VALUE IN HEALTH 22 (2019) 23–30

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Themed Section: Evolution of EuroQoL

Overview, Update, and Lessons Learned From the International EQ-5D-5L Valuation Work: Version 2 of the EQ-5D-5L Valuation Protocol

Elly Stolk, PhD [1,], Kristina Ludwig, PhD [1,2], Kim Rand, PhD [3,4], Ben van Hout, PhD [5], Juan Manuel Ramos-Goñi, PhD [1]*

[1]EuroQol Research Foundation, Rotterdam, The Netherlands; [2]Health Economics and Health Care Management, Bielefeld University, Bielefeld, Germany; [3]Department of Health Management and Health Economics, University of Oslo, Oslo, Norway; [4]Health Services Research Centre, Akershus University Hospital, Lørenskog, Norway; [5]School of Health and Related Research, University of Sheffield, Sheffield, UK

ABSTRACT

A standardized 5-level EuroQol 5-dimensional questionnaire (EQ-5D-5L) valuation protocol was first used in national studies in the period 2012 to 2013. A set of problems encountered in this initial wave of valuation studies led to the subsequent refinement of the valuation protocol. To clarify lessons learned and how the protocol was updated when moving from version 1.0 to the current version 2.1 and 2.0, this article will (1) present the challenges faced in EQ-5D-5L valuation since 2012 and how these were resolved and (2) describe in depth a set of new challenges that have become central in currently ongoing research on how EQ-5D-5L health states should be valued and modeled.

Keywords: discrete choice experiment, EQ-5D-5L, health valuation, time trade-off, utilities

Introduction

Experience accumulated during the valuation of the 3-level EuroQol 5-dimensional questionnaire (EQ-5D-3L), the original version of the EQ-5D, combined with the added complexities of the 5-level EQ-5D (EQ-5D-5L), led to the realization that there was a requirement for a new and standardized approach to valuation. Variation between protocols used to value the EQ-5D-3L has provided useful information about successful practices in health state valuation, but has hampered comparability of value sets. The introduction of the EQ-5D-5L[1] was an opportunity to take advantage of these experiences and introduce a standardized valuation protocol, supporting best practices and promoting comparability of valuation studies.[2] This formed the backdrop to systematic experimentation with valuation methods in the years around the introduction of the EQ-5D-5L and to subsequent decisions about the EQ-5D-5L valuation protocol. This protocol proved a major advance in that it was embedded in software. Continued use of time trade-off (TTO) preserved consistency with EQ-5D-3L value sets, and the discrete choices provided a convenient way to collect additional information about people's values for health and to exploit opportunities for hybrid modeling.

A description of version 1.0 of the EQ-5D-5L valuation protocol is provided by Oppe et al.[3,4] The protocol includes composite TTO (cTTO) and a discrete choice experiment (DCE). cTTO is a modified version of the conventional TTO variant that was used in the seminal Measurement and Valuation of Health study[5] and most subsequent EQ-5D-3L valuation studies. For the evaluation of health states considered to be better than dead, cTTO offers respondents the conventional task comprising a series of adaptive choices between *x* years in full health and 10 years in the disease state. In an iterative procedure, *x* is varied to identify the respondent's point of indifference where the health state value is given by *x*/10. When a respondent considers a health state to be worse than dead, lead-time TTO is used. Respondents are then offered a series of choices between *x* years in full health and a fixed life of 10 years in full health followed by 10 years in the target

* *Address correspondence to:* Elly Stolk, EuroQol Research Foundation, Marten Meesweg 107, 3068AV Rotterdam, The Netherlands. E-mail: stolk@euroqol.org

https://doi.org/10.1016/j.jval.2018.05.010

state. As before, x is varied until indifference is reached, and the health state is given by $(x - 10)/10$.[6,7]

Before the cTTO task, respondents receive a cTTO warm-up task featuring the health state "being in a wheelchair." Interviewers are instructed to use this example to explain the cTTO task and show the range of possible answers (ie, better and worse than dead). Next, respondents receive 10 real cTTO tasks and 7 choice sets from a DCE. The DCE task requires respondents to compare 2 health states and indicate which health state is better. The protocol enables analysts to implement hybrid models that draw on both types of data when generating their value sets, following recent literature.[8–11] The EQ-5D-5L valuation protocol was introduced with software for computer-assisted personal interviews, named the EuroQol Valuation Technique (EQ-VT), and interviewer instruction manuals.

The first use of the EQ-5D-5L valuation protocol (also referred to as the EQ-VT protocol) in national studies was in the period 2012 to 2013. A set of problems encountered in the cTTO data of this first wave of valuation studies led to the subsequent refinement of the valuation protocol. To clarify lessons learned and to show how the protocol was updated accordingly, this article will (1) present the challenges faced in EQ-5D-5L valuation since 2012 and how these have been addressed and (2) describe a set of new challenges that have become central in ongoing research with respect to how EQ-5D-5L health states could be valued and modeled.

Issues With cTTO Data

After the first tranche of EQ-5D-5L valuation studies, which were conducted in England, the Netherlands, China, Canada, and Spain, concerns were raised over observations of high rates of inconsistent responses, clustering of values, low values for mild states, and few worse than dead responses.[12–14] To define the problem, the valuation data have been analyzed in depth. This was made possible because the EQ-VT software captures the entire path that is followed to reach a value and the time stamps between mouse clicks. Exploiting the richness of these data, Ramos-Goñi et al[15] found that some interviewers systematically omitted explanation of the lead-time part of the cTTO task and elicited no worse than dead values. Furthermore, in some interviews, a very short time was spent on explaining the cTTO task and in obtaining a single cTTO value, which could indicate that respondents minimized effort and expedited the cTTO tasks by reducing the number of iterations. Because the iterative procedure requires a different number of steps to reach specific values (Fig. 1), lack of effort may partially account for the relatively low values for mild states (it takes more steps to reach high values) and clustering of values (a limited number of values can be attained when a cTTO task is completed with few iterations). Occurrence of these issues was found to vary across interviewers, suggesting that interviewer behavior affected the tendency for respondents to use such shortcuts.

When respondents make choices that result in quick task completion, this is indicative of respondent behavior that complies with the requirements of the task, but may still be detrimental to the precision of the answers that are obtained,[15] a general phenomenon referred to as *satisficing*.[16] Indirectly, this may also account for the large number of inconsistent valuations. Although one would expect that the worst state described by the EQ-5D-5L descriptive system would also receive the lowest value, roughly 20% of respondents gave at least 1 health state a lower value.[17] An inconsistency could itself be assigned to different causes, such as task complexity, random error, or learning effects, but it could also reflect inadequate efforts from respondents who did not feel compelled to expend resources on providing optimal answers.

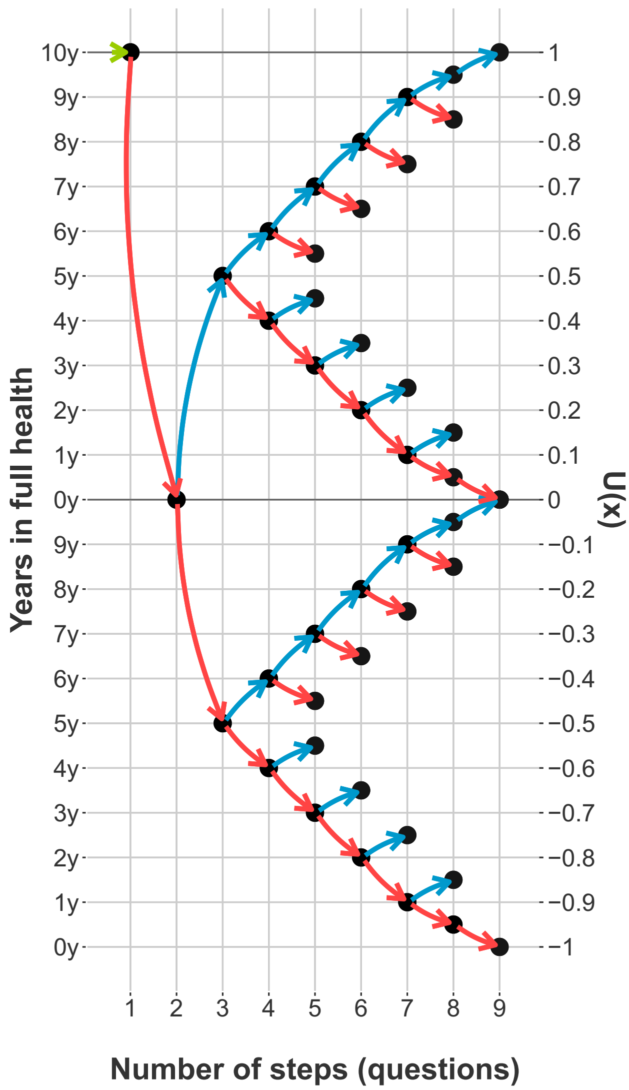

Fig. 1 – Routing implementation in the cTTO task. The left y-axis represents the bid in life A (number of years y in full health) that is compared with 10 y in the disease state *h* (better than dead) or 10 y in the disease state preceded by 10 y in full health (worse than dead). The right y-axis reflects the value of *h* that is obtained when the offered bid leads to an indifference statement. The arrows indicate how y is raised or lowered if a respondent selects A (red arrow) or B (blue arrow) to the bid previously offered. cTTO indicates composite time trade-off. The green arrow points to the opening bid.

The findings seem to reflect low levels of task engagement on the part of respondents or interviewers, with detrimental effects on the quality of the data. Herein we describe how this shaped our research in subsequent years, when we considered implications for the EQ-VT protocol and for data handling.

Updated Protocol for Implementing the cTTO Task (Version 1.1)

Table 1 presents how the protocol evolved over the years, resulting in the current version 2.1 that is presented in Table 2. The

Table 1 – Overview of valuation studies done using EQ-VT

Valuation study	EQ-VT elements								Extra elements tested in 2013 to inform decisions about EQ-VT version 2.0			
	EQ-VT v 1.0				EQ-VT v 1.1		EQ-VT v 2.0	EQ-VT v 2.1				
	Self-reported health	Background questions	cTTO	DCE	Practice states	QC monitoring	FB Module	Dynamic question after wheelchair example	Better than dead/worse than dead split	Ranking task	Routing and Iteration	Comparator full health
Canada	✓	✓	✓	✓								
China	✓	✓	✓	✓								
England	✓	✓	✓	✓								
Netherlands	✓	✓	✓	✓								
Spain	✓	✓	✓	✓								
Japan	✓	✓	✓	✓	✓	✓				✓		
South Korea	✓	✓	✓	✓	✓	✓						
Thailand	✓	✓	✓	✓	✓	✓						
Uruguay	✓	✓	✓	✓	✓	✓						
Australia	✓	✓	✓	✓	✓	✓	✓					
Germany	✓	✓	✓	✓	✓	✓	✓		✓	✓		
Hong Kong	✓	✓	✓	✓	✓	✓	✓		✓			
Indonesia	✓	✓	✓	✓	✓	✓	✓					
Ireland	✓	✓	✓	✓	✓	✓	✓					
Malaysia	✓	✓	✓	✓	✓	✓	✓					
Poland	✓	✓	✓	✓	✓	✓	✓					
Portugal	✓	✓	✓	✓	✓	✓	✓					
Singapore	✓	✓	✓	✓	✓	✓	✓				✓	
Taiwan	✓	✓	✓	✓	✓	✓	✓					
Ethiopia	✓	✓	✓	✓	✓	✓	✓	✓				
France	✓	✓	✓	✓	✓	✓	✓	✓				
Philippines	✓	✓	✓	✓	✓	✓	✓	✓				
Unites States	✓	✓	✓	✓	✓	✓	✓	✓				
Vietnam	✓	✓	✓	✓	✓	✓	✓	✓				
Methodological studies done in 2013 to inform decisions about EQ-VT v 2.0												
The Netherlands	✓	✓	✓	✓	✓	✓	✓		✓			
Spain	✓	✓	✓	✓	✓	✓	✓			✓		
England	✓	✓	✓	✓	✓	✓						✓
Norway	✓	✓	✓	✗	✓	✓	✓				✓	

Note. The check mark shows that an element was included; the cross mark indicates that an element of the protocol version used in the study was dropped.
cTTO indicates composite time trade-off; DCE, discrete choice experiment; EQ-VT, EuroQol Valuation Technology; QC, quality control.

identifier conveys the type of change between versions: the first position is changed for introduction of a significant task element, whereas the decimal changes with the way of implementing the tasks. In version 1.1 of the EQ-VT protocol, we implemented several suggestions aimed at capturing better data from the existing valuation tasks. We added 3 practice states following the wheelchair example to familiarize respondents with the cTTO task and with the severity range of health state descriptions. In addition, confirmatory pop-ups were implemented to validate answers before storing them. Furthermore, we started to monitor interviewer performance during data collection to enable timely intervention if problems were detected. We introduced a quality control (QC) procedure to review protocol compliance and interviewer effects while the study was ongoing.[15] Also, the QC report flagged up interviews as being of potentially poor quality when any 1 of the following 4 criteria was met.

Table 2 – Elements of the EQ-5D-5L valuation protocol 2.0.

Start interview

1. General welcome
2. Introduction
   - Self-reported health on the EQ-5D-5L descriptive system
   - Self-reported health on the EQ-VAS
   - Background questions
3. cTTO
   - Instructions and example of cTTO task, 3 practice states
   - cTTO valuation of 10 EQ-5D-5L states
   - cTTO debriefing/structured feedback
   - cTTO feedback module
4. Discrete choice
   - Instructions of discrete choice task
   - Discrete choice valuation of 7 pairs of EQ-5D-5L states
   - Discrete choice debriefing/structured feedback
5. General thank you and goodbye

End interview

Accompanying: a quality control process

cTTO indicates composite time trade-off; EQ-5D-5L, 5-level EuroQol 5-dimensional questionnaire; EQ-VAS, EuroQol visual analogue scale.

1. no explanation of the worse than dead task (lead time) in the wheelchair example;
2. too short time period spent on the wheelchair example (<3 minutes);
3. clear inconsistency in the cTTO ratings (55555 not the lowest and at least 0.5 higher than the state with the lowest value); and
4. 10 cTTO tasks completed in less than 5 minutes.

Initial QC reports were used to evaluate whether interviewers met minimum quality requirements. If the rate of flagged interviews of an interviewer was 40% or more in the first 10 interviews, then those interviews were considered ineligible and the interviewer would be retrained. After another set of 10 interviews, interviewer performance was evaluated again to decide whether the interviewer could continue or should be discharged, with removal of that interviewer's data as a consequence. A lenient threshold value of 40% was used because a good interview could not be identified unambiguously (flagged interviews could hold genuine responses), and to allow interviewers to grow into their roles when they built up experience. Later, QC reports allowed the team to continuously reflect on interviewers' performance and to discuss potential improvements. The QC process was implemented with the recommendation that each interviewer should contribute about 100 interviews to allow the establishment of effective feedback loops.

Although all new requirements increase study cost and may lead to removal of data, the effectiveness of the process is undisputed. Interviewer effects, clustering, and inconsistencies were dramatically reduced in studies that adhered to the updated protocol.[15]

EQ-VT Version 2.0: Introduction of the Feedback Module

Several other suggestions for remedying data issues were presented in a modified cTTO task, each connecting different possible causes for data issues with strategies for their mitigation. A research program was carried out to test the proposed cTTO task modifications (included in Table 1) and to assist in informing an update of the valuation protocol to version 2.0. Teams from Spain, Japan, the Netherlands, Hong Kong, Norway, Singapore, Germany, and England were involved in the research program, each allocating participants to a control group that received version 1.1 of the protocol or to an experimental group that received a modified protocol.[17] This work has resulted in the inclusion of a feedback module in version 2.0 of the EQ-VT protocol.

The feedback module (Fig. 2) shows respondents what rank ordering of health states would be inferred from their cTTO responses. The health state with the highest value is presented at the top, the lowest at the bottom, and ties side by side. The color of the box containing a single health state indicates whether that particular state was considered better (light blue) or worse than dead (dark blue). Respondents can indicate disagreement with the implied rank ordering by clicking on the offending health state(s). No attempt is made to derive new cTTO values. Shah et al[17] reported that 1 in 3 respondents reconsidered 1 or more of their initial responses. This in turn lowered the number of inconsistencies.

The research program provided no support for the other modifications tested: separation of the better and worse than dead task in cTTO, reintroduction of a ranking task for warm-up

Fig. 2 – Example of feedback module in EQ-VT. EQ-VT indicates EuroQol Valuation Technology.

purposes, relabeling the comparator state in cTTO, and the use of a different iteration procedure. These modifications had limited effects on the data, which could reflect diminished scope for further improvements in cTTO after implementing QC. cTTO, combined with the QC process outlined earlier, thus appears to offer researchers a solid basis for constructing their EQ-5D-5L value function.[17]

Currently, we are at version 2.1 of the software. This new version was released in order to promote compatibility of the software with interviewer instructions. In the normal flow of a cTTO task, a better than dead or a worse than dead response is obtained, but interviewers are supposed to show both parts of the evaluation space using the wheelchair example. This was easy to forget and difficult to do. After completing the cTTO task for the wheelchair example, we now include a dynamic question: Depending on the respondent's response for the wheelchair, he is asked to imagine a health state that is much better or much worse in order to move to the other part of the evaluation space.

Discrete Choice Experiment

In the EQ-VT protocol, the cTTO task has been accompanied by a DCE task to collect additional information about health state values. The DCE task comprises pairwise comparisons of 2 EQ-5D-5L health states and produces by itself no knowledge of how the derived values relate to the full health (1) to death (0) scale. Thus, DCE values are derived on a latent scale and are not suited to replace the cTTO task as a basis for value set generation. Nevertheless, assuming that an individual's cTTO and DCE responses both reflect the same underlying preferences for health, and that DCE- and cTTO-based value functions are the same up to a linear transformation ($\beta_{TTO} = \theta \times \beta_{DCE}$), it may be considered appropriate to complement cTTO data with DCE data. This enables a search for a common set of parameter values using a hybrid model that maximizes the likelihood of all observations collected in the cTTO and DCE tasks.[18,19]

In the Netherlands, Canada, South Korea, Uruguay, and China, value algorithms were developed on the basis of only cTTO data,[13,20–23] whereas in Spain and England the hybrid model was adopted.[12,14,24] The decisions were often based on considerations specific to the local context. For example, lack of software and of documentation initially constituted barriers to implementation of the hybrid model, and properties of the local data made some teams push to use a hybrid model while others chose not to do so.

Theoretically, much depends on the interpretation of possible differences between cTTO- and DCE-derived values. The two methods tended to lead to agreement on the utility loss associated with each severity level,[13,24] but perfect agreement should not be expected. The approaches derive values using different types of tasks, matching, or choice, and on the basis of different measurement models; in cTTO, all health states are valued against time, whereas in DCE values are derived on the basis of how dimension severity levels are traded off against each other. Hence, there can be concerns with the commensurability of the methods or the axioms that underpin them. Nevertheless, if one assumes that both valuation methods capture important aspects of preferences for health and that neither method does so perfectly, these concerns could be alleviated. Under these assumptions, hybrid models using cTTO and DCE data are superior in the sense that they make use of all the available data.[10] The inclusion of more data also can increase the power of the model and result in more precise parameter estimates.

By enabling hybrid models, DCE data serve an important purpose in the EQ-VT protocol. The DCE data generated have served this purpose well, although some aspects of the experimental design may be enhanced. The DCE includes 196 pairs that have been selected using a Bayesian efficient design algorithm that requires priors for the model's parameters, with additional hand-picked pairs comparing 2 mild states.[3,25] When modeling the data, violations of logical orderings of the dimension severity levels were observed in several countries, particularly pertaining to severity levels 2 and 3. This distinct problem may be resolved with an update of the design (eg, based on better priors).[26] Nevertheless, work on this has been postponed because we have been contemplating a bigger update of the DCE task, as described in a later section.

Analytical Advances

Although key features of the raw cTTO data have been analyzed in recent years, researchers have also come to investigate what these findings imply for modeling of the data. EQ-5D-3L TTO data were often modeled using simple linear regression. Modeling approaches of EQ-5D-5L cTTO data are new in that they commonly account for censoring, heteroskedasticity, truncation, and/or preference heterogeneity.[13,14,24,27] The modeling advances were driven by considerations obtained from carefully investigating aspects of the cTTO task and the data it provided, and by matching these to the assumptions underlying the regression models. These considerations can be categorized into 3 groups related to (1) mechanics of the cTTO task itself, (2) individual respondent behavior, and (3) characteristics of the complete cTTO data set, as presented in Table 3. Even when improved QC significantly curtails the range of issues that occur in cTTO data, the issues described are multicausal and will likely keep recurring, but at a reduced level. The new modeling approaches thus appear to be generally applicable and to represent best practice.

An obvious reason to adopt a framework for censored data is that cTTO data are left-censored at −1, but it is innovative that we also consider the presence of other types of censoring. Table 3 presents several factors contributing to the view that cTTO responses can have low levels of accuracy and hence may be better construed as indicating ranges of values within which the point of indifference is likely to reside, rather than discrete indifference points. For example, left-censoring may not be limited to the bottom value of the scale but could also occur at 0 for people who were not properly introduced to the worse than dead task. Satisficing can occur if people complete the task half-heartedly or provide crude responses because their preferences are not well defined. Then, the quality of the data is downgraded and values are potentially biased; for example, values for mild states can be downward-biased because it requires many moves in the cTTO task to reach high values (see Fig. 1). Frameworks for censored data have the potential to mitigate such bias and are increasingly used.[14,24,27]

A further innovation was the introduction of models that accommodate heteroskedasticity and non-normality.[14,21,27–29] In health state valuation, variability increases with severity; there is little disagreement that mild health states are good, but opinions diverge about how bad moderate and severe states really are (ie, heteroskedasticity of error terms).[30] A cause for non-normality of the error terms is that cTTO values have a maximum of 1, so that the value range and error distributions are truncated. When relatively mild health states are valued, many values at 1 or close to 1 will be obtained, resulting in a skewed error distribution. Outliers, if not handled properly, can result in estimates that are too low, especially for mild states. For this reason, the Uruguayan EQ-5D-5L tariff was produced using robust regression,[20] but models for censored data can also accommodate heteroskedastic data. Models for censored data can address several considerations simultaneously, which explains their growing popularity in

Table 3 – Overview of phenomena that characterize cTTO data and how they can be modeled.

Phenomenon	Cause	Consequence	Possible solution
cTTO task mechanics	Censoring of values below −1	For observations at −1 the true value equals −1 or a lower value	Tobit or interval regression
	Left-skewed value distributions at values close to 1	Models that assume normally distributed errors can produce biased estimates	Tobit or interval regression
	The smallest tradable unit (in EQ-VT 6 mo) limits precision of assessed values	True values equal observed values ±0.025	Interval regression
Individual respondent behavior	Satisficing: no effort from the participants to precisely express their values	Routing biases observed values	Interval regression
	Respondents do not have well-defined preferences	Routing biases observed values	Interval regression
	Time preferences make an individual's values cluster at the top or bottom of the scale*	No discriminative ability within cluster	Tobit or interval regression
Characteristics of the complete cTTO data set	The SD around observed values increases with worsening quality of life	Models that assume identical errors produce biased estimates	Heteroskedastic models
	Some people consider all health states to be better than dead; others have worse than dead preferences	Values of different subgroups may require different treatment (Feng et al[24])	Latent class models

cTTO indicates composite time trade-off; EQ-VT, EuroQol Valuation Technique.
* As discussed by Boye et al[44] and by Van der Pol and Shiell.[45]

handling EQ-VT valuation data, as illustrated by some of the latest articles on EQ-5D-5L value sets.[14,27–29]

Another way of dealing with heteroskedastic data considers that it might result from the presence of population subgroups with different preferences. In particular, the distinction between people who consider all health states to be better than dead and people who consider some health states to be worse than dead causes variance in error terms to increase with severity. Values obtained from respondents within a subgroup may be similar, but across subgroups the values can be very different. This recognition motivated researchers from England to analyze cTTO data using a latent class model.[12,24] Their model assumed that all respondents assigned relative weights to the dimension severity levels in a similar way, but differed in their views about the position of dead, and hence their value functions had different slopes. A latent class model was used to identify subgroups and to estimate a parameter for the slope in each group ($n = 3$ in England) and the probability of respondents being in each group. A population value set was derived by computing the weighted average slope using these 6 numbers.

A 5-Year View of EQ-5D-5L Valuation

Although the advances in collecting, interpreting, and analyzing cTTO health state valuation data have boosted confidence that teams can successfully produce a valid EQ-5D-5L tariff, the EQ-VT protocol development may not end at version 2.1. Currently, we are looking at the possibilities of promoting the feasibility of valuation studies in a wider context. Interest in health state valuation studies appears to be increasing, including from small countries with fewer resources. Arguably, reliance on the cTTO task as the principal valuation method is a drawback of the current protocol, because cTTO is costly and inherently difficult to implement by investigators who have little experience with health valuation. This motivated research into the potentialities of taking more advantage of DCE.

One approach that aimed at deriving greater benefits from DCE involves further exploitation of the hybrid model. Rowen et al[11] indicated that for the application of a hybrid model on DCE and cTTO data, it is not necessary to have a high number of health states in the cTTO task. Slimming down the cTTO part of EQ-VT has obvious benefits, albeit restricted because face-to-face interviews will still be required and the learning curve of cTTO is not avoided. Alternatively, we have considered the scope for DCE as a primary health state valuation method. Although the current DCE task derives values on a latent scale, a variant of this DCE task could include the duration of a health state as an extra attribute. This would allow health state values to be defined as the product of the quantity and quality of life, with the derived values anchored at full health and dead.[31] An advantage of the "DCE duration" approach is that it derives values from the same conceptual model as cTTO, but from a task that avoids the complex iterative procedure.[31–33] In terms of benefits, exploiting the hybrid model could yield results faster, whereas DCE duration might achieve greater benefits.

Although the promises of DCE duration were widely recognized when the EQ-VT protocol was developed, the method has not yet been implemented as a possible substitute for cTTO because of concerns with respect to the low values that were obtained in some initial applications.[31,34,35] We considered that the discrepancies might be, at least partially, explained by differences between the 2 approaches in anchoring at dead. Because cTTO requires observed values, this method includes a task to assess the strength of preferences of health states that are classified as worse than dead (ie, lead-time TTO). In contrast, the way in which the stimuli in the DCE duration task are shown implies that choices never indicate directly whether a health state has a worse than dead value. In DCE, extrapolation is required and this comes with extra uncertainty and has a potential for bias if the assumptions underlying the extrapolations are wrong. Although some research funded by the EuroQol Research Foundation in this area is still ongoing, preliminary results show that DCE duration estimates are sensitive to model specification and in particular to assumptions made regarding duration preferences.[36,37] Models applied to cTTO commonly assume that the same proportional trade-offs would be made regardless of whether the remaining life expectancy is long or short (<2 years), which may not hold.[38,39] Nevertheless, violations of this assumption can be a bigger problem for DCE duration than for cTTO, because of the required extrapolation. The consequences of these findings need to be considered.

Conclusions

The lessons learned from EQ-5D-5L valuation have resulted in a detailed valuation protocol, paired with a quality assurance process and novel analytical approaches. The updated protocol has enabled teams from all over the world to establish EQ-5D-5L value sets.

The collective efforts that led to the refinement of the protocol demonstrate that the EuroQol Group is dedicated to learn about valuation and willing to revise protocol elements to improve the validity and reliability of the results. Although TTO has been a preferred method for health state valuation for 2 decades, we have obtained important new insights concerning how respondent behavior and features of the task work together to define the level of precision of cTTO responses. These insights have emphasized the importance of the interviewer's role in motivating respondents to give accurate responses and reacting to certain respondent behaviors. We came to realize that demands placed on the interviewer are high, making it unlikely that any interviewer training, script, or software for performing interviews will be sufficient to guarantee proper interviewer performance. This motivated the introduction of a QC process and of new modeling approaches.

Reflecting the nature of the most scientific endeavor, improvement of valuation methods is a gradual process, and value sets derived from procedures that were state-of-the-art at the time could have benefited from later developments. It is likely that 3-level value sets suffer from several of the issues described here, and that the most recent 5-level valuation studies are improvements over the older 5-level value sets. Similarly, strategies used to produce value sets for other questionnaires such as the 6-dimensional health state short form or the Health Utilities Index have been subject to modification in the light of new empirical results.[40–43] It follows that value sets should not be considered as permanent entities, but should have an expected life cycle with anticipated updates or replacements that reflect gradual changes in society, health preferences, and improved methods. How often value sets should be updated is a question without any definite answer.

Given the changes in perception about what works in valuation, it is appropriate to recognize that the rigorous approach to EQ-5D-5L valuation studies inspires trust. Yet we can also note that the kind of insights that guide our current valuation work did not exist when the EQ-5D-3L was originally valued. It is wise to remain modest with respect to claims concerning the qualities of older value sets, until we have returned to them and scrutinized the methods used in constructing these value sets. Such research is warranted to provide users and stakeholders with recommendations regarding the use of the instruments in analysis and healthcare decision making.

REFERENCES

1. 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.
2. Devlin NJ, Krabbe PFM. The development of new research methods for the valuation of EQ-5D-5L. Eur J Health Econ 2013;14(suppl 1):S1–3.
3. 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.
4. Oppe M, Rand-Hendriksen K, Shah K, Ramos-Goñi JM, Luo N. EuroQol protocols for time trade-off valuation of health outcomes. Pharmacoeconomics 2016;34(10):993–1004.
5. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35(11):1095–108.
6. Janssen BMF, 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.
7. Devlin NJ, Tsuchiya A, Buckingham K, Tilling C. A uniform time trade off method for states better and worse than dead: feasibility study of the "lead time" approach. Health Econ 2011;20:348–61.
8. Oppe M, Ramos-Goñi JM, van Hout B. Modeling EQ-5D-5L valuation data. EuroQol Group. http://eq-5dpublications.euroqol.org/download?id=0_53782&fileId=54196. Accessed June 6, 2017.
9. Ramos-Goñi JM, Craig BM, Oppe M, et al. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. Value Health 2018;21(5):596–604.
10. Craig BM, Busschbach JJ. The episodic random utility model unifies time trade-off and discrete choice approaches in health state valuation. Popul Health Metr 2009;7:3.
11. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. Med Decis Making 2015;35(3):328–40.
12. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-442 5L value set for England. Health Econ 2018;27(1):7–22.
13. Versteegh MM, Vermeulen KM, Evers SMAA, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. Value Health 2016;19(4):343–52.
14. Ramos-Goñi JM, Pinto-Prades JL, Oppe M, Cabasés JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. Med Care 2017;55(7):e51–8.
15. Ramos-Goñi JM, Oppe M, Slaap B, Busschbach JJV, Stolk E. Quality control process for EQ-5D-5L valuation studies. Value Health 2017;20(3):466–73.
16. Craig BM, Runge SK, Rand-Hendriksen K, Ramos-Goñi JM, Oppe M. Learning and satisficing: an analysis of sequence effects in health valuation. Value Health 2015;18(2):217–23.
17. Shah K, Rand-Hendriksen K, Ramos 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. EuroQol Group. http://eq-5dpublications.euroqol.org/download?id=0_53918&fileId=54332. Accessed June 6, 2017.
18. Oppe M, van Hout B. The optimal hybrid: experimental design and modeling of a combination of TTO and DCE. Paper presented at: The 27th Plenary Meeting of the EuroQol Group, September 16-17, 2010. Athens, Greece, http://eq-5dpublications.euroqol.org/download?id=0_53738&fileId=54152. Accessed June 6, 2017.
19. Ramos-Goñi JM, Craig B, Oppe M, Van Hout B. Combining continuous and dichotomous responses in a hybrid model. 2016. EuroQol working paper series 16002, https://euroqol.org/publications/working-papers/. Accessed June 6, 2017.
20. Augustovski F, Rey-Ares L, Irazola V, et al. An EQ-5D-5L value set based on Uruguayan population preferences. Qual Life Res 2016;25(2):323–33.
21. Kim SH, Ahn J, Ock M, et al. The EQ-5D-5L valuation study in Korea. Qual Life Res 2016;25(7):1845–52.
22. Xie F, Pullenayegum E, Gaebel K, et al; Canadian EQ-5D-5L Valuation Study Group. A time trade-off-derived value set of the EQ-5D-5L for Canada. Med Care 2016;54(1):98–105.
23. Luo N, Liu G, Li M, Guan H, Jin X, Rand-Hendriksen K. Estimating an EQ-5D-5L value set for China. Value Health 2017;20(4):662–9.
24. Feng Y, Devlin NJ, Shah KK, Mulhern B, van Hout B. New methods for modelling EQ-5D-5L value sets: an application to English data. Health Econ 2018;27(1):23–38.
25. Oppe M, van Hout B. The "power" of eliciting EQ-5D-5L values: the experimental design of the EQ-VT. 2017. EuroQol working paper 17003, https://euroqol.org/publications/working-papers/. Accessed June 6, 2018.

26. Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health 2013;16(1):3–13.
27. Ludwig K, Graf von der Schulenburg JM, Greiner W. German value set for the EQ-5D-5L. Pharmacoeconomics 2018;36(6):663–74.
28. Wong ELY, Ramos-Goñi JM, Cheung AWL, Wong AYK, Rivero-Arias O. Assessing the use of a feedback module to model EQ-5D-5L health states values in Hong Kong. Patient 2018;11(2):235–47.
29. Purba FD, Hunfeld JAM, Iskandarsyah A, et al. The Indonesian EQ-5D-5L value set. Pharmacoeconomics 2017;35(11):1153–65.
30. Cameron AC, Trivedi PK. Microeconometrics Using Stata. Revised ed. College Station, TX: Stata Press; 2010.
31. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. J Health Econ 2012;31(1):306–18.
32. Flynn TN, Louviere JJ, Marley AA, Coast J, Peters TJ. Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. Popul Health Metr 2008;6:6.
33. Norman R, Mulhern B, Viney R. The impact of different DCE-based approaches when anchoring utility scores. Pharmacoeconomics 2016;34(8):805–14.
34. Gu Y, Norman R, Viney R. Estimating health state utility values from discrete choice experiments—a QALY space model approach. Health Econ 2014;23(9):1098–114.
35. Craig BM, Greiner W, Brown DS, Reeve BB. Valuation of child health-related quality of life in the United States. Health Econ 2016;25(6):768–77.
36. Jakubczyk M, Craig BM, Barra M, et al. Choice defines value: a predictive modeling competition in health preference research. Value Health 2018;21(2):229–38.
37. Lim S, Jonker M, Oppe M, Donkers B, Stolk E. Severity-stratified discrete choice experiment designs for health state evaluations. Pharmacoeconomics 2018;36(11):1377–89.
38. 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(8):979–90.
39. Stiggelbout AM, Kiebert GM, Kievit J, Leer JW, Habbema JD, De Haes JC. The "utility" of the time trade-off method in cancer patients: feasibility and proportional trade-off. J Clin Epidemiol 1995;48:1207–14.
40. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system: health utilities index mark 2. Med Care 1996;34(7):702–22.
41. Feeny DH, Furlong WJ, Torrance GW, et al. Multi-attribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care 2002;40(2):113–28.
42. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21(2):271–92.
43. Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 health survey. J Clin Epidemiol 1998;51(11):1115–28.
44. 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.
45. Van der Pol M, Shiell A. Extrinsic goals and time tradeoff. Med Decis Making 2007;27(4):406–13.