark Oppe<sup>1</sup> · Kim Rand-Hendriksen<sup>2,3</sup> · Koonal Shah<sup>4</sup> ·  
Juan M. Ramos-Goñi<sup>1</sup> · Nan Luo<sup>5</sup>

Published online: 15 April 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** The time trade-off (TTO) valuation technique is widely used to determine utility values of health outcomes to inform quality-adjusted life-year (QALY) calculations for use in economic evaluation. Protocols for implementing TTO vary in aspects such as the trade-off framework, iteration procedure and its administration model and method, training of respondents and interviewers, and quality control of data collection. The most widely studied and utilized TTO valuation protocols are the Measurement and Valuation of Health (MVH) protocol, the Paris protocol and the EuroQol Valuation Technology (EQ-VT) protocol, all developed by members of the EuroQol Group. The MVH protocol and its successor, the Paris

protocol, were developed for valuation of EQ-5D-3L health states. Both protocols were designed for a trained interviewer to elicit preferences from a respondent using the conventional TTO framework with a fixed time horizon of 10 years and an iteration procedure combining bisection and titration. Developed for valuation of EQ-5D-5L health states, the EQ-VT protocol adopted a composite TTO framework and made use of computer technology to facilitate data collection. Training and monitoring of interviewers and respondents is a pivotal component of the EQ-VT protocol. Research is underway aiming to further improve the EuroQol protocols, which form an important basis for the current practice of health technology assessment in many countries.

✉ Nan Luo  
nan\_luo@nuhs.edu.sg

Mark Oppe  
oppe@euroqol.org

Kim Rand-Hendriksen  
kim.rand-hendriksen@medisin.uio.no

Koonal Shah  
KShah@ohe.org

Juan M. Ramos-Goñi  
jramos@euroqol.org

<sup>1</sup> EuroQol Research Foundation, Rotterdam, The Netherlands

<sup>2</sup> Health Services Research Centre, Akershus University Hospital, Lørenskog, Norway

<sup>3</sup> Dept. of Health Management and Health Economics, University of Oslo, Oslo, Norway

<sup>4</sup> Office of Health Economics, London, UK

<sup>5</sup> Saw Swee Hock School of Public Health, National University of Singapore, 12 Science Drive 2, Block MD1, #11-01D, Singapore 117549, Singapore

## Key Points for Decision Makers

The time trade-off (TTO) technique has been widely used to obtain health state values for use in the economic evaluation of health technologies.

The use of different variants of TTO creates two practical problems: incomparability across studies and difficulty in choosing among the variants.

The EuroQol Group, a multi-disciplinary group of researchers from all over the world, has developed three standardized TTO-based protocols for valuation of EQ-5D health states: the MVH protocol, the Paris protocol and the EQ-VT protocol. For the valuation of the 3-level version of the EQ-5D (EQ-5D-3L), the Paris protocol is recommended. For the valuation of the 5-level version of EQ-5D (EQ-5D-5L), the EQ-VT protocol is recommended.

## 1 Introduction

Quality-adjusted life years (QALYs), a generic measure of health combining quality and length of life, are increasingly used to quantify outcomes in the economic evaluation of health technologies. According to the principles underpinning the QALY, the quality-of-life (QOL) weights, also described as health state utility values, must lie on an interval scale anchored at 0 (death) and 1 (full health) [1]. A second assumption underlying the basic QALY model is that the QOL weight for a given health state does not depend on how long it is experienced (constant proportional time preference) [2].

A variety of survey-based methods can be used to estimate these values based on the stated preferences of individuals of interest. Methods include standard gamble (SG), time trade-off (TTO), visual analogue scale (VAS), ranking exercises and, more recently, discrete-choice experiments (DCEs) (see Ryan et al. [3] for a systematic review of preference-elicitation techniques). Choice-based methods are typically preferred by economists, since these reflect the view that the value an individual places on something should be estimated by what they would be willing to forego to obtain it (or in the case on impaired health, in order to avoid it). The SG technique, which is grounded in expected utility theory, has traditionally been considered the gold standard valuation method. However, in general, people have difficulties assessing probabilities, but can relate more easily to time [1]. Hence, an alternative technique—TTO [4]—has been widely used to value different states of health. It has been shown that TTO and SG result in different values and that the total bias in TTO valuation is smaller than that in SG [5].

TTO can be implemented in numerous ways, although the approach that comes closest to obtaining the unobservable true utilities cannot be known. Over the last decades, a growing number of different variants of TTO have been used, which creates two practical problems: incomparability across studies and difficulty in choosing between the variants. Studies have shown that the design of the TTO task influences how respondents value health [6]. Therefore, values from studies using different TTO tasks may not be comparable [7]. The difficulty in choosing between the variants is due to the absence of a ‘gold standard’. While continuous scientific explorations are warranted to inform future choices about valuation techniques, the need for comparability suggests that standardization of the tasks used in valuation protocols is desirable. Analogous to the ‘reference case’ approach proposed by the US panel on cost-effectiveness on health and medicine [8], standardization can ensure the comparability of different studies and therefore consistent decision making.

Among currently available protocols, those developed by the EuroQol Group (<http://www.euroqol.org>) are the most widely studied and utilized. The EuroQol Group first met in 1987 to test the feasibility of jointly developing a standardized generic instrument for describing and valuing health-related QOL (HR-QOL). From the outset, the EuroQol Group has been multi-country, multi-centre, multi-disciplinary and not for profit. Over several years, the EuroQol Group developed a generic HR-QOL instrument, called the EQ-5D [9, 10]. It uses a standardized health state descriptive system consisting of five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression, each of which has three levels of severity. Since then, two additional versions of the EQ-5D have been created: the EQ-5D-Y, for use in children and adolescents aged 8–15 years [11]; and the EQ-5D-5L, with five severity levels per dimension [12]. In this paper, we refer to the original (3-level) version of EQ-5D as EQ-5D-3L.

In addition to the instruments, the EuroQol Group has also developed protocols for valuation of health states defined by its instruments. Three of these valuation protocols included TTO as the main valuation task: the Measurement and Valuation of Health (MVH) protocol [13]; the so called ‘Paris’ protocol [14] and the EuroQol Valuation Technology (EQ-VT) protocol [15]. All three protocols focus on the data-collection procedure using TTO. Because the EQ-5D and the TTO-based valuation studies are used in many different countries, the protocols do not include requirements for sampling frameworks or for modelling the collected TTO data, as these can vary by country. In addition, experimental design aspects, such as the algorithm for the selection of the health states in TTO experiments, are not part of the EuroQol Group’s valuation protocols.

The purpose of this paper is to provide a detailed introduction and critique of the three valuation protocols developed by the EuroQol Group focusing on TTO data collection, and to provide recommendations for their usage. We start with an overview of the different variants of the TTO that have been included in the EuroQol Group’s protocols and their most important characteristics. We then describe the three EuroQol valuation protocols. We conclude by discussing research that may help further improve the current EuroQol protocols.

## 2 The Time Trade-Off (TTO) Tasks

The objective of the TTO is to determine the length of lifetime the respondent would be willing to forego to live in a better health state (typically ‘full health’) and avoid living in a bad health state. This is achieved by presenting respondents with a series of choice tasks, each involving

two alternative hypothetical lives. The two lives are presented such that the respondent is forced to choose between a longer life in the health state of interest and a shorter life in better health. Depending on which life is chosen, the amount of time in better health is altered until a point of preferential indifference is reached. This constitutes the basic framework of all TTO tasks. However, the described framework can be implemented in a variety of ways. TTO implementations may differ in trade-off framework, iteration procedure and its mode and method of administration, training of respondents and interviewers and quality control of data collection.

## 2.1 TTO Frameworks

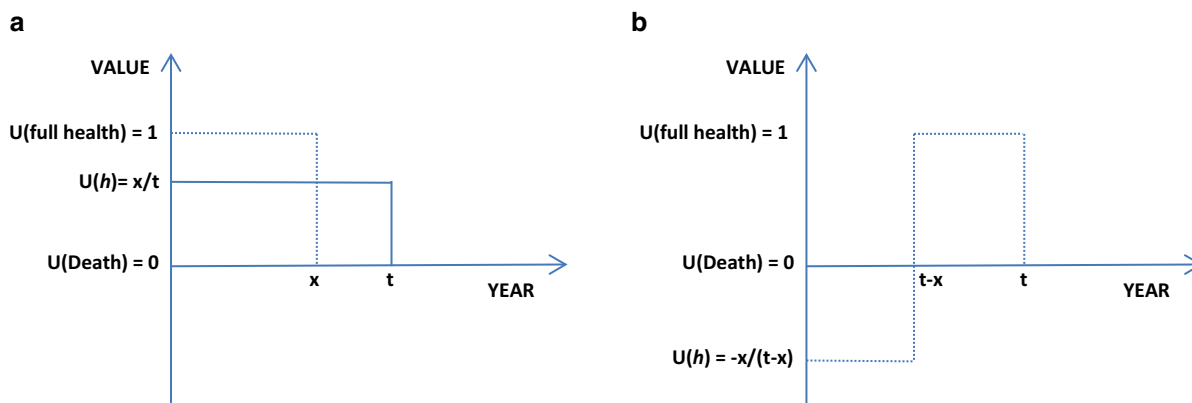
### 2.1.1 Conventional TTO

In conventional TTO (also referred to as classic or traditional TTO), health states considered better or worse than death are valued using different approaches. For an impaired state  $h$ , which is considered better than death (BTD), the respondent faces a series of choices between two hypothetical lives: one involving  $x$  years of healthy life, followed by death (alternative 1); the other involving  $t$  years in  $h$  (where  $x \leq t$ ), followed by death (alternative 2). Time  $t$  is fixed, whereas time  $x$  is varied until the respondent's point of indifference is identified. If the respondent prefers alternative 2 to alternative 1,  $x$  is increased to make alternative 1 more attractive; if the respondent prefers alternative 1 to alternative 2,  $x$  is reduced to make alternative 1 less attractive. This iterative procedure continues until the respondent is unable to choose between the two lives. The value of  $h$ ,  $U(h)$ , is calculated according to how much healthy time the respondent is willing to forgo at this point of indifference, and is given by  $x/t$ . This is shown schematically in Fig. 1a.

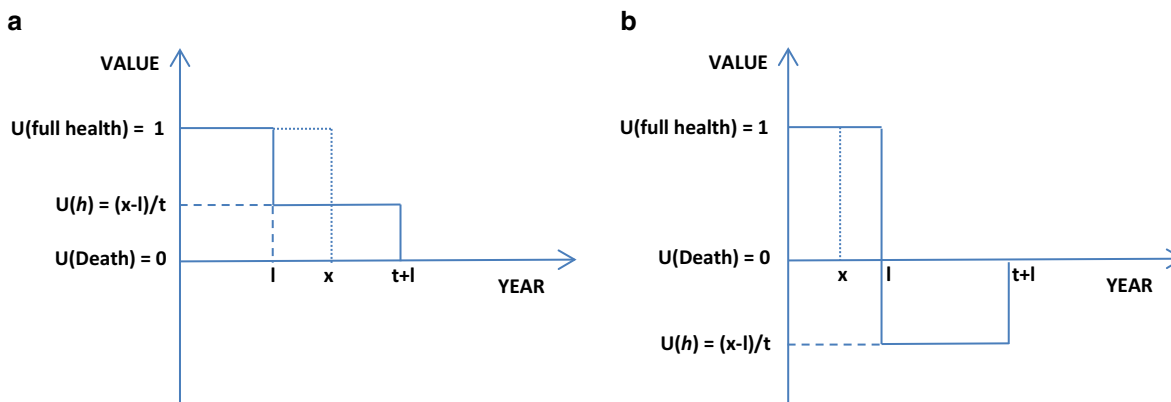
In conventional TTO, if  $h$  is considered by the respondent to be worse than death (WTD), the respondent is presented with a different type of choice: between a life involving  $t - x$  years in  $h$ , followed by  $x$  healthy years and then death (alternative 1); and immediate death (alternative 2). The value of  $x$  is varied until the respondent's point of indifference is identified, at which point the  $U(h) = -x/(t - x)$ . This is illustrated in Fig. 1b. As can be seen, the approaches used to value health states BTD and WTD are fundamentally different. The use of different approaches to produce values on a single scale has been a major point of critique for the conventional TTO [16]. A second point of critique for the conventional TTO is that the scale has an upper bound of 1 but a lower bound of negative infinity if all time  $x$  is traded off in the WTD procedure. While the BTD and WTD procedures are visually similar, the value of the health state  $U(h)$  in the BTD procedure is a linear function of the length of life in full health ( $x$ ), whereas a (visually similar) change in the WTD procedure involves changing both the length of life in  $h$  and the length of life in full health, such that  $U(h)$  is a rapidly steepening function. Three types of solutions have been proposed to deal with this scale issue: rescaling the WTD values to a scale ranging from 0 to  $-1$  [17–19]; changing the measure of central tendency from the mean to the median [19, 20]; and using geometry-based (angular) methods [21]. The choice of method for handling WTD values has been shown to have a substantial impact on the resulting values [19, 22]. A thorough discussion of the relative merits of these procedures is beyond the scope of this paper, but all of them have been criticized for having limited theoretical underpinnings [21, 22].

### 2.1.2 Composite TTO

Robinson and Spencer [16] proposed an alternative approach to avoid having to use different tasks to elicit



**Fig. 1** Conventional time trade-off.  $h$  impaired health state,  $U(h)$  value of state  $h$ ,  $x$  time in full health,  $t$  time in state  $h$



**Fig. 2** Lead-time time trade-off.  $h$  impaired health state,  $U(h)$  value of state  $h$ ,  $x$  time in full health,  $l$  lead time,  $t$  time in state  $h$

BTD and WTD values: lead-time TTO. This approach involves adding healthy life-years ('lead time';  $l$ ) before both of the lives being compared in the BTD task of conventional TTO (Fig. 2a). This allows the respondent to trade off these additional years when he or she considers  $h$  to be WTD (Fig. 2b). Research conducted using lead-time TTO showed severe framing effects and made clear that respondents had difficulties with the task. One finding was that it was not clear to respondents that trading into the lead time implied that the health state was considered to be worse than death [15, 23, 24]. The composite TTO was proposed as a compromise between the conventional TTO and lead-time TTO [25]. For health states BTD, the BTD task of the conventional TTO is used, while for health states WTD, lead-time TTO is used. This makes the distinction between BTD and WTD explicit to respondents, compared with lead-time TTO, and makes the BTD and WTD tasks more comparable than with conventional TTO. The value of  $h$  for BTD is calculated according to the conventional TTO,  $U(h) = x/t$  ( $x \leq t$ ) (Fig. 1a). The value of  $h$  for WTD is calculated according to lead-time TTO,  $U(h) = (x - l)/t$  ( $x \leq t + l$ ) (Fig. 2b). As with conventional TTO, composite TTO can be criticized for eliciting BTD and WTD values through different procedures, but the switch of WTD procedure to more closely resemble the BTD procedure is believed to be an improvement over conventional TTO [25].

## 2.2 Iteration Procedures and Mode and Method of Administration

### 2.2.1 Iteration Procedures

The variation of the length of life in alternative 1 (the  $x$  value) requires the determination of a starting point, an iteration algorithm and a termination rule (on what condition to end the iteration process). The starting point can be a fixed or random value, whereas the iteration algorithm and

the termination rule are usually standardized for all respondents. Most iteration algorithms can be described as being in one of three categories: titration, in which the length of life is sequentially altered by fixed increments/decrements; bisection, in which the length of life is always the midpoint of the remaining scale section (bisected); or 'ping-pong', where high and low values are alternately presented [6]. Combinations of different iteration procedures can also be used, for instance starting with a few steps of bisection, followed by titration. The rule for termination can be the identification of the indifference point or the boundaries of a range surrounding the indifference point.

### 2.2.2 Mode and Method of Administration

Traditionally, one-on-one personal interviews have been used for the collection of TTO data. Nowadays, due to developments in technology, online data collection has become very popular, as it can substantially reduce the costs of conducting a valuation study. Studies conducted by the EuroQol Group on the impact of administration mode on TTO data collection showed that group-based interviews and online data collection resulted in a substantially larger proportion of respondents stating their point of indifference after just one or two steps in the iteration procedure than one-on-one personal interviews [15, 26–28]. Because this is a clear indicator for a reduction in task comprehension and/or engagement of the respondents, the recently developed EuroQol protocols continue to be based on one-on-one personal interviews.

Before the digital age, a TTO time board was used by interviewers to guide respondents through the iterative procedure. Computer technology can be used to make TTO tasks easier to administer. For example, computer-assisted personal interviewing (CAPI) can help ensure more consistent interview conditions and protocol compliance by automatizing the iteration procedure and minimizing variations. Furthermore, electronic data collection avoids the

risk of coding errors, as there is no need to enter the obtained data into a database that is needed for analyses.

### 2.2.3 Visual Aids

The concept of TTO (i.e. trading-off QOL and length of life) may not be straightforward to individuals from whom preferences are to be elicited. The iteration procedures used to search for TTO values add to the complexity of the task. To improve task comprehension and conduct of the TTO tasks, visual aids in the form of TTO time boards are usually used in TTO-based valuation studies. In general, the visual aids take the form of graphs or props and use straight lines or graduated scales to illustrate the length of the different lives. The design of the visual aids and their use by the interviewers supports illustration of the variation of the length of lives so the iteration procedure can be followed by respondents without confusion. In a CAPI setting, digital versions of the TTO time boards are typically used instead of physical props.

### 2.3 Warm-Up Tasks for Respondents

TTO is difficult in that respondents need to understand the rather abstract concept of a ‘health state’, the difference between QOL and length of life, and the concept of trade-offs. The task is further complicated by requiring the respondent to imagine hypothetical health states of which many have no prior experience, over time horizons that may be unrealistic or difficult to contemplate. To prepare the respondents for TTO tasks, warm-up tasks and training of respondents such as a simple example or exercise are typically conducted before respondents are asked to embark on the formal valuation tasks. They can also help identify respondents who are unable to understand TTO tasks.

### 2.4 Training for Interviewers

Health state valuation is difficult not only for the respondent but also for the interviewer. Interviewers must be able to explain all aspects of the task to respondents. The pre-specified iteration procedure the interviewer is required to use may be complex. In such a setting, the interviewer can unwittingly (or deliberately) influence responses through idiosyncratic application of the interview protocol. Supporting materials such as a thorough interviewer script, visual aids and response recording sheets or standardized software can help to mitigate interviewer effects. Training of interviewers aims to familiarize them with the study protocol and reduce the potential for interviewer effects. A multiday training workshop is recommended, covering the study background and objectives, detailed discussion of the

interviewer script, and opportunities for demonstration, practice and feedback.

### 2.5 Quality Control of Data Collection

Quality control is an important element in TTO-based valuation studies. The purpose is to ensure protocol adherence to the extent that interviewers properly explain the aim and all elements of the task, mitigating interviewer effects. Measures for quality control should focus on interviewer behaviour. The measures should include evaluation of protocol proficiency and examination of interview times invested in each task (a proxy indicator of thoroughness). In addition, the measures should aim to evaluate respondents’ comprehension of and engagement in the TTO tasks. All TTO-based studies should include this component. This is nowadays typically included in clinical trials [29], but—to our knowledge—not in valuation studies. The EuroQol Group now integrates it in all computer-based TTO valuation studies.

## 3 The EuroQol Valuation Protocols

Table 1 summarizes the main characteristics of the three protocols. The MVH and Paris protocols were developed in sequence for valuation of the EQ-5D-3L health states using conventional TTO. These protocols reflected the state of the art in TTO-based valuation at the time (1993 and 2009, respectively). The EQ-VT protocol is built on the experience obtained from the earlier protocols and on a set of multinational pilot studies [15]. It was launched in 2012. All three protocols feature warm-up exercises to familiarize the respondent with the EQ-5D health states and the TTO technique. In all three protocols, the respondent is first asked to classify their own health with EQ-5D to introduce them to the concept of health states. In the MVH and Paris protocols, the respondent is then asked to rank a set of health states and then to value these using a VAS. In the EQ-VT protocol, the warm-up tasks do not involve ranking or VAS valuation, but focus on practicing the valuation of health states using composite TTO prior to formal valuation tasks. All three protocols feature study designs capable of providing the necessary data for modelling using a multitude of regression techniques. For example, all three require each respondent to value ten or more different health states using TTO, allowing for individual-level models such as random-effects models. In addition, all protocols include a second valuation technique. The MVH and Paris protocols use VAS valuation prior to the TTO task, whereas the EQ-VT protocol includes a DCE after the TTO tasks [15]. For the purposes of this paper, we concentrate on the trade-off framework,

**Table 1** Summary of the characteristics of EuroQol protocols

	Protocol		
	MVH	Paris	EQ-VT
<b>TTO framework</b>			
Framework	Conventional TTO		Composite TTO
Anchor states	11111 ('no problems in walking about; washing or dressing oneself; and performing usual activities; no pain or discomfort; not anxious or depressed') and 'death'		'Full health' and 'death'
Time horizon	10 years for both BTD and WTD states		10 years for BTD states; 20 years for WTD states
Structure of the hypothetical lives	For BTD states: Life A—living in full health for $x$ years followed by death (alternative 1); Life B—living in the impaired state for 10 years followed by death (alternative 2) For WTD states: Life A—living in the impaired state for $10 - x$ years, then full health for $x$ years followed by death (alternative 1); immediate death (alternative 2)		For BTD states: Life A—living in full health for $x$ years followed by death (alternative 1); Life B—living in the impaired state for 10 years followed by death (alternative 2); For WTD states: Life A—living in full health for $x$ years followed by death (alternative 1); Life B—living in full health for 10 years, then the impaired state for 10 years followed by death (alternative 2)
<b>Iteration procedures and mode and method of administration</b>			
Iteration starting point	10 years		
Iteration algorithm	Bisection for the first three steps followed by upward/downward titration with 1-year increments and a correction of 6 months when preference reversal occurs	Bisection for the first three steps followed by upward/downward titration with 1-year increments without any correction	Bisection for the first three steps followed by upward/downward titration with 1-year or 6-month increments and corrections of 6 months whenever preference reversal occurs
Iteration termination rule	Indifference/indifference range = 6 months	Indifference/indifference range = 1 year	Indifference
Mode of administration	Face-to-face, one-on-one personal interviewing		Face-to-face, one-on-one computer-assisted personal interviewing
Method of administration and visual aids	Paper-and-pencil with a time board as the visual aid		Electronic data collection and digital presentation of the visual aid
<b>Training and quality control</b>			
Respondent training	Ranking and VAS-based warm-up tasks plus explanation of TTO by interviewer		One example and three practices
Interviewer training	Interviewer training workshop (unstandardized)		Interviewer training workshop (unstandardized)
Quality control	No quality control procedure was formalised and included in the protocol		Standardised quality control software

*BT*D better than death, *EQ-VT* EuroQol valuation technology, *MVH* measurement and valuation of health, *TTO* time trade-off, *VAS* visual analogue scale, *WTD* worse than death

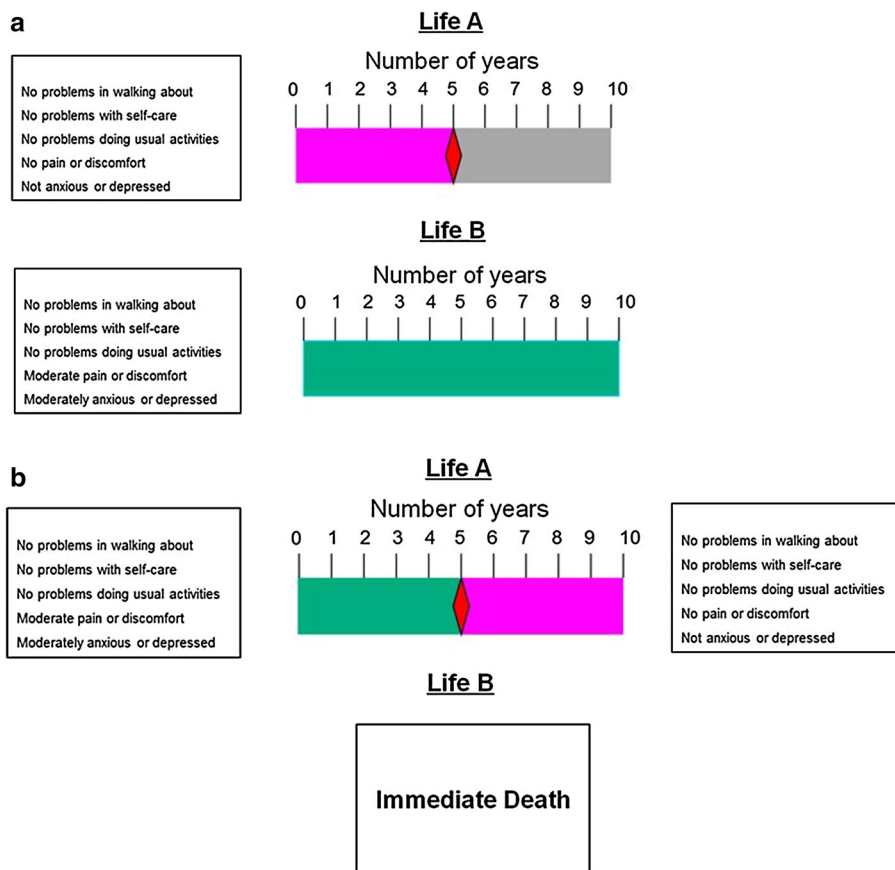
iteration procedure and its mode and method of administration, training and quality control measures when describing the EuroQol protocols.

### 3.1 The MVH Protocol

The MVH protocol was developed in 1993 for eliciting the utility values of EQ-5D-3L health states from a general UK population sample [13]. The conventional TTO using a fixed time horizon of  $t = 10$  years was adopted. The two

anchor states used are the best health state defined by EQ-5D-3L (health state '11111', i.e. no problems in any of the five dimensions) and immediate death. The iteration procedure of the MVH protocol uses a mix of three algorithms: bisection for the first three steps followed by an upward/downward titration procedure. The respondent is first asked to choose between 10 years in 11111 (life A) and 10 years in the target state (life B). Second, he or she is asked to choose between 0 years in good health (life A), i.e. immediate death, and 10 years in the target state (life

**Fig. 3** The visual aids used in the MVH and Paris protocols. **a** Visual aid for valuation of states considered to be better than death; **b** visual aid for valuation of states considered to be worse than death. *MVH* measurement and valuation of health



B). Depending on the choice in the second task, the length of life A uses the mid-point of the BTD or the WTD scale (5 years in both cases) in the third task.

Following the third choice, the procedure continues with 1-year incremental changes to life A, followed by a 6-month correction at preference reversal. If life B is considered BTD, the third choice will compare 5 years in full health (life A) with 10 years in the target state (life B), followed by one-year adjustments. If  $x$  denotes the length of life A at the point of indifference, the value of the target state  $h$  is calculated as  $U(h) = x/10$ . If life B is considered WTD, a different task is presented. In this case, the respondent is faced with a choice between a composite life A, which begins with  $t - x$  years in the target state  $h$ , followed by  $x$  years in '11111', for a fixed total of  $t = 10$  years, and a life B of 0 years (immediate death). In the third choice question, life A is described as 5 years in  $h$ , followed by 5 years in '11111'. If life A is preferred, life A will be altered to 6 years in  $h$ , followed by 4 years in '11111'. If life B is preferred, the time in  $h$  will be decreased to 4 years, followed by 6 years in '11111'. If  $10 - x$  is the number of years in the WTD state, and  $x$  is the number of years in full health, the value of the target state is  $U(h) = -x/$

$(10 - x)$  when indifference is achieved. The iteration terminates when the respondent states preferential indifference, or when the point of indifference could be inferred to lie between two life 'A's differing by 6 months in length, at which the mid-point is assumed to be the point of indifference. The full iteration scheme is illustrated in Appendix Fig. 5a.

The MVH protocol is designed for use in one-on-one, face-to-face interviews. A data-collection form incorporated with step-by-step instructions and a standardized script [30] is prepared for trained interviewers to strictly follow the above-mentioned elicitation technique. A specially designed visual aid called a 'time board' is prepared for the interviewer to present the anchor and target states and explain the valuation tasks. As illustrated in Fig. 3, there are two horizontal graduated bars on side A of the time board representing the two hypothetical lives for valuing BTD states; a similar bar with two sections on side B is used to illustrate life A for valuing WTD states. All health states are presented by attaching cards to the time board, and the different lengths of life A are presented using a sliding pointer. The MVH protocol includes neither specific quality control components nor training guidelines for interviewers.

As stated previously, there are many different options with respect to the details of the TTO. All of these have their own pros and cons. Since there is no ‘gold standard’ for valuation, it is unavoidable that, when designing a TTO protocol, some decisions are not based on empirical evidence, such as the decision to use a fixed 10-year time horizon.

Several issues of concern became apparent from the use of the MVH protocol, including within-respondent logical inconsistency in valuation [31, 32], great variance among respondents [18, 33], interviewer effects [33] and a non-continuous value distribution [18, 33, 34]. Also, it should be noted that the MVH protocol was not officially standardized or recommended. Many studies used MVH-like protocols that were similar but not identical, thus hampering the comparability of the resultant value sets. TTO-based valuation studies using these protocols are reviewed elsewhere [35].

### 3.2 The Paris Protocol

The Paris protocol is an updated version of the MVH protocol for valuation of the EQ-5D-3L health states, refined to improve the data-collection process [14]. The main difference between the two protocols is that the Paris protocol uses a simplified iteration procedure and a different selection of health states. In the Paris protocol, iteration is terminated either when indifference is stated or, if indifference is not stated, when the interval surrounding the indifference point is narrowed to 1 year. This means that only integer years are used as  $x$  values in the iteration procedure. As illustrated in Appendix Fig. 5b, the number of  $x$  values is half of that for the MVH protocol. Accordingly, the interview and the data-collection forms are less complex than those of the MVH protocol. The rationale for using the simplified iteration procedure is to improve efficiency. Increasing the unit of measurement from 6 months to 1 year would not be expected to have a major effect on the mean and standard deviation of observed values for health states given the wide variations in values across individual respondents.

The Paris protocol has been used in a number of EQ-5D-3L valuation studies [36–40] since being proposed in 2009, but we are not aware of any empirical research on the comparative merits of the Paris and MVH protocols.

### 3.3 The EQ-VT Protocol

The EQ-VT protocol adopts the composite TTO. The iteration procedure is built on that of the MVH protocol: initial comparison to 10 years in full health, separation into BTD/WTD, bisection of the BTD/WTD scales, and 1-year

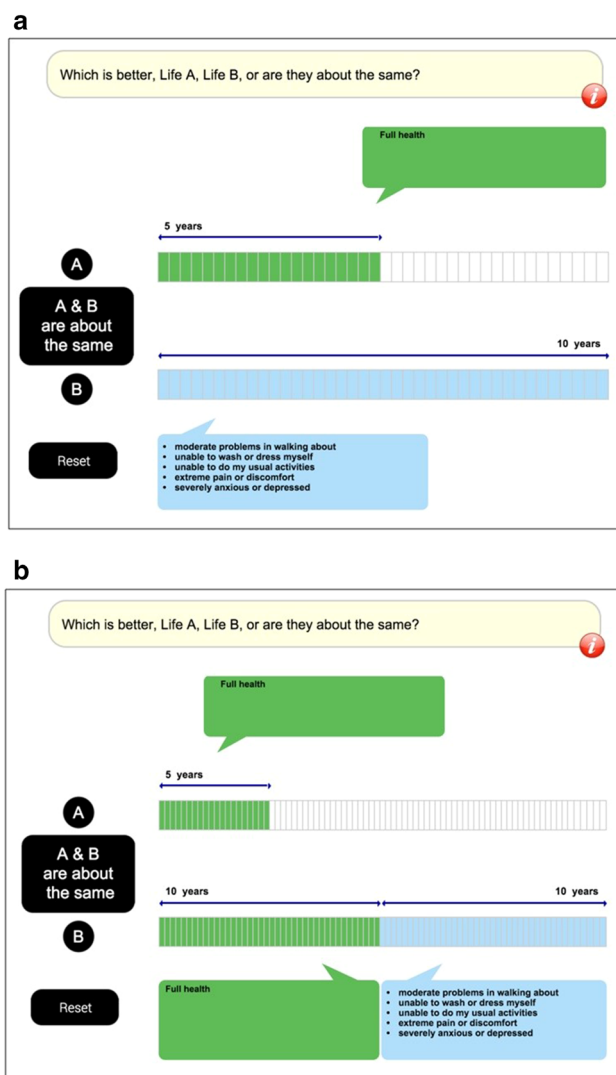
incremental adjustments followed by a 6-month correction at preference reversal. Unlike the MVH and Paris protocols, the composite TTO task does not terminate until the respondent states indifference, allowing endless adjustments by 1-year increments, followed by 6-month corrections whenever the direction of preference is reversed. The EQ-VT protocol also allows easy variation of the  $x$  values within and across the BTD/WTD scales. The possible utility value ranges from  $-1$  to  $1$ , with the smallest difference between values being  $0.05$  (see Appendix Fig. 5c).

The EQ-VT protocol is designed for use in computer-assisted personal interviews. A visual aid similar to those in the MVH protocol is presented on the screen to illustrate the composite TTO questions (Fig. 4). All components of the protocol, including an interviewer guide, were developed to provide standardized interview conduct. Multiple training and quality control components are included in the EQ-VT protocol. First, a recommended interviewer-training procedure has been developed. Second, a training task is incorporated into the interview to make sure the respondent understands the concept of TTO. The interviewer first shows how TTO works using as an example the state of being ‘in a wheelchair’. This training task is followed by three practice tasks where the respondent is asked to value EQ-5D-5L health states of varying severity of problems.

The EQ-VT protocol was informed by a multi-country, multi-stage research program [41]. Two major aspects of the design in the EQ-VT protocol were based on empirical research. The composite TTO was adopted after scientific investigation of several TTO variants, including conventional TTO, lead-time TTO and lag-time TTO [27], using differing visualizations and time horizons. Furthermore, the decision to use face-to-face interviews was made after testing internet surveys and group interviews [28]. Empirical studies were also conducted to inform decisions on other aspects of the protocol design such as the visual aid [26] and the anchor state ‘full health’ [42].

A number of EQ-5D-5L valuation studies have used the EQ-VT protocol since its first version (version 1.0) became available in 2012. While those studies showed that the protocol is feasible, reliable and sensitive to variations in EQ-5D-5L health states, some issues have also emerged. Protocol adherence by the interviewers and data distribution were two areas worthy of attention. Analyses of interviews performed using the protocol indicated the presence of interviewer effects, with respect to both protocol compliance and TTO values obtained [43]. To address these issues, quality control software (QC tool) was developed and implemented in the second version of the protocol, which allows for real-time monitoring of protocol compliance and interviewer performance from the start of





**Fig. 4** The visual aids used in the EQ-VT protocol. **a** Visual aid for valuation of states considered to be better than death; **b** visual aid for valuation of states considered to be worse than death. EQ-VT EuroQol valuation technology

interviewer training and during the entire data-collection process. Interviewers failing to follow the protocol can be identified and retrained during data collection or removed from the study.

As stated previously, the EQ-VT includes a DCE in addition to the TTO task. When the development of the EQ-VT started, it was recognized that TTO as a valuation technique has its limitations and that other valuation techniques might be needed to replace or to be used in conjunction with TTO to make the valuation studies more affordable and feasible. Based on the promising results of a pilot study [44], DCE, which is rooted in random utility

theory, became one of the main candidates. In a DCE, respondents are shown multiple (usually two) EQ-5D health states and asked to indicate which one they prefer, arguably making the valuation task easier to understand for respondents than TTO. However, this reduction in task complexity comes with a cost: health state values based on DCEs are on an arbitrary scale based on the relative distances between health states and not on a scale anchored at 0 (death) and 1 (full health) as is required by the QALY model. Until this anchoring problem is properly resolved, DCE as a standalone technique is not viable for generating utilities for use in QALY calculations and therefore cannot replace the TTO.

However, both DCE and TTO attempt to measure the same concept in different ways and both types of data seem to contain information relevant to this concept (i.e. the utility function). Therefore, the data resulting from the two elicitation techniques could be seen as complementary rather than competing. Put another way, the assumption is that respondents have a unique utility function that generates both types of responses. This leads to the idea of combining the TTO and DCE data into a single modelling framework: the hybrid model. The hybrid model is a maximum likelihood model where the 'hybrid likelihood' is the product of the likelihoods of the TTO data and the DCE data. The  $\beta$ 's of the TTO model and those of the DCE model are connected via a link function to account for the differences between the scales [43]. A hybrid model maximizes the use of the available data from a valuation study using the EQ-VT protocol.

#### 4 Ongoing and Future Research

Given the nature of health state values, it will never be possible to determine if or when a valuation method has reached perfection. Thus, the EQ-VT protocol, representing the current state of the art of the EuroQol Group's TTO-based valuation methods, is undergoing continuous scientific investigation for ways to further improve its performance and user friendliness [45]. Some of the research ideas that are currently being investigated or may be implemented in the future are briefly described below. These involve testing alternative iteration procedures and time horizons used in the composite TTO, as well as comparison with other TTO valuation protocols.

Several of the challenges associated with TTO are tied to the iteration procedure. The sequential change of lengths of the presented lives can be difficult to explain and understand, and when exposed to an adaptive sequence of choices,

respondents can develop response strategies that result in biased values. Thus, the development of a TTO or TTO-like valuation procedure that does not rely on iteration might be beneficial for the EuroQol valuation protocols.

The chosen time horizon is likely to influence resulting TTO values [46, 47]. The choice of a 10-year horizon has not been explored to a great extent in any of the three EuroQol protocols. Hence, further research is warranted to test alternative time horizons for the EuroQol valuation protocols, especially in light of the assumption of constant proportional time preference. Both fixed and variable time horizons may be worth testing. Since shorter timespans may be more plausible for certain health problems, since it is unlikely that patients are left in such states for a long time in clinical practice, shorter time spans could merit further investigation. Also, a life expectancy of 10 years may be unrealistic for certain respondents such as the elderly or patients with severe conditions [48]. However, the use of a short time horizon may create the issue of non-equivalence in resultant utility values with the current time horizon. When investigating this issue, attention should be paid to the effect of time horizon on severe health states, as some evidence suggests a maximum endurable time may exist for such states [49].

Last but not least, it would be beneficial to compare the performance of the EuroQol protocols and TTO valuation protocols developed outside the EuroQol Group. The TTO technique has been widely used to elicit utility values of health states not described by the EQ-5D. Although the valuation protocols used in those studies were not tested as thoroughly as the EuroQol protocols, some aspects of their design may be superior. A systematic review of the currently available TTO protocols and head-to-head comparisons of promising protocols with the EuroQol protocols in valuation of EQ-5D health states could provide valuable clues for how to further improve the EuroQol protocols.

## 5 Concluding Remarks

The EuroQol protocols provide standardized approaches to valuing EQ-5D health states using the TTO technique. These protocols are products of decades of research and development by an international group of multi-disciplinary researchers and they form an important basis for the current practice of health technology assessment in many countries. Hence, it is important for health researchers and policy makers to understand these protocols and to further improve them.

**Author contributions** NL conceptualized the paper. All authors contributed to the drafting and editing of the paper.

### Compliance with Ethical Standards

**Funding** This work is partially supported by the EuroQol Research Foundation.

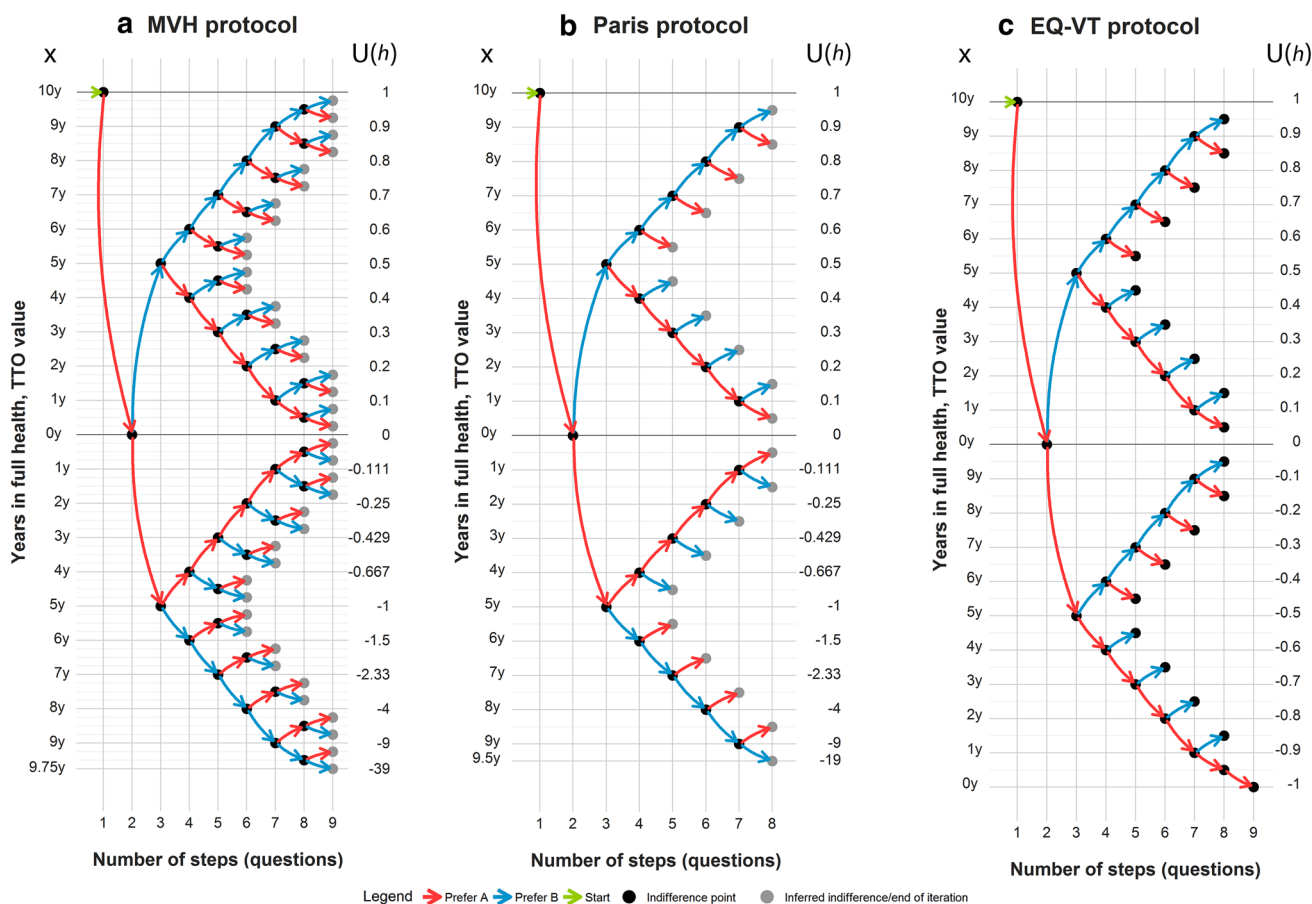
**Conflict of interest** All authors are members of the EuroQol Group.

**Disclaimer** The views of the authors expressed in the paper do not necessarily reflect the views of the EuroQol Group. Part of the content of this paper was presented in a workshop at the 2014 International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Asia-Pacific Congress in Beijing, China.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix

See Fig. 5.



**Fig. 5** The iteration schemes of the EuroQol protocols.  $h$  target health state,  $U(h)$  value of state  $h$ ,  $x$  time in full health. For states considered better than death, all three protocols present life A as  $x$  years in full health and life B as 10 years in the target state. For states considered worse than death, the MVH and Paris protocols present life A as 10- $x$  years in the target state followed by  $x$  years in full health, and life B as immediate death; in the EQ-VT, life A is

presented as  $x$  years in full health and life B as 10 years in full health followed by 10 years in the target state. In the MVH and Paris protocols, the tasks proceed along the directions of the arrows until indifference or the end of the branch in the graph is reached. In the EQ-VT, the tasks terminate only when indifference is stated, and the direction of the tasks can be reversed at any point

## References

- Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press; 2007.
- Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Oper Res*. 1980;28:206–24.
- Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, Napper M, Robb CM. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess*. 2001;5(5):1–186.
- Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res*. 1972;7(2):118–33.
- Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *J Health Econ*. 1997;16(2):155–75.
- Lenert LA, Cher DJ, Goldstein MK, Bergen MR, Garber A. The effect of search procedures on utility elicitation. *Med Decis Making*. 1998;18(1):76–83.
- Arnesen T, Trommald M. Are QALYs based on time trade-off comparable? A systematic review of TTO methodologies. *Health Econ*. 2005;14(1):39–53.
- Gold MR, Siegel JE, Russell LB, Weinstein MC. Estimating costs in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, et al., editors. *Cost-effectiveness in health and medicine*. New York: Oxford University Press; 1996.
- EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208.
- Brooks R. EuroQol Group: the current state of play. *Health Policy*. 1996;37(1):53–72.
- Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res*. 2010;19(6):875–86.
- Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–36.
- Williams A. The measurement and valuation of health: a chronicle. Discussion paper 136. York: Center for Health Economics, The University of York; 1995.

14. Kind P. A revised protocol for the valuation of health states defined by the EQ-5D-3L classification system: learning the lessons from the MVH study. York: Centre for Health Economics, The University of York; 2009.
15. Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health*. 2014;17(4):445–53.
16. Robinson A, Spencer A. Exploring challenges to TTO utilities: valuing states worse than dead. *Health Econ*. 2006;15(4):393–402.
17. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. *Med Decis Making*. 1994;14(1):9–18.
18. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43(3):203–20.
19. Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. *Med Care*. 2007;45(3):238–44.
20. Shaw JW, Pickard AS, Yu S, Chen S, Iannacchione VG, Johnson JA, Coons SJ. A median model for predicting United States population-based EQ-5D health state preferences. *Value Health*. 2010;13(2):278–88.
21. Craig BM, Oppe M. From a different angle: a novel approach to health valuation. *Soc Sci Med*. 2010;70(2):169–74.
22. Augestad LA, Rand-Hendriksen K, Kristiansen IS, Stavem K. Impact of transformation of negative values and regression models on differences between the UK and US EQ-5D time trade-off value sets. *Pharmacoeconomics*. 2012;30(12):1203–14.
23. Devlin N, Buckingham K, Shah K, Tsuchiya A, Tilling C, Wilkinson G, van Hout B. A comparison of alternative variants of the lead and lag time TTO. *Health Econ*. 2013;22(5):517–32.
24. Augustovski F, Rey-Ares L, Irazola V, Oppe M, Devlin NJ. Lead versus lag-time trade-off variants: does it make any difference? *Eur J Health Econ*. 2013;14(Suppl 1):S25–31.
25. Janssen BM, Oppe M, Versteegh MM, Stolk EA. Introducing the composite time trade-off: a test of feasibility and face validity. *Eur J Health Econ*. 2013;14(Suppl 1):S5–13.
26. Luo N, Li M, Stolk EA, Devlin NJ. The effects of lead time and visual aids in TTO valuation: a study of the EQ-VT framework. *Eur J Health Econ*. 2013;14(Suppl 1):S15–24.
27. Versteegh MM, Attema AE, Oppe M, Devlin NJ, Stolk EA. Time to tweak the TTO: results from a comparison of alternative specifications of the TTO. *Eur J Health Econ*. 2013;14(Suppl 1):S43–51.
28. Shah KK, Lloyd A, Oppe M, Devlin NJ. One-to-one versus group setting for conducting computer-assisted TTO studies: findings from pilot studies in England and the Netherlands. *Eur J Health Econ*. 2013;14(Suppl 1):S65–73.
29. Buyse M. Centralized statistical monitoring as a way to improve the quality of clinical data. *Applied Clinical Trials*. 2014. <http://www.appliedclinicaltrials.com/centralized-statistical-monitoring-way-improve-quality-clinical-data>. Accessed 2 Feb 2016.
30. US Agency for Healthcare Research and Quality. US valuation of the EuroQol EQ-5 health states. Available from URL: <http://archive.ahrq.gov/professionals/clinicians-providers/resources/rice/EQ5Dproj.html>. Accessed 28 Sep 2015.
31. Badia X, Roset M, Herdman M. Inconsistent responses in three preference-elicitation methods for health states. *Soc Sci Med*. 1999;49(7):943–50.
32. Lamers LM, Stalmeier PF, Krabbe PF, Busschbach JJ. Inconsistencies in TTO and VAS values for EQ-5D health states. *Med Decis Making*. 2006;26(2):173–81.
33. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–108.
34. Stalmeier PF, Busschbach JJ, Lamers LM, Krabbe PF. The gap effect: discontinuities of preferences around dead. *Health Econ*. 2005;14(7):679–85.
35. Szende A, Oppe M, Devlin N. EQ-5D Value sets: inventory, comparative review and user guide. Dordrecht: Springer; 2007.
36. Lee YK, Nam HS, Chuang LH, et al. South Korean time trade-off values for EQ-5D health states: modeling with observed values for 101 health states. *Value Health*. 2009;12(8):1187–93.
37. Liu GG, Wu H, Li M, Gao C, Luo N. Chinese time trade-off values for EQ-5D health states. *Value Health*. 2014;17(5):597–604.
38. Chevalier J, de Pouvourville G. Valuing EQ-5D using time trade-off in France. *Eur J Health Econ*. 2013;14(1):57–66.
39. Ferreira LN, Ferreira PL, Pereira LN, Oppe M. The valuation of the EQ-5D in Portugal. *Qual Life Res*. 2014;23(2):413–23.
40. Menezes RM, Andrade MV, Noronha KV, Kind P. EQ-5D-3L as a health measure of Brazilian adult population. *Qual Life Res*. 2015;24(11):2761–76.
41. Devlin N, Krabbe P. The development of new research methods for the valuation of EQ-5D-5L. *Eur J Health Econ*. 2013;14(Suppl 1):S1–3.
42. Shah K, Mulhern B, Longworth L, Janssen MF. An empirical study of two alternative comparators for use in time-trade off studies. *Value Health*. 2016;19(1):53–9.
43. Ramos-Goñi JM, Pinto-Prades JL, Oppe M, Cabasés JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modelling of EQ-5D-5L health states using a hybrid approach. *Med Care*. 2014 (**Epub ahead of print**).
44. Stolk EA, Oppe M, Scalone L, Krabbe PF. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. *Value Health*. 2010;13(8):1005–13.
45. Shah K, Rand-Hendriksen K, Ramos-Goñi JM, Prause AJ, Stolk E. Improving the quality of data collected in EQ-5D-5L valuation studies: a summary of the EQ-VT research methodology programme. In: Proceedings of the 31st Scientific Plenary Meeting of the EuroQol Group; 2014. p. 1–18. [http://www.euroqol.org/uploads/media/EQ14-CH01\\_Shah.pdf](http://www.euroqol.org/uploads/media/EQ14-CH01_Shah.pdf). Accessed 2 Oct 2015.
46. Dolan P, Gudex C. Time preference, duration and health state valuations. *Health Econ*. 1995;4(4):289–99.
47. Dolan P. Modelling valuations for health states: the effect of duration. *Health Policy*. 1996;38(3):189–203.
48. Boye KS, Matza LS, Feeny DH, Johnston JA, Bowman L, Jordan JB. Challenges to time trade-off utility assessment methods: when should you consider alternative approaches? *Expert Rev Pharmacoecon Outcomes Res*. 2014;14(3):437–50.
49. Stalmeier PF, Chapman GB, de Boer AG, van Lanschot JJ. 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*. 2001;17(4):488–96.