

REVIEW



An overview of the time trade-off method: concept, foundation, and the evaluation of distorting factors in putting a value on health

Anna K. Lugnér^a and Paul F.M. Krabbe^b

^aTheta Research, Zeist, The Netherlands; ^bDepartment of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

ABSTRACT

Introduction: Preference-based instruments measuring health status express the value of specific health states in a single number. One method used is time trade-off (TTO). Health-status values are key elements in calculating quality-adjusted life years (QALYs) and are pertinent for resource allocation. Since they are used in economic evaluations of healthcare, searching for a theoretical foundation of TTO in economics is justified.

Area covered: This paper provides an overview of TTO, including its relation to economic theory, and discusses biases and distortions, compiled from recent and older research. Inconsistencies between TTO and random utility theory were detected; The TTO is confounded by time preferences and by respondents' life expectancies. TTO is cognitively challenging, therefore guidance during the interviews is needed, producing interview effects. TTO does not measure one thing at a time, nor are the values independent of other states that are being valued in the same task. That is, TTO does not exhibit theoretical measurement properties such as unidimensionality and the invariance principle.

Expert opinion: We conclude that the TTO may be a pragmatic method of eliciting health state values, but the limitations in regard to measurement theory and practical elicitation problems makes it prone to inconsistencies and arbitrariness.

ARTICLE HISTORY

Received 2 February 2020

Accepted 3 June 2020

KEYWORDS

Economic evaluation; health related quality of life; health state valuation; pharmacoeconomics; QALY; utilities

1. Introduction

It is generally acknowledged that apart from life years or survival, the 'quality of life' is crucial to any health evaluation. Since the introduction of the concept of the quality-adjusted life year (QALY), many attempts have been made to define quality-of-life and health. Since the ideas differ of what defines health and what exactly is being measured in its evaluation, the definition of quality-of-life remains elusive. In the literature, the term quality-of-life is used interchangeably with subjective health status, perceived health status, health-related quality of life, and general well-being [1–4]. If psycho-social and social functioning are included, the concept is often referred to as HRQoL (health-related quality of life). The criteria for quality-of-life are largely subjective, as opposed to standard health indicators such as survival, serum cholesterol levels, and bone mineral density. We use the term health status and acknowledge that this subjective phenomenon requires specific methods to be measured.

Health outcome instruments that measure health status can be developed within various measurement frameworks, e.g. indicating the frequency, intensity, or level of a specific health domain (e.g. mobility, pain) [5]. However, when comparing health outcomes across different populations, conducting disease modeling studies, and performing economic evaluations of various healthcare interventions, preference-based instruments are more useful. Preference-based measures expresses preferences in a single

numeric metric (which we refer to as a 'value' of a particular health state), and so, these measures incorporate weights that reflect the importance attached to a set of specific health aspects. Therefore, preference-based measures reflect the overall quality of an individual patient's (perceived) health status.

One of the prevalent preference-based methods to derive a value of a health state is time trade-off (TTO). The widespread use of the EQ-5D instrument, in which the value sets are mainly derived from TTO [6], has certainly contributed to the extensive use of TTO to construct value sets for health states. The AQoL-8D and its predecessors (4D, 6D and 7D versions) developed in Australia are other instruments that are largely based on TTO values [7]. The EQ-5D is commonly recommended for use in health economic evaluations of healthcare as a basis for resource allocation [8].

In the area of clinical decision-making, individual patients are often involved in eliciting values for health states that concern possible outcomes related to their own disease and optional treatment modalities. We refrain from incorporating such TTO values in our review because clinical decision-making is an entirely different area of research, with different goals and different premises [9]. TTO is also often used as a stand-alone method to value health state for a specific disease. In this overview, we specifically deal with the more general use (economic evaluation) of TTO where values for a (large) set of health states are derived from a (large) sample of respondents from the general population.

Article highlights

- Methods that are less biased than the time-trade off are needed to derive values and utilities for health.
- The research paradigm of health economics that underpins present protocols to value health states has too many flaws to adequately quantify a subjective phenomenon such as health.
- HRQOL values in cost-effectiveness studies (e.g. QALYs) are more valid if based on patient-reported and patient-centered values.

Through the years a large amount of literature has accumulated that discusses the merits and shortcomings of TTO, both on theoretical and empirical grounds. This overview provides insight into its development, including its relation to economic theory, and discusses the biases and distortions of the method. To put the TTO into its historical context that would explain the broad use of the method, we start with a brief overview of the QALY as a summary measure of health and preference-based measurement. Problems associated with the TTO are then summed up and discussed from the perspective of economic reasoning, and in light of the empirical evidence of its validity.

2. Quality-adjusted life years

In the early 1970s Fanshel and Bush introduced the QALY concept, at the time called ‘function years’, in a paper evaluating a tuberculin skin-testing program [10]. The QALY concept is a descriptive summary measure where the two main components of health – mortality and morbidity (i.e. health status) – are combined. Most of the methodological aspects pertinent to the measurement of health are discussed in that paper [10], and, although there has been some variation in the terminology used these aspects are still valid. The concepts of the index day and health day were mentioned by Torrance and colleagues in 1971 and ’72 [11,12]. Also, Grogono and Woodgate (1971) published an approach to measure health [13]. In their approach, ten domains specified health and with a simple weighting procedure they produced ‘health-years’. A few years before, Klarman and colleagues (1968) [14] compared treating options for chronic renal disease, measured as life years gained. By means of arbitrary adjustments in their epidemiological cohort model analyzes of ‘quality of life’ were performed. The term ‘quality-adjusted life year’ was introduced in 1976 by Zeckhauser and Shepard [15].

The quality-adjustment factor in the QALY requires a numeric metric with an upper bound of 1.0 corresponding to full health, with poorer health states being arrayed along a single-dimensioned continuum at scale values less than 1.0. Being dead has a value of 0.0 on this scale [16]. Various terms are used to refer to this quality metric such as value, utility, strength of preference, or index.

3. Preference-based measurement

Preference-based methods of valuation or measurement such as standard gamble (SG) and TTO (see next section) capture the overall health condition of individuals and reflect the value

of a health state. The core of a preference-based measurement framework consists of a response task comparing at least two objects with the objective of expressing which object is preferred (is better). These objects can be different health states or health states with different duration, or the objects can be the risk of dying or full recovery. A health state is often described as a small set of attributes, whereby each attribute entails a limited number of levels of severity. The respondents score the set of attributes as a whole, and not the individual attributes separately. In doing this, the ability to read and mentally processing all of the attributes simultaneously is required [17]. In comparing complete attribute sets, which differ according to levels of severity (i.e. the health state), a preference for one health states is evoked. Health states can also be compared with a specified health outcome (e.g. immediate death or living in full health for a specified number of years). People’s choices are then assumed to be based on trading off one health state (with specific levels of severity) against another health state, with other levels of severity or health outcomes. In short, preference-based methods come down to teasing out the ‘true’ values that people assign to specific health states.

Parallel to the development of the QALY concept, the development of preference-based methods in health economics started in the 1970s, both in North America [12,18–20] and in the UK [21]. Different preference-based methods to derive values trace their ancestry to different scientific fields, thereby entailing their own specific methodological consequences and different applications. Preference-based methods have been introduced from decision science [11], health economics [22–24], marketing [25], psychometrics [26], public health [27], and clinimetrics [28]. Which method should be used to derive health-state values is still under debate, however [29,30].

4. Standard gamble and game theory

The first theory on how to assign a numeric metric to utility was worked out by means of game theory [31,32]. Before this breakthrough, preferences were expressed as rankings without a cardinal or ratio measurement level (metric) implied. The renowned theory of expected utility is an axiomatic theory of rational choice under uncertainty. Individual behavior may be characterized as if choices were being made to maximize the expected utility. When the choices (i.e. preferences) obey the axioms of completeness and transitivity, the utility of different alternatives can be calculated [33].

The method of eliciting utilities as described by the von Neumann and Morgenstern (vNM) theory has become known as the standard gamble (SG). The introduction of SG is hard to trace, but it is clear that it was neither mentioned nor developed by vNM. Rather, it is based on Thurstone’s much earlier concept of probabilistic preferences [34–37]. In SG, a rational individual would choose the one lottery with the highest expected outcome [38]. In the original vNM theory this outcome was monetary, ‘supposed to be unrestrictedly divisible and substitutable, freely transferable and identical, even in the quantitative sense ...’ [31, p.8].

Expected utility theory uses monetary outcomes, since ‘the aim of all participants in the economic system, consumers as

well as entrepreneurs, is money, or equivalently a single monetary commodity' [31, p.8]. Preferences (utility) of goods are often expressed as the amount of money people are willing to pay. By replacing money as an outcome, SG has been used to elicit the utility of a health state or health intervention. One of the earliest published applications of SG in the context of deriving utilities for specific health states involved an experiment where two physicians would draw imaginary pills out of a box that would cure their patient, but with a risk of drawing a pill that would kill the patient [39]. Similar methods were used in clinical or patient decision making, whereby individual patients performed tasks to derive values for specific possible treatment outcomes [40,41]. In the setting of economic evaluation, the valuation procedure is often different; a wide range of hypothetical general health states, in contrast to disease-specific states, are assessed with a preference-based method. Whether existential outcomes such as health also can be captured by this approach is not clear. In the SG method the object of interest seems to have been transposed into a normative framework of valuation that may be a reliable approach to quantify normal goods, such as cars and holidays, but may be less appropriate to deal with the valuation of health [42]. Furthermore, unlike money, health cannot be transferred from one individual to another, which makes it different from other commodities (goods), making the transformation from SG to health status utilities cumbersome.

5. Time trade-off

5.1. Foundation

Due to the cognitive difficulties with the appraisal and processing of probabilities, which is necessary in the SG, as recognized by Torrance and colleagues in 1972, the TTO was developed for the specific application of valuing health states [12]. In his dissertation (1971), Torrance states that the SG and TTO methods give 'equivalent and reliable results but the time trade-off method was found easier to administer' [11, p. 125]. Torrance also refers to communication with Bush, proving his awareness of work in other research groups [11, p.33]. Bush and colleague Fanshel discussed the basic approach of TTO in 1970, referring to the notion as 'weighting through equivalence in time' [10, p.1043]. For Torrance the TTO was a 'short-cut' method of obtaining values equivalent to those estimated using SG (which was theoretically sound according to the vNM utility theory). TTO values were not intended to be used directly, nor was the method meant to replace SG entirely. The purpose of TTO was to estimate a functional relationship between SG and TTO values to interpolate TTO values between utilities estimated using SG. Interestingly, Torrance introduced TTO but did not use it for his later work (HUI-2, HUI-3). There does not seem to be consensus on whether the two methods (SG and TTO) give similar values [43] or not [44].

One reason for the popularity of TTO is that it is a preference-based method with a format and operationalization procedure that has a connection to economic reasoning (e.g. trade-off: sacrifice one alternative in order to receive another). Furthermore, TTO prompts face validity as it mimics the QALY

concept and resonates with the medical context [43]. Notably, the values deduced by TTO are often referred to as utilities, but TTO does not conform to the expected utility theory in the sense that there is no element of risk involved in the assessment. As such, some researchers claim that formally TTO tasks produce values, and not utilities.

The wide use of the TTO in economics evaluations has prompted efforts to find a theoretical foundation of TTO in economics. The starting point in economics is that there are unlimited human wants which are to be met by limited resources. Essentially, this is what economists call scarcity, and it is closely related to people's choices for goods and services. When an individual is faced with a number of alternative options to choose from, he will choose the one that he prefers. This is equivalent to choosing the option that gives an individual the highest utility, i.e. what gives an individual the highest benefit, what makes him or her feel better. Under conditions of scarcity, the choice is a trade-off that involves giving up one commodity in return for gaining another one.

5.2. Valuation task

In a TTO valuation task the respondents are asked to trade off duration of life against health status. The trade-off entails choosing a shorter life spent in full health or living longer but in a lesser state of health. Often this is done by using an iterative process to offer the respondent different lengths of life before they indicate indifference. Intuitively, individuals would prefer to spend a shorter time in full health than a longer time in a lesser health state, and therefore they would trade off life years for better health. The number of years sacrificed in full health represents the value of the lesser state.

In the operationalization of TTO, as applied in the Measurement and Valuation of Health Study (MVH) [45], the first, large population study using TTO, the preferences are elicited by confronting a respondent with a suboptimal health state of a given duration (x , often 10 years). As the competing alternative, a better health state (conventionally perfect health) is offered but with a shorter duration ($y < 10$). In the TTO exercise, the 10-year period is conventionally followed by death. The respondent is asked to state the duration spent in perfect health (y) at which he/she is indifferent between the duration y and the 10 years in the lesser health state. The value of the lesser state can then be established as $y/10$ (Figure 1).

5.3. Position of dead

TTO was originally developed to assess values for states considered better than being dead [46,47]. An important development of TTO was the elaboration by Torrance in 1984 of the method to accommodate states regarded as worse than being dead [48]. An important claim often made is that health states must be valued on a scale where the value of being dead is 0.0, because the absence of life is considered equivalent to zero QALYs [12,43]. Inevitably this rule leads to assigning negative values for very bad health states.

The MVH study used a specific methodology for worse-than-dead states. First, respondents were asked if a state is

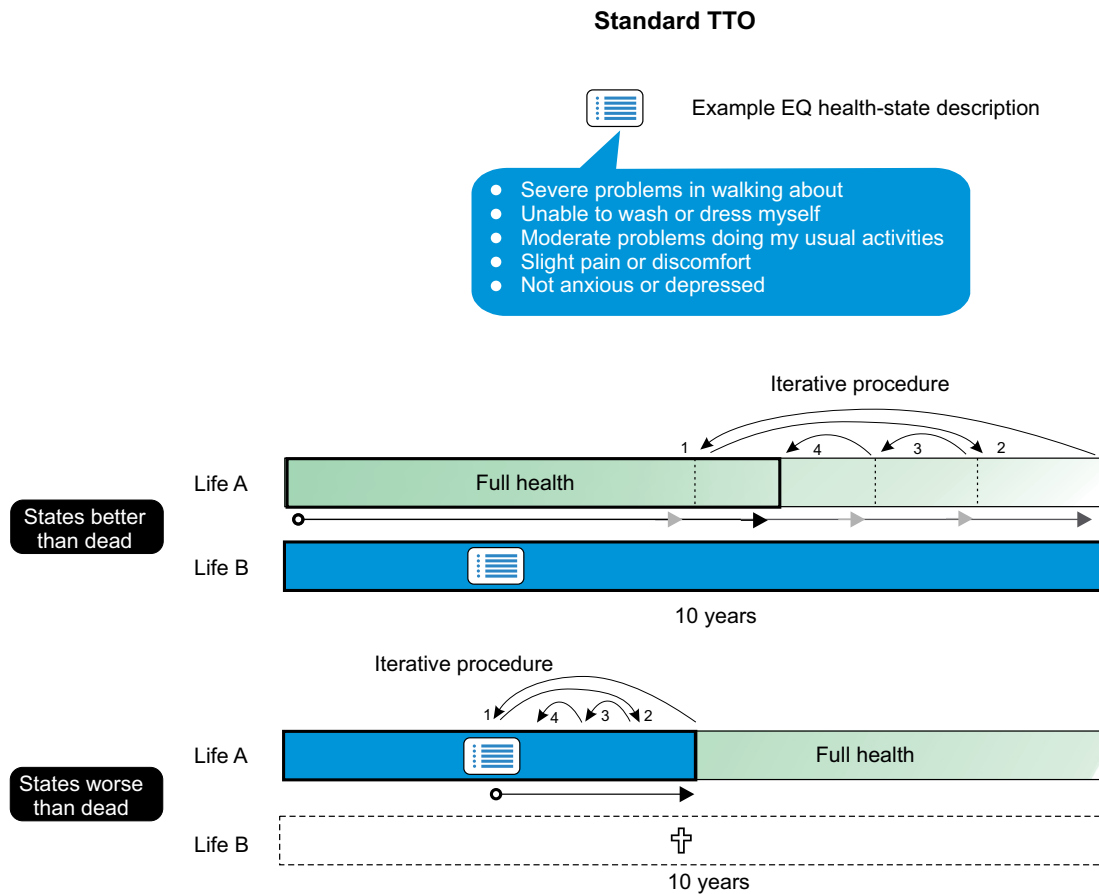


Figure 1. Representation of the time trade-off (TTO) methodology as performed in the MVH study. The iterative procedure to arrive at the point (number of years) where a respondent is indifferent between a shorter life in full health, State A (dark green), and 10 years in a lesser health state, State B (light blue) is illustrated with the arrows. For states perceived as worse than dead by the respondent, the time spent in that state is followed by some time in full health. The iteration regards the number of years in full health that are needed to compensate for the years spent in a state worse than dead.

regarded as better or worse than dead. If worse, the respondent is asked how many years (t) spent in that state followed by a period of perfect health, summing to 10 years, would be equivalent to immediate death. In an iterative process t is decided upon. Assigning the values 0 and 1 to being dead and full health, respectively, the valuation for the worse-than-dead state is $1-(10/t)$. This follows from the calculation that the time spent (t) in the state plus the remaining time spent in full health ($10-t$) would add up to 0. This implies that the value for a worse-than-dead state theoretically falls in the interval $(-\infty, 0)$.

In lead-time TTO (see also below), values for all states can be elicited in one procedure, avoiding the need to engage in a different procedure for states worse than dead [22,23]. The composite TTO, with its lead-time TTO part for the states worse than dead, may solve some of the problems. However, the main problem with the original TTO has not been resolved, namely that different procedures were used for states worse and better than dead [49].

5.4. Alternative TTO versions

Alternative versions of TTO have been developed in an effort to optimize the method for producing more credible values or to deal with temporary health states [50]. In particular such

efforts have been undertaken for the EQ-5D instrument. Recent developments around the introduction and valuation of the EQ-5D-5 L (with 5 levels instead of 3) instrument still rely on TTO, but in a slightly different form [51]. Adjustments in the TTO procedure concern states worse than dead. Valuing health states with the lead-time TTO involves imagining a period of time in full health before the respondent moves into a state of less than full health (Figure 2). In contrast to lead-time TTO, lag-time TTO involves imagining a health state that is less than full health, starts immediately, and is followed by a lag-time of good health. Lag-time TTO produced lower values than lead-time TTO, and the difference increased in longer time frames [52]. The most recent EuroQol TTO version combines a conventional TTO to elicit values for states regarded better than dead and a lead-time TTO variant for states worse than dead; it is referred to as composite TTO [24]. Unfortunately, valuation studies based on the composite TTO showed almost no discrimination among values for states worse than dead [53].

6. Time trade-off distortions

6.1. Theory

There have been some attempts to fit the TTO into the economic utility theoretical framework [54,55]. In those

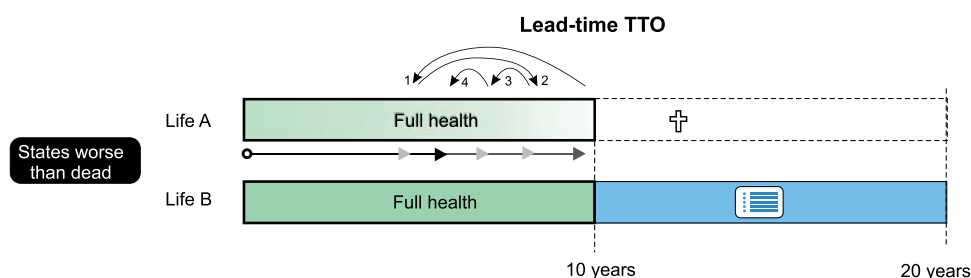


Figure 2. In the new EQ-VT protocol, lead-time TTO is used for states worse than dead. Here the respondent needs to picture a life in full health for 10 years before ending up in the state worse than dead for 10 years. . Maximally 10 years of life can be sacrificed to arrive at the point where the x years in full health are equivalent to a life of 10 years in full health followed by 10 years in a state worse than dead.

investigations, the axioms of rational preferences are assumed to hold. Yet, experiments and tests for specific conditions and properties give extensive empirical evidence of choices made by individuals that are inconsistent with the underlying theory of rational individuals [56–61].

Most inconsistencies between TTO and expected utility theory have been found when examining the possible influence of duration. The assumption of constant proportional trade-offs entails that the proportion of remaining life span that one would trade for a specified improvement in quality is independent of the remaining duration of life (e.g. TTO). There is mixed evidence of whether this assumption holds [61,62]. Another assumption is utility independence. Lack of utility independence is undermining the results of TTO; its absence means that the value of the health state depends on the time spent in it [62]. The lack of utility independence also implies that the value given in the valuation task depends on the order of the states presented in the task [47]. Overall, there is substantial empirical evidence that an individual's choice behavior does not align with the assumptions of economic utility theory [63–67]. Many of these violations of the axioms are related to the element of time, which will be discussed next.

6.2. Internal distortions

6.2.1. Time preference

In the first application of TTO, the values referred to a kidney-dialysis regime, and one of the conclusions was that the value of a health state depends on the amount of time spent in it [12]. Subsequently, the authors stressed the importance of carefully determining the duration of the state before measuring its value, and they emphasized the need to control duration during the measurement process. Through the years, more evidence has been gathered showing that duration affects the values measured [68–72]. TTO is not only confounded by time preferences and the framing effect due to the chosen duration; it is also confounded by the respondents' different life expectancies [73–76]. Attempts have more recently been made to explore and test these time duration effects [77], some based on the well-known and interesting prospect theory [78]. This theory explains that valuations in a TTO task differ depending on the respondents' own perceived life expectancy (which can be longer or shorter than the standard 10 years in TTO). However, hands-on methods to

correct or adjust TTO values based on findings of prospect theory are yet unavailable and will likely complicate TTO assessments even further.

6.2.2. Indifference procedure

Economics, as a discipline, has a tradition of assuming rationality and acting on the basis of perfect information. This tradition is reflected in the implicit assumption that the response received from a respondent will precisely reflect a considered choice or valuation for a specific outcome [33]. Regarding valuations, respondents may arrive at this response in one step. In practice, however, valuing health states with TTO is done through an iterative procedure to make the task more manageable.

The quarter-year increments used in the MVH protocol resulted in a large number of health-state predictions having negative average values (the lowest possible value is –39) [6,79]. In the EuroQol Valuation Technology (EQ-VT) protocol for the EQ-5D-5 L [51,80], which is based on a computer-assisted personal interview mode, the iterative process produces an accumulation of indifferences in three specific iteration steps, leading to a large number of valuations of –1.0, 0.0, and 1.0 for different states [81,82]. Where the participants' responses indicate a value of –1.0, we know their value is *at most* –1.0, since that observation is censored. This suggests that using these individual values to compute average values needs to take account of this censoring in some way [83,84]. The EQ-5D-5 L value set derived in England, reports that respondents were more likely to value a state at 0.0, 0.5, or 1.0 rather than any intermediate number, and there were very few values in the range of –0.5 to 0.0 [52]. These discontinuities suggest that the participants were providing quite crude responses. Indeed, participants who valued all health states using either a single high value (such as 1.0) or a single low value (such as 0.0), are possibly giving a signal that the states are either 'good' or 'bad' rather than providing numbers that have cardinal meaning. It cannot be denied that the iteration steps influence the values, but the exact size of the influence can hardly be defined.

6.3. External distortions

6.3.1. Cognitive understanding

An important empirical concern is the extent to which respondents understand and respond accurately to the judgmental tasks of the TTO (Figure 3). Qualitative research has shown

"POINT TO GREEN BAR FOR LIFE A: This green bar describes what we call Life A. In Life A you will live 10 years from now – and during those 10 years you will be in full health. After the 10 years you will die. POINT TO BLUE BAR FOR LIFE B: This blue bar describes what we call Life B. In Life B you will also live for 10 years after which you will die. However, in these 10 years you have problems with mobility which mean that you require a wheelchair. POINT TO QUESTION IN THE YELLOW BOX: Try to imagine what it would be like for someone like you to have to choose between Life A and Life B. Which would you choose? The choice is between Life A, 10 years of full health; and Life B, 10 years with the mobility problems."

Figure 3. TTO instruction as used in the EQ-5D-5 L valuation protocol.

that respondents often have difficulty distinguishing between the health states to be valued, understanding the hypothetical nature of health states, and conceptualizing being dead; the research also points out that the respondents may have their own religious or spiritual beliefs about being dead [85–87]. These cognitive distortions include those relating to the framing of issues, anchoring, time inconsistency, making choices in the presence of uncertainty (what will be happening after the 10/20 years of the TTO representation?), overconfidence, and many others [88–90].

6.3.2. Interviewer effect

Experience shows that the assessment of TTO requires the presence of a trained interviewer and/or specialized computer programs. The current EQ-VT TTO protocol dictates trained

interviewer assistance. Respondents are now confronted with complex and strenuous instructions offered by the interviewer, on top of the already complex representation (the health profile bars, written instructions, health-state descriptions). Acknowledging these difficulties, solutions were sought by the EuroQol Group in quality control tools to enhance protocol compliance to reduce interviewer effects. These tools are based on continuous data monitoring and various checks during data collection [91]. Nevertheless, recent studies have made it clear that trained interviewers do not seem to explain the tasks in the same structured way. This leads to interviewer-dependent values [92,93]. Such interviewer effects have also been traced in the MVH study for the VAS values [94]. Figure 4 sums up some of the issues discussed in this paragraph.

7. Conclusion

In this paper TTO is discussed as a method to derive values for health states. Because the method was developed in connection to the framework of expected utility theory and is an adaptation of the SG method of measuring utility, the values elicited with TTO would supposedly obey the axioms of economic theory. However, the research described in the literature shows that this is not the case. Compelling arguments against using TTO have been raised by several authors, most of them health economists themselves. In fact, TTO seems to be associated with many problems, both theoretical (e.g. axiomatic violations, problems in dealing with states worse than dead, time preference) and practical (e.g. difficult for people to perform, trained interviewer assistance required). From a measurement perspective, TTO has been criticized for its susceptibility to framing issues (e.g. duration of the time frame, mode of administration, indifference procedure). The TTO method is a rather pragmatic approach to arrive at values

Theoretical problems (time horizon)	Administration and framing
Life years affects health state values Duration and life-expectancy respondent <i>(Versteegh et al. 2013, van Nooten et al. 2014, Matza et al. 2015)</i>	Values affected by setting Internet, group session, self-administered <i>(Shah et al. 2013)</i>
Utility independence Do not hold for different time horizons <i>(Miyamoto & Eraker 1988)</i>	Interviewer effect Values affected by interviewer <i>(Edelaar-Peeters et al. 2014, Shah et al. 2015)</i>
Constant proportional time trade-off May not hold <i>(Dolan & Stalmeier 2003)</i>	Cognitive understanding Respondent interpretation differ from intended <i>(Green et al. 2000, Edelaar-Peeters et al. 2014)</i>
Violation expected utility Loss aversion <i>(Bleichrodt 2002)</i>	Iteration steps Tradeable unit and number of steps <i>(Dolan 1997, Janssen et al. 2013, Versteegh et al. 2013, Oppe et al. 2014)</i>
New versions do not solve the issues Lead Time TTO, Lag Time TTO, Composite TTO <i>(Attema et al. 2013, Mulhern et al. 2014)</i>	

Figure 4. Issues that influence the values elicited by the TTO, selected references.

for health states, but methodological, procedural, and analytical problems are prominent in the applications of TTO [95]. We believe that what is needed for the measurement of a subjective phenomenon such as people's perceived health status is a simpler and theoretical underpinned measurement model.

8. Expert opinion

TTO is cognitively challenging. As a result, there is a need for guidance during interviews, which in turn produces interview effects. The presence of significant interviewer effects [52] gives reason to doubt the straightforward interpretation of valuation data as 'true' representations of individuals' values of health states. There are reasons to suspect that many of those participating in TTO valuation tasks do not understand the tasks, are not fully engaging with the tasks at hand, or simply want to finish the valuation exercise as quickly as possible. All of these factors might have serious effects on the generated values. Furthermore, valuations provided by those who do understand and are engaged are influenced by the framing of the tasks and the interviewer's approach [96]. The need to introduce quality control on collected values during TTO valuation studies, and drop specific responses, indicates that the TTO cannot be considered a robust measurement method.

The combination of health-state descriptions augmented with the duration of these states has introduced biases and distortions into the health state values due to the time element. This is due to various factors: time preference, duration may become a dominant attribute, life stage concerns, violation of constant proportional time trade-off, and maximal endurable time [52,74,97–99]. One solution that has been suggested is to harmonize TTO methods [49]. A method of correcting for time preference in the analysis of TTO data has been suggested [100] but seldom used [101]. Moreover, although standardizing may seem an attractive solution, it will only treat the symptoms. In the case of the EQ-5D-5L, a discussion is started, challenging its leading use [102]. The recent extension of this instrument has even led to an in-depth investigation by the UK government [103]. An official report commissioned by NICE and the Department of Health and Social Care addresses several concerns related to the UK-values estimated for the 5-level version of the EQ-5D [104]. Many of these concerns seems associated to the principal valuation technique used, the TTO. Yet, these results may be typical for the UK EQ-5D-5 L valuation study, as in other countries that used later versions of the EQ-VT protocol, the TTO exercise produced results that were more balanced with smaller interviewer effects.

The current convention in QALY computation is that the lower anchor on the scale should be 0.0 and is defined as a state equivalent to being dead [105]. The argument for anchoring dead and not for the worst health state on 0.0, is because the absence of life is considered equivalent to zero QALYs [12,43]. Recently, a discussion has started whether positioning dead as zero is a theoretically requirement in the QALY approach [106]. These authors show that the arguments for dead as a zero anchor are not very strong. Yet, the existential notion 'dead' remains commonplace as the anchoring label in QALY computations and therefore the use of TTO.

To some extent it is understandable that valuation methods such as the TTO include dead since it is a natural part of the life course (the end of it), and therefore it may be intuitive to use 'dead' as the lowest point on the scale to express the quality of health (and even life). As a consequence, health states considered as 'worse than dead' have to be quantified or assessed one way or the other. Inevitably, this means that very bad health states that are considered as worse than dead are assigned negative values. Moreover, the issue is not only that states worse than dead may confront researchers with methodological problems, but even the fact that people are confronted with the notion 'dead' may itself introduce all kind of prejudices (biases) in their responses. Using the 'dead' concept in health measurement methods is a contentious issue [107]. The concept of dead is so confrontational and hard to grasp that some find it astonishing that health economists make dead a central element in their valuation methods [108]. In addition, it seems that 'dead' cannot be perceived as a manifestation of health status as it has a distinct connotation.

Various transformations have been proposed to position states that are better or worse than dead on a single metric scale. These methods assume that health-state values are independent of the duration of the states. However, there are some indications that this assumption does not hold, at least for severe health states [109–111]. The way variants of TTO have been introduced to solve problems regarding valuing states worse than dead is in the absence of a clear analytical framework [112]. Lead-time TTO is an example of a seemingly pragmatic solution to a serious methodological problem [83].

Furthermore, TTO does not take into account crucial requirements in measurement theory such as unidimensionality and the invariance principle [90,113–120]. Unidimensionality means that there is one dominant, relevant factor or dimension that is measured (e.g. valued). To fulfill the principle of invariance, the valuation of health states should be: i) independent of the group of respondents that performed the valuation task, and ii) independent of the set of health states being valued [90,118]. Both unidimensionality and invariance seem to be compromised in the TTO method, mainly because it is measuring two distinct elements: health status and life years [99]. To achieve unidimensionality, measuring only one thing at a time is standard practice in the natural sciences, such as physics, and it amounts to controlling all potential disturbance factors. Nunnally and Bernstein [119] put it concisely: 'a measure should generally concern some one thing – some distinct, unitary attribute' (p.4). In regard to the invariance principle, the TTO does not seem to be based on this important measurement consideration. Another conceptual measurement issue is the reference to TTO as a choice method. In fact, TTO could better be described as a matching-method. In matching tasks respondents are effectively asked to provide a number that will make them indifferent between the options [121,122].

Five years from now, patient-centered measurement will become more relevant than the current use of societal values of health states, derived by means of TTO or SG. Nowadays, for most preference-based instruments values are provided by respondents from the general population. The trend is that patient's values will more and more be used [123–126]. 'If the

public at large believes that those who experience health states are more reliable informants concerning the value of those health states, then the public would presumably choose to rely on these informants rather than on its own uninformed attitudes.' [89, p.91]. So, one reason to use patient values is that patients are likely to be more adequately informed than healthy people. Another reason is that they may be more adept at imagining certain health states and thereby be better positioned and motivated to make an informed judgment about the impact on perceived health of such states, especially when taking severely impaired health states into account [127]. Until now, there have been no convincing ways of adapting the TTO method to derive or estimate patients' values, mainly because of ethical issues. For example, using the TTO with its time element would imply asking critically ill individuals to value the time they might have left. If the patient community wants to have a central role in defining value, robust processes and other methods than TTO are needed to incorporate the patient voice in a form of value assessment that is free from adaptation and other biases.

The evolution of other measurement models is ongoing [120,128,129]. In research areas other than health economics, the evaluation of health is treated most often as an attitude [89,120]. Attitudes denote a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor [130–132]. One such an alternative method to quantify health states may be the discrete choice model [25,35]. This choice model is based on (paired) comparisons of two or more hypothetical health states. This method has also attracted attention by the EuroQol Group (EQ-5D instrument) [133]. Another comparison task may offer an attractive alternative, namely the multi-attribute profile best-worst task (or case 2), one of the three best/worst versions in the Best-Worst scaling framework [128]. In the profile task, respondents are confronted with one multi-attribute health-state description and are asked to indicate the best and the worst attribute levels. A drawback of best-worst scaling is that there is not yet a uniform statistical procedure to analyze the choice data. Another recently introduced patient-centered valuation method uses patients' input on all stages: selection of items for the classification system, describing own health condition, and assigning a value to their own health condition. This system is based on a combination of item response theory and discrete choice methods [126,129].

Other methods to measure subjective phenomena, far less complicated than the TTO, lack the possibility to produce measures that are anchored on dead = 0. Therefore, an attractive alternative is to combine two methods. Recently, a two-step procedure was developed for a preference-based health instrument for infants (Infant Quality of life Instrument: IQI) [134]. Coefficients for the levels were obtained from a DCE. These coefficients were then normalized using another DCE including 'dead' as a choice option. In this way, the values were rescaled from full health (1.0) to dead (0.0). A major benefit of such an approach is that the coefficients for the items are based on a simple and robust measurement method which will increase the reliability and validity of these coefficients. A similar two-step approach is proposed for the youth version of the EQ-5D, using a composite TTO to normalize the scale from step 1,

a conventional DCE [135]. Instead of a DCE with dead or time elements included, or a TTO to localize the position of dead, other methods for step 2 can be considered too [136].

The challenge is to develop inventive methods that overcome several shortcomings of the TTO and other conventional methods. Five years from now, modern software will be more widely used, simplifying the complexity of preference-based measurement, and putting the classification and valuation into the hands of patients. The full capacity of modern software entails more than just simply translations from paper-and-pencil form to an electronic mode. Simple to use apps are available, creating convenient, attractive tools for patients and researchers. The use of such apps in combination of modern measurement models will reduce the likelihood that patient responses are biased due to adaptation, strategic performance, or other mechanisms.

Funding

This work was funded by The EuroQol Research Foundation. The views expressed by the authors do not necessarily reflect the views of the EuroQol Research Foundation.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewers disclosure

Peer reviewers on this manuscript have no relevant financial relationships or otherwise to disclose.

References

Papers of special note have been highlighted as either of interest (*) or of considerable interest (**) to readers.

1. Gill TM, Feinstein ARA. Critical appraisal of the quality of quality-of-life measurements. *JAMA*. 1994;272(8):619–626.
2. Bonomi AE, Patrick DL, Bushnell DM, et al. Validation of the United States' version of the World Health Organization Quality of Life (WHOQOL) instrument. *J Clin Epidemiol*. 2000;53(1):1–12.
3. Sullivan M. The new subjective medicine: taking the patient's point of view on health care and health. *Soc Sci Med*. 2003;56:1595–1604.
4. Hamming JF, de Vries J. Measuring quality of life. *Br J Surg*. 2007;94:923–924.
5. Krabbe PFM. The measurement of health and health status: concepts, methods and applications from a multidisciplinary perspective. San Diego: Elsevier/Academic Press; 2016.
6. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35:1095–1108.
7. Richardson J, Sinha K, Iezz A, et al. Modelling utility weights for the Assessment of Quality of Life (AQoL) 8D. *Qual Life Res*. 2014;23:2395–2404.
8. Rowen D, Azzabi Zouraq I, Chevrou-Severac H, et al. International regulations and recommendations for utility data for health technology assessment. *Pharmacoeconomics*. 2017;35:11–19.
9. Sox HC, Higgins MC, Owens DK. Medical decision making. Oxford: John Wiley & Sons; 2013.
10. Fanshel S, Bush JW. A health-status index and its applications to health-services outcomes. *Oper Res*. 1970;18:1021–1066.
11. Torrance GW. A Generalized Cost-effectiveness Model for the Evaluation of Health Programs. [dissertation]. Buffalo (NY): State University of New York at Buffalo; 1971.

12. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res.* 1972;7:118–133.
13. Grogono AW, Woodgate DJ. Index for measuring health. *Lancet.* 1970;290:1024–1026.
14. Klarman HE, Francis JO, Rosenthal GD. Cost effectiveness analysis applied to the treatment of chronic renal disease. *Med Care.* 1968;6:48–54.
15. Zeckhauser R, Shepard DS. Where now for saving lives? *Law Contemp Probs.* 1976;40:5–25.
16. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *New Engl J Med.* 1977;296:716–721.
17. Selivanova A, Krabbe PFM. Eye tracking to explore attendance in health-state descriptions. *PLoS ONE.* 2018;13(1):e0190111.
18. Patrick DL, Bush JW, Chen MM. Methods for measuring levels of well-being for a health status index. *Health Serv Res.* 1973;8:228–245.
19. Bergner M, Bobbitt RA, Pollard WE, et al. The sickness impact profile: validation of a health status measure. *Med Care.* 1976;14:57–67.
20. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. *Health Serv Res.* 1976;11:478–507.
21. Rosser R, Kind P. A scale of valuations of states of illness: is there a social consensus? *Int J Epidemiol.* 1978;7:347–358.
22. Robinson A, Spencer A. Exploring challenges to TTO utilities: valuing states worse than dead. *Health Econ.* 2006;15:393–402.
23. Devlin NJ, Tsuchiya A, Buckingham K, et al. A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. *Health Econ.* 2011;20:348–361.
24. Janssen BM, Oppe M, Versteegh MM, et al. Introducing the composite time trade-off: a test of feasibility and face validity. *Eur J Health Econ.* 2013;14:S5–13.
25. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *PharmacoEconomics.* 2008;26:661–677.
26. Krabbe PFM. A generalized measurement model to quantify health: the Multi-Attribute Preference Response Model. *PLoS ONE.* 2013;8(11):e79494.
- **Comprehensive review of the basic measurement conditions required in science followed by an introduction of two prominent measurement models for subjective phenomena: the discrete choice model and the Rasch item response model. The former model can deal with multiple predictors (health attributes) and is based on expressing preference for one hypothetical health states over the other. The latter model deals with holistic descriptions but uses the individual respondents as a reference. It is shown that these two models can be merged.**
27. Murray CJ, Lopez AD. Quantifying disability: data, methods and results. *Bull World Health Organ.* 1994;72:481–494.
28. Wright J, Feinstein A, Alvan R. A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. *J Clin Epidemiol.* 1992;45:1201–1218.
29. Nord E. Methods for quality adjustment of life years. *Soc Sci Med.* 1992;34:559–569.
30. Salomon J. Techniques for valuing health states. In: Culyer AJ, editor. *Encyclopedia of Health Economics.* Amsterdam: Elsevier; 2014. p. 454–458.
31. Neumann von J, Morgenstern O. *Theory of games and economic behavior.* The 2004 edition. Princeton: Princeton University Press; 1944.
32. Moscati I. *Measuring utility: from the marginal revolution to behavioral economics.* New York: Oxford University Press; 2019.
33. Mas-Colell A, Whinston MD, Green JR. *Microeconomic Theory.* New York: Oxford University Press; 1995.
34. Thurstone LL. A law of comparative judgment. *Psychol Rev.* 1927;34:273–286.
- **Landscape publication. Many latter developments are based on the basic principles presented in this paper. Difficult to read and grab, because the terminology and statistical notation is different from what is used today.**
35. Thurstone LL. *The Measurement of Values.* Chicago: University of Chicago Press; 1959.
36. Moscati I. Early Experiments in Consumer Demand Theory: 1930–1970. *Hist Political Econ.* 2007;39(3):359–401.
37. McFadden D. The new science of pleasure: consumer choice behavior and the measurement of well-being. In: Hess S, Daly A, editors. *Handbook of Choice Modelling.* Cheltenham: Edward Elgar Publishing Limited; 2014;7–48.
38. Gafni AG. The standard gamble method: what is being measured and how it is interpreted. *Health Serv Res.* 1994;29:207–224.
39. Ginsberg AS, Offensend FL. An application of decision theory to a medical diagnosis-treatment problem. *IEEE Trans Syst Sci Cybern.* 1968;4:355–362.
40. McNeil BJ, Weichselbaum R, Pauker SG. Speech and survival: trade-offs between quality and quantity of life in laryngeal cancer. *New Engl J Med.* 1981;305:982–987.
41. McNeil BJ, Parker SG, Sox HC, et al. On the elicitation of preferences for alternative therapies. *N Engl J Med.* 1982;306:1259–1262.
42. Sunstein CR. *Valuing life: humanizing the regulatory state.* Chicago: The University of Chicago Press; 2014.
43. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. *Value Health.* 2009;12:S5–9.
- **Very readable introduction of the concepts and ideas associated to the computation and estimation of quality-adjusted life years.**
44. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ.* 2002;11:447–456.
45. Dolan P, Gudex C, Kind P, et al. The time trade-off method: results from a general population study. *Health Econ.* 1996;5:141–154.
46. Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Oper Res.* 1980;28:206–224.
47. Krabbe PFM, Bonsel GJ. Sequence effects, health profiles and the QALY model: in search of realistic modeling. *Med Decis Making.* 1998;18:178–186.
48. Torrance GW. Health states worse than death. In: van Eimeren W, Engelbrecht R, Flagle CD, editors. *Third international conference on system science in health care.* Berlin: Springer Verlag; 1984. p. 1085–1089.
49. Attema AE, Edelaar-Peters Y, Versteegh MM, et al. Time trade-off: one methodology, different methods. *Eur J Health Econ.* 2013;Jul (Suppl 1):53–64,14.
50. Wright DR, Wittenberg E, Swan JS, et al. Methods for measuring temporary health states for cost-utility analyses. *PharmacoEconomics.* 2009;27(9):713–723.
51. Oppe M, Devlin NJ, van Hout B, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health.* 2014;17(4):445–453.
52. Mulhern B, Bansback N, Brazier J, et al. Preparatory study for the revaluation of the EQ-5D tariff: methodology report. *Health Technol Assess.* 2014; report no.18.12:p. 1–191.
53. Gandhi M, Rand K, Luo N. Valuation of health states considered to be worse than death—an analysis of composite time trade-off data from 5 EQ-5D-5L valuation studies. *Value Health.* 2019;22(3):370–376.
54. Miyamoto JM, Eraker SA. Parameter estimates for a QALY utility model. *Med Decis Making.* 1985;5:191–213.
55. Buckingham K, Devlin N. A theoretical framework for TTO valuations of health. *Health Econ.* 2006;15:1149–1154.
56. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, et al. The measurement of patients' values in medicine. *Med Decis Making.* 1982;2:449–462.
57. Treadwell JR, Lenert LA. Health values and prospect theory. *Med Decis Making.* 1999;19:344–352.
58. Starmer C. Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *J Econ Lit.* 2000;38:332–382.
59. Bleichrodt H, Abellan-Perpinan JM, Pinto-Prades JL, et al. Resolving inconsistencies in utility measurement under risk: tests of generalizations of expected utility. *Manage Sci.* 2007;53:469–482.
60. Bleichrodt H, Johannesson M. The validity of QALYs: an experimental test of constant proportional trade-off and utility independence. *Med Decis Making.* 1997;17:21–32.

61. Dolan P, Stalmeier P. The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. *J Health Econ.* 2003;22:445–458.
62. Bleichrodt H, Pinto JL, Abellan-Perpiñan JM. A consistency test of the time trade-off. *J Health Econ.* 2003;22(6):1037–1052.
63. Attema AE, Brouwer WBF. The way that you do it? An elaborate test of procedural invariance of TTO, using a choice-based design. *Eur J Health Econ.* 2012;13:491–500.
64. Beresniak A, Medina-Lara A, Auray JP, et al. Validation of the underlying assumptions of the quality-adjusted life-years outcome: results from the ECHOUTCOME European project. *PharmacoEconomics.* 2015;33:61–69.
65. Loomes G, McKenzie L. The use of QALYs in health care decision making. *Soc Sci Med.* 1989;28:299–308.
66. Spencer A. A test of the QALY model when health varies over time. *Soc Sci Med.* 2003;57:1697–1706.
67. Treadwell JR. Tests of preferential independence in the QALY model. *Med Decis Making.* 1998;18:418–428.
68. Matza LS, Boye KS, Feeny DH, et al. The time horizon matters: results of an exploratory study varying the timeframe in time trade-off and standard gamble utility elicitation. *Eur J Health Econ.* 2016;17:979–990.
69. van Nooten FE, Koolman X, Busschbach JJ, et al. Thirty down, only ten to go?! awareness and influence of a 10-year time frame in TTO. *Qual Life Res.* 2014;23(3):377–384.
70. Lin MR, Yu WY, Wang SC. Examination of assumptions in using time tradeoff and standard gamble utilities in individuals with spinal cord injury. *Arch Phys Med Rehabil.* 2012;93:245–252.
71. Stiggelbout AM, Kiebert GM, Kievit J, et al. The “utility” of the time trade-off method in cancer patients: feasibility and proportional trade-off. *J Clin Epidemiol.* 1995;48:1207–1214.
72. van Nooten FE, Koolman X, Brouwer WB. The influence of subjective life expectancy on health state valuations using a 10 year TTO. *Health Econ.* 2009;18:549–558.
73. Richardson J. Evaluating Summary Measures of population Health. In: Murray CJL, Salomon JA, Mathers CD, et al., editors. Summary measures of population health: concepts, ethics, measurement and applications. Geneva: World Health Organization; 2002. p. 147–157.
- **Very readable overview and discussion on the research tradition of (health) economist in relation to the valuation of health by a critical health economist himself**
74. Dolan P. Modelling valuations for health states: the effect of duration. *Health Policy.* 1996;38:189–203.
75. Miyamoto JM, Eraker SA. A multiplicative model of the utility of survival duration and health quality. *J Exp Psychol Gen.* 1988;117(1):3–20.
76. Froberg DG, Kane RL. Methodology for measuring health-state preferences - I: measurement strategies. *J Clin Epidemiol.* 1989;42:345–354.
- **Clear introduction and overview of various methods to measure subjective phenomena, including some methods that were used in the early years of health valuation but are now abandoned or no longer in fashion.**
77. Lipman SA, Brouwer WBF, Attema AE. The corrective approach: policy implications of recent developments in QALY measurement based on prospect theory. *Value Health.* 2019;22(7):816–821.
78. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science.* 1981;211:453–458.
79. Augestad LA, Rand-Hendriksen K, Kristiansen IS, et al. 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;30:1203–1214.
80. Oppe M, Rand-Hendriksen K, Shah K, et al. EuroQol Protocols for Time Trade-Off Valuation of Health Outcomes. *PharmacoEconomics.* 2016;34:993–1004.
81. Shah KK, Lloyd A, Oppe M, et al. 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;14:65–73.
82. Luo N, Minghui L, Stolk A, et al. The effects of lead time and visual aids in TTO valuation: a study of the EQ-VT framework. *J Health Econ.* 2013;14:S15–S24.
83. Devlin N, Shah K, Feng Y, et al. Valuing Health-Related Quality of Life: an EQ-5D-5L Value Set for England. *Health Econ.* 2018;27(1):7–22.
84. Feng Y, Devlin NJ, Shah KK, et al. New methods for modelling EQ-5D-5L value sets: an application to English data. *Health Econ.* 2018;27:23–38.
- **Recent report of the modeling work done for the 5-level version of the English EQ-5D. Compared to the study for the 3-level version (Dolan 1997) the modeling has become even more complex**
85. Shah KK, Mulhern B, Longworth L, et al. An empirical study of two alternative comparators for use in time trade-off studies. *EuroQol Working Paper Series Number 15001 June 2015*, available at: <http://www.euroqol.org/about-eq-5d/working-paper-series.html>.
86. Green C, Brazier J, Deverill M. Valuing health-related quality of life. A review of health state valuation techniques. *PharmacoEconomics.* 2000;17:151–165.
87. Edelaar-Peeters Y, Stiggelbout AM, Hout van den WB. Qualitative and quantitative analysis of interviewer help answering the time tradeoff. *Med Decis Making.* 2014;34(5):655–665.
88. Kahneman D. Article commentary: judgment and decision making: a personal view. *Psychol Sci.* 1991;2(3):142–145.
89. Hausman DM. Valuing health: well-being, freedom, and suffering. Oxford: Oxford University Press; 2015.
90. Froberg DG, Kane RL. Methodology for measuring health-state preferences - II: scaling methods. *J Clin Epidemiol.* 1989;42:459–471..
- **Clear introduction and overview of various valuation methods, including some methods that were used in the early years of health valuation but are now abandoned or no longer in fashion.**
91. Ramos-Goñi JM, Oppe M, Slaap B, et al. Quality control process for EQ-5D-5L valuation studies. *Value Health.* 2017;20(3):466–473.
92. Shah K, Lloyd A, Devlin N. Participants’ responses to valuation tasks and implications for valuing EQ-5D-5L. In: Oxford 2011 EuroQol Proceedings. [cited 2020 Jun 9]. Available from: <https://euroqol.org>
93. Yang Z. Inconsistency in the valuations of Euroqol Eq-5d-5l Health States in China was More Related to Interviewer and to Interview Process than to Respondents’ Characteristics. *Value Health.* 2015;18:PA737–A738.
94. Krabbe PFM. Good day sunshine: about biases, irregularities and inconsistencies in the valuation of health states. In: York 2002 EuroQol Proceedings. [cited 2020 Jun 9]. Available from: <https://euroqol.org>
95. Arnesen TM, Norheim OF. Quantifying quality of life for economic analysis: time out for time trade off. *J Med Humanit.* 2003;29(2):81–86.
96. Devlin N, Shah K, Buckingham K. What is the normative basis for selecting the measure of “average” preferences for use in social choices? Office of Health Economics; 2017. Research Paper 201717/01.
97. Essink-Bot ML, Bonsel GJ. How to derive disability weights. In: Murray CJL, Salomon JA, Mathers CD, et al., editors. Summary measures of population health: concepts, ethics, measurement and applications. Geneva: World Health Organization; 2002. p. 449–465.
98. Salomon JA, Murray CJ. A multi-method approach to measuring health-state valuations. *Health Econ.* 2004;13:281–290.
99. Krabbe PFM. Valuation structures of health states revealed with singular value decomposition. *Med Decis Making.* 2006;26:30–37.
100. Johannesson M, Pliskin JS, Weinstein MC. A note on QALYs, time tradeoff, and discounting. *Med Decis Making.* 1994;14:188–193.
101. Neumann PJ, Sanders GD, Russell LB, et al. Cost-effectiveness in health and medicine. Oxford: Kindle Edition Oxford Scholarship; 2016.
102. Round J. Once bitten twice shy: thinking carefully before adopting the EQ-5D-5L. *PharmacoEconomics.* 2018;36:641–643.
103. EuroQol Group. visit May 6, 2020. Available from: <https://euroqol.org/update-on-the-eq-5q-5l-value-set-for-england>

104. Hernández-Alava M, Pudney S, Wailoo A (2018) "Quality review of a proposed EQ-5D-5L value set for England" Policy Research Unit in Economic Evaluation of Health and Care Interventions. Universities of Sheffield and York. EEPUR Research Report 060. [cited 2020 Jun 9]. Available from: <http://www.eepru.org.uk/validation-of-the-eq-5d-5l-valuation-set/>
105. Weinstein MC, Fineberg HV. Clinical decision analysis. Philadelphia: WB Saunders; 1980.
106. Sampson C, Devlin N, Parkin D Drop dead: is anchoring at 'dead' a theoretical requirement in health state valuation? EuroQol Plenary meeting, Lisbon, 20-21 sept, 2018.
107. Norman R, Mulhern B, Viney R. The impact of different DCE-based approaches when anchoring utility scores. *Pharmacoeconomics*. 2016;34:805–814.
108. Kamm FM. Morality, Mortality. Volume I: death and whom to save from it. New York: Oxford University Press; 1993.
109. Sutherland H, Llewellyn-Thomas H, Boyd NF, et al. Attitudes toward quality of survival the concept of "maximal endurable time". *Med Decis Making*. 1982;2:299–309.
110. Stalmeier P, Lamers L, Busschbach J, et al. On the assessment of preferences for health and duration: maximal endurable time and better than dead preferences. *Med Care*. 2007;45:835–841.
111. Scalone L, Stalmeier PFM, Milanis S, et al. Values for health states with different life durations. *Eur J Health Econ*. 2015;16(9):917–925.
112. Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. *Med Care*. 2007;45:238–244.
113. Luce RD, Tukey JW. Simultaneous conjoint measurement – a new type of fundamental measurement. *J Math Psychol*. 1964;1:1–27.
114. Coombs CH. A Theory of Data. New York: John Wiley & Sons; 1964.
 - **Compelling book with a theory about different types of data and what type of information is captured by these types of data. Also presenting a theory (Unfolding) how to deal with data collected by specific response tasks where we may assume an 'ideal point'. Stimulating book to reflect on assessment tasks, data and analysis.**
115. Suppes P, Krantz DM, Luce RD, et al. Foundations of measurement Vol. II: geometrical, threshold, and probabilistic representations. Mineola: Dover Publications; 1971.
116. Krantz DH, Luce RD, Suppes P, et al. Foundations of measurement, Vol. I: additive and polynomial representations. New York: Academic Press; 1971.
117. Luce RD, Krantz DH, Suppes P, et al. Foundations of measurement, Vol. III: representation, axiomatization, and invariance. New York: Academic Press; 1990.
118. Engelhard G Historical views of invariance: evidence from the measurement theories of Horndike, Thurstone, and Rasch. *Educ Psychol Meas*. 1992;2:275–291.
 - **This paper explains what is key in 'true' measurement.**
119. Nunnally JC, Bernstein IH. Psychometric Theory. New York: McGraw-Hill; 1994.
120. Arons AMM, Krabbe PFM. Probabilistic choice models in health-state valuation research: background, theory, assumptions and relationships. *Expert Rev Pharmacoecon Outcomes Res*. 2013;13:93–108.
 - **Presentation of various probabilistic choice models that may be relevant in the area of health valuation.**
121. Fischer GW, Carmon Z, Ariely D, et al. Goal-based construction of preferences: task goals and the prominence effect. *Manage Sci*. 1999;45:1057–1075.
122. Carson RT, Louviere JJ. A common nomenclature for stated preference elicitation approaches. *Environ resour econ*. 2011;49:539–559.
123. Calvert M, Kyte D, Price G, et al. Maximising the impact of patient reported outcome assessment for patients and society. *BMJ*. 2019;364:k5267.
124. Frank L, Basch E, Selby JV. The PCORI perspective on patient-centered outcomes research. *JAMA*. 2014;312(15):1513–1514.
125. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167.
126. Krabbe PFM, van Asselt ADI, Selivanova A, et al. Patient-centered item selection for a new preference-based generic health status instrument: CS-Base. *Value Health*. 2019;22:467–473.
127. Jonker MF, Attema AE, Donkers B, et al. Are health state valuations from the general public biased? A test of health state preference dependency using self-assessed health and an efficient discrete choice experiment. *Health Econ*. 2017;26(12):1534–1547.
128. Louviere JJ, Flynn TN, Marley AAJ. Best-Worst Scaling: theory, methods and applications. Cambridge: Cambridge University Press; 2015.
129. Groothuis-Oudshoorn CGM, van der Heuvel E, Krabbe PFM. An item response theory model to measure health: the multi-attribute preference response model. *BMC Med Res Methodol*. 2018;18:62.
 - **Clear presentation of this new valuation method. This article introduces and explains the MAPR model conceptually and mathematically. Results of a small empirical study are also presented to illustrate the procedures of the MAPR model and possible extensions of the model are discussed.**
130. Eagly AH, Chaiken S. The psychology of attitudes. Forth Worth: Harcourt, Brace & Jovanovich; 1993.
131. McFadden D. Rationality for economists? *J Risk Uncertainty*. 1999;19:73–105.
132. Kahneman D, Ritov I, Schkade D. Economic preferences or attitude expression?: an analysis of dollar responses to public issues. *J Risk Uncertainty*. 1999;19:203–235.
133. Krabbe PFM, Devlin NJ, Stolk EA, et al. Multinational evidence of the applicability and robustness of discrete choice modeling for deriving EQ-5D-5L health-state values. *Med Care*. 2014;52(11):935–943.
134. Krabbe PFM, Jabrayilov R, Detzel P, et al. A two-step procedure to generate utilities for the Infant health-related Quality of life Instrument (IQI). *PLoS ONE*. 2020;15(4):1–14.
135. Goñi JMR, Oppe M, Stolk E, et al. International valuation protocol for the EQ-5D-Y-3L. *Pharmacoeconomics*. 2020. DOI:10.1007/s40273-020-00909-3.
136. van Hoorn R, Donders A, Oppe M, et al. The better than dead method: feasibility and interpretation of a valuation study. *Pharmacoeconomics*. 2014;32(8):789–799.