
Julija Stoniute, BA1, David John Mott, MSc1,2, Jing Shen, PhD1,*

1Health Economics Group, Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK; 2Office of Health Economics, London, UK

ABSTRACT

Background: The time trade-off (TTO) technique is commonly used to elicit health state utilities. Nevertheless, when the health states being valued are temporary, the TTO approach may be unsuitable. A variant of TTO—chained TTO—has been suggested to be used when the health states are temporary, but little research has been done on how chained TTO should be conducted. Objectives: To systematically review the use of chained TTO in valuing temporary health states. Methods: A systematic literature search was conducted using the following major databases: Ovid MEDLINE(R), Embase, EBM Reviews, and PsycINFO. Abstracts (full articles if necessary) were screened by two independent reviewers, with a third reviewer resolving any disagreements. Results: The resulting number of articles for review was low (n = 9). All the reviewed studies used face-to-face interviews, most had small sample sizes (<100), and all studies valued a small number of health states (<7), with time horizons typically ranging from 4 weeks to 1 year. All studies discussed methodological issues of using chained TTO, and some compared the results with those generated using other preference elicitation methods. Conclusions: Chained TTO appears to be feasible, consistent, and responsive and allows the valuation of temporary health states that would improve the efficiency and accuracy of decision making in health and health care. Nevertheless, the evidence is limited due to the low number of relevant studies in the literature. Further research is needed to examine the performance and validity of chained TTO compared with conventional TTO in the valuation of temporary health states. Keywords: health state utilities, preference elicitation, systematic review, temporary health states, time trade-off.

Copyright © 2017, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

One of the most common forms of economic evaluation in health and health care is cost-utility analysis (CUA), in which benefits are typically expressed using quality-adjusted life-year (QALYs) [1]. A common valuation exercise to elicit health state utilities used in deriving QALYs is the time-trade-off (TTO) method, which is usually used to value chronic health states. However, there are concerns over its applicability in situations in which the health states are temporary [2]. There has not been a clear definition of a temporary health state, but the convention in the literature is that a health state lasting less than 1 year [2] or simply a short-term state followed by a return to health [3] can be considered temporary. The concern with using TTO for temporary health states is largely due to one of the restrictive assumptions of the QALY model—constant proportional TTO [4]. This assumption requires that the value of a health state is unaffected by the duration used in the valuation task; however, it has been shown that this does not always hold [5]. In the case of temporary health states, the constant proportional TTO assumption becomes more problematic because of the potential use of short time horizons, which provide respondents with a scenario that includes an imminent death. In 1986, Torrance [6] outlined a variant of TTO that could be used to value temporary health states referred to as “chained TTO.” It is a two-stage preference elicitation method that avoids trading life with a short time span in the first stage, which has the benefit of making the task more realistic in the context of temporary health states.

There has been no scientific consensus on the optimal specification of the chained TTO task. This makes conducting the chained TTO difficult in practice, reduces the comparability of studies that use this method, and also discourages potential uses of this method because of the barrier of not having a well-established method guide. The first step to overcoming the aforementioned issues would be to systematically review and appraise the existing practice of chained TTO in the valuation of temporary health states. The aim of this article is, therefore, to fill this gap in the literature and provide suggestions for future research. The rest of the article is organized as follows. The second section of the article describes conventional and chained TTO, the third section outlines the literature search strategy, the fourth section reviews the retrieved articles, the fifth section...
discusses the findings, and the last section concludes and suggests areas for further research.

**Conventional and Chained TTO**

Conventional TTO involves a choice between two hypothetical scenarios: respondents are asked to choose between a health state \(i\) for a given duration \(T\), and perfect health for a duration between 0 and \(T\), both followed by death [6]. The respondent is asked which scenario he or she prefers, with the duration in perfect health being varied between 0 and \(T\) until the respondent is indifferent between the two scenarios. The duration in perfect health when the respondent expresses this indifference \((X)\) is then used to generate the utility value for \(i\) using Equation 1:

\[
U_i = \frac{X}{T}.
\]

In contrast, chained TTO consists of two stages. In the first stage, respondents choose between two hypothetical scenarios: a (temporary) health state \(i\) for a given duration \(T\), and an anchor health state \(j\) for a duration between 0 and \(T\), both followed by a return to perfect health [6]. The anchor state must be worse than the temporary health state but better than death \((i > j > 0)\). The length of time in the anchor state \(j\) is varied until the respondent is indifferent between the two scenarios at duration \(X_j\). Hence, the first stage elicits the respondent’s preference for health state \(i\) relative to anchor state \(j\).

The second stage values the anchor state \(j\) using the conventional TTO method, in which respondents are asked to choose between the anchor health state \(j\) for the given duration \(T\), and perfect health for a duration between 0 and \(T\), both followed by death. The length of time in perfect health is varied until the respondent is indifferent between the two scenarios at duration \(X_j\). Utility \(U_j\) for the temporary health state \(i\) is calculated using Equation 2:

\[
U_i = 1 - \left(1 - U_j \right) \frac{X_j}{T},
\]

where utility \(U_j\) for the anchor state \(j\) is calculated using Equation 3, as in conventional TTO:

\[
U_j = \frac{X_j}{T}.
\]

It should be noted that chained TTO has also been used to value chronic health states [7], because it is thought that using an anchor health state may improve the sensitivity of TTO when trying to detect differences in utility between health states that are similar. In this review, we focus only on those studies that use the chained TTO method to value temporary health states.

**Search Strategy**

Established methodological approaches for undertaking systematic reviews in health care were followed throughout the review process [8,9]. A literature search was conducted of published studies from the earliest possible date up to July 2016 in the major databases: Ovid MEDLINE(R), Embase, PsycINFO, and EBM Reviews. Initially, titles, abstracts, and keywords were searched using only the terms listed in column A of Table 1 to identify all TTO studies, the purpose of which was to create a database of TTO studies for future work, with the added benefit that additional search terms could be added with ease at a later date if required for different study purposes. The results were then imported into reference management software EndNote version X7 [10] and duplicates were subsequently removed. After this step, titles, abstracts, and keywords were searched again in the EndNote database of TTO studies using the terms listed in column B of Table 1 to identify chained TTO studies. The eligibility of the identified records for inclusion in the review was then determined according to the following inclusion and exclusion criteria.

**Inclusion criteria:**

1. Studies published in English;
2. Chained TTO used to value temporary states only;
3. Primary data collected.

**Exclusion criteria:**

1. Review articles;
2. Non–health-related studies.

Abstracts were screened on the basis of the aforementioned criteria, and full-text screening was conducted if it was not possible to assess eligibility from the abstract alone. The screening was conducted independently by two reviewers and compared for consistency, with a third reviewer resolving any disagreements.

**Results**

**Literature Search Results**

After the initial search using terms from column A of Table 1, 3715 studies were identified, and 2025 studies remained after removing duplicates. The subsequent search using terms from column B of Table 1 then resulted in the identification of 285 potential studies. After screening the abstracts, 169 studies were excluded, leaving 116 studies for full-text examination, which resulted in 8 studies being identified for the review. Subsequent reference and citation searches of the eight studies resulted in an additional relevant study being identified—this particular article [11] did not come up in our search because it does not mention “time trade-off” in the title, abstract, or keywords. Another five studies that used chained TTO to value chronic health states rather than temporary health states were identified but not included in the review. Figure 1 illustrates this search process using a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram [9].

The following characteristics were extracted from each study: lead author, title, publishing journal, clinical area, study location, methods used, sample characteristics, information about the health states, methodological issues raised, comparisons with other methods (if applicable), and the key conclusions. In addition, the chained TTO values were typically analyzed by calculating mean [11–15] or median [16–18] or both [19] the values generated. We were unable to prospectively assess the study quality because of the lack of established criteria for assessing such studies. Nevertheless, given the few studies identified, we
were able to analyze and assess each study on its individual merit. The key characteristics of the chained TTO studies that value temporary health states are presented in Table 2.

**Study Context**

Breast cancer was the most prevalent disease area used in the studies identified (four of nine studies), and three of these were from the same lead author. The study objectives varied from solely methodological purposes to valuing temporary health states in an economic evaluation. All three studies by Jansen et al. [12,13,16] used chained TTO to value health states related to treatments for breast cancer and compared the results with chained standard gamble (SG). The first study [12] explored the feasibility of chained TTO. The second [13] explored how preferences for radiotherapy may change after patients experience postoperative radiotherapy. The third [16] explored how preferences for chemotherapy may change after the patients underwent adjuvant chemotherapy. Johnston et al. [19] used chained TTO to evaluate various scenarios related to the results of breast cancer screening, in which “false-positive” and “true-negative” screening results were considered to be temporary. Cook et al. [11] valued health states related to postprocedural symptoms of two gallstone disease treatments for use in a CUA and explored the effects of using different time horizons for the anchor health state. Locadia et al. [17] valued health states related to the possible outcomes after an episode of venous thromboembolism and compared the results of the chained TTO with those of conventional TTO and visual analogue scale (VAS). Gu et al. [15] used chained TTO to elicit hip replacement preferences from outpatients and surgeons. McNamee [14] used chained TTO to derive QALY gains from various temporary health profiles relating to the treatment of multiple sclerosis and compared those with QALY gains from selected EuroQol five-dimensional questionnaire (EQ-5D) health profiles using general population tariff values. Finally, Spencer [18] explored the impact of using anchor states in chained procedures (chained TTO and chained SG) in two types of chaining: states chained to death (as in the chained TTO method described in this article) and states chained to normal health (the method typically used when chained TTO is applied in chronic health states). EQ-5D health profiles were used. Sample sizes were typically low, between 30 and 96, with the exception of the Johnston et al. [19] study, which had a sample size of 440.

**The Chained TTO Task**

None of the included studies reported the average length of time required for the chained TTO task. However, one study that used chained TTO to value chronic health states (as opposed to temporary health states) had a mean interview duration of 85 minutes [20]. The number of temporary health states valued in the chained TTO exercise varied from 1 [19] to 6 [11]. There were different approaches to the choice of the anchor state. Four
<table>
<thead>
<tr>
<th>Study</th>
<th>Clinical area</th>
<th>Sample details</th>
<th>HS descriptions</th>
<th>No. of temporary HS valued (in stage 1)</th>
<th>Time horizon for temporary HS (in stage 1)</th>
<th>Anchor HS used (in stage 2)</th>
<th>Time horizon for anchor HS (in stage 2)</th>
<th>Other methods used in the study to value temporary HS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cook et al. [11]</td>
<td>Gallstone disease</td>
<td>General population (n = 96), United Kingdom</td>
<td>Vignettes</td>
<td>6</td>
<td>12 wk</td>
<td>Worst HS (by respondent ranking)</td>
<td>12 mo*</td>
<td>VAS</td>
</tr>
<tr>
<td>Jansen et al. [12]</td>
<td>Breast cancer</td>
<td>Patients (n = 68), The Netherlands</td>
<td>Vignettes</td>
<td>3</td>
<td>6 mo</td>
<td>Hospitalization</td>
<td>3 mo</td>
<td>Chained SG</td>
</tr>
<tr>
<td>Johnston et al. [19]</td>
<td>Breast cancer</td>
<td>Patients (n = 440), United Kingdom</td>
<td>Vignettes</td>
<td>1</td>
<td>1 y</td>
<td>Worst HS (by respondent ranking)</td>
<td>1 y</td>
<td>VAS</td>
</tr>
<tr>
<td>Jansen et al. [13]</td>
<td>Breast cancer</td>
<td>Patients (n = 55), The Netherlands</td>
<td>Vignettes</td>
<td>3</td>
<td>6 mo</td>
<td>Hospitalization</td>
<td>3 mo</td>
<td>VAS, chained SG</td>
</tr>
<tr>
<td>Jansen et al. [16]</td>
<td>Breast cancer</td>
<td>Patients (n = 94), The Netherlands</td>
<td>Vignettes</td>
<td>2</td>
<td>6 mo</td>
<td>Hospitalization</td>
<td>6 mo</td>
<td>VAS, chained SG</td>
</tr>
<tr>
<td>Locadia et al. [17]</td>
<td>Venous thromboembolism</td>
<td>Patients (n = 54), The Netherlands</td>
<td>Vignettes</td>
<td>5</td>
<td>4 wk/3 mo†</td>
<td>Hospitalization</td>
<td>4 wk and 3 mo†</td>
<td>VAS, conventional TTO</td>
</tr>
<tr>
<td>Spencer [18]</td>
<td>General health</td>
<td>General population (n = 30), United Kingdom</td>
<td>EQ-5D-3L states (21211 and 21222)</td>
<td>2</td>
<td>10 y</td>
<td>EQ-5D-3L states (21222 and 22232, respectively)</td>
<td>10 y</td>
<td>Conventional TTO, conventional SG, chained SG</td>
</tr>
<tr>
<td>McNamee [14]</td>
<td>Multiple sclerosis</td>
<td>Patients (n = 50), United Kingdom</td>
<td>Vignettes</td>
<td>2</td>
<td>3 y</td>
<td>Worst HS (by respondent ranking)</td>
<td>3 y</td>
<td>Conventional TTO</td>
</tr>
<tr>
<td>Gu et al. [15]</td>
<td>Hip replacement</td>
<td>Patients (n = 50) and health care providers (n = 16), United States</td>
<td>Vignettes</td>
<td>4</td>
<td>6 wk/3 mo/4 mo†</td>
<td>Constant severe pain</td>
<td>15 y</td>
<td>None</td>
</tr>
</tbody>
</table>

EQ-5D-3L EuroQol five-dimensional questionnaire; HS, health state; SG, standard gamble; TTO, time trade-off; VAS, visual analogue scale.

* The HS most commonly ranked as the worst was also valued with a 12-y time horizon in stage 2, for comparison. Results were not sensitive to the time horizon and a time of 12 mo was used to calculate the utilities.

† One HS, “treatment with oral anticoagulants,” had a time horizon of 3 mo in stage 1.

‡ Various time horizons were used as appropriate.
studies used a ‘hospitalization after a serious accident’ scenario as an anchor state [12,13,16,17]. Jansen et al. [12] found that 34% (n = 24) of their respondents preferred the anchor health state to one of their temporary health states, which meant that the utility for this state could not be calculated for these respondents. Similar issues were reported by Locadia et al. [17]. Three studies avoided this issue by not selecting a specific anchor state before the interviews. Instead, the studies asked respondents to rank the health states before the chained TTO task and used the worst health state as decided by the respondent as the anchor state [11,14,19]. One study [18] preselected the anchor states but avoided this issue because three-level EQ-5D states were used, which, using existing utility tariffs [21], can be ranked with reasonable confidence without respondent input.

The time horizon used varied from 4 weeks [17] to 1 year [19], with the exception of the Spencer [18] and McNamee [14] studies, which used longer time horizons. These are considered exceptions because of the wide range of methods used by Spencer [18] and the fact that McNamee [14] used health profiles that consisted of a series of consecutive health states across a given period of time rather than individual health states. In three of the studies [11–13], the time horizons used when valuing the anchor health states were not consistent with the time horizons used when valuing the temporary health states. Locadia et al. [17] used two time horizons to value their anchor states. They found no difference in the mean and median utilities for their anchor health state when using the two time horizons, and as a result the utilities of the temporary health states would not differ depending on the utility of the anchor state that was used in their calculation. Cook et al. [11] decided that they would not use the same time horizon in both stages, given that the stage 1 time horizon of 12 weeks would present a scenario in which the respondent faced an impending death in stage 2. They found no statistically significant difference (at the 1% level) in the utility for their most common anchor state when using 12 months and 12 years as time horizons in stage 2.

The consistency of chained TTO was examined in two studies by comparing the rank order of the health states implied by the average utilities elicited using chained TTO with rank orders generated using other valuation methods such as SG and VAS [12,17]. No studies reported significant concerns with the consistency of chained TTO. The chained TTO method was also found to be responsive—it was shown to be able to detect changes between different health states. McNamee [14] found that chained TTO was more able to account for differences in preferences than conventional TTO when used to value health profiles consisting of multiple states. Jansen et al. [16] reported that chained TTO found a significant difference between a health state experienced by the respondent and its hypothetical description, whereas this was not detected by VAS. Similarly, by applying chained TTO, Jansen et al. [13] found that utilities were significantly higher once a health state had been experienced, whereas chained SG and VAS could not detect significant changes. These comparisons suggest that chained TTO is a responsive preference elicitation method.

All the analyzed studies show that chained TTO is a feasible method in eliciting short-term utilities. Specifically, Jansen et al. [12] concluded that the method seemed feasible, given that only 2 respondents (out of 70) stopped their interviews and that there were only 5% missing answers.

Last, chained TTO was also found in studies [7,18,22] that valued chronic health states. In addition, there were examples of a slightly different variant of chained TTO being used to value chronic [20] and temporary health states [23]. The “iterative” method used in these studies differed from the method outlined earlier in the second section in that a series of health states were valued relative to one another (creating a “chain”) rather than all states being valued relative to the same anchor state. The study that valued temporary health states using this method [23] was excluded for consistency, but had it been included in this review, little would have changed. Much like the included studies, the sample size was low (n = 28), the health states were not from a generic descriptive system, and few temporary states were valued (n = 3).

**Comparisons with Other Utility Elicitation Methods**

Conventional TTO featured in three of the reviewed studies, but comparisons between these methods were made in only two studies. Locadia et al. [17] found that chained TTO utilities were larger than those elicited using conventional TTO, whereas Spencer [18] concluded that there were no significant differences between utilities derived from the two methods. Four studies included both a chained TTO and a chained SG [12,13,16,18], Jansen et al. [12] found that mean utilities from chained TTO and chained SG were almost identical, with the exception of the anchor health state, which was found to have a lower utility in TTO relative to the SG. As the anchor state was valued using conventional methods (i.e., stage 2 of the chained procedure), the authors concluded that this finding disagreed with results from comparison studies of conventional TTO and SG. In both the studies by Jansen et al. [13,16], the main objective was related to the change in utility after personal experience of a health state, and hence comparisons between chained TTO and chained SG were not made explicitly. Finally, although Spencer [18] included both chained TTO and chained SG, the focus was on how the biases that typically affected the conventional variants might be altered when an anchor state was introduced. Much like the findings with regard to biases in the conventional variants [24], it was concluded that there were counteracting issues within the chained TTO method that might make it preferable to chained SG, although it was suggested that the results should be interpreted with caution.

**Discussion**

**Summary of Main Findings**

This article systematically reviewed chained TTO studies that valued temporary health states. One of the major findings was that there were very few published chained TTO studies (n = 9) in the literature. The applications of chained TTO were diverse: from being used in a CUA [11] to methodological articles that explored the impact of using anchor states [18]. Chained TTO studies typically had a small sample size, which could be explained by the fact that it has a relatively complex format compared with conventional TTO. Furthermore, the data collection format (interviewer-led face-to-face interviews in all studies) and the source of participants (patients rather than the general public in seven of nine studies) could also be the reason for the low sample sizes. Nonetheless, the general conclusion is that chained TTO appears to be a feasible, consistent, and responsive method. Nevertheless, although the method appears to have been successfully applied to a number of therapeutic areas with various objectives, several methodological issues remain with regard to its use.

**Methodological Issues**

**Anchor states**

The selection of a suitable anchor state is vital for chained TTO to be applied successfully. For respondents to be willing to trade off time in the anchor health state in stage 1, this state must be considered worse than the temporary health state being valued.
Furthermore, the anchor state must also be considered better than dead to avoid the complication of dealing with negative utility values in stage 2. Hence, chained TTO studies require substantial piloting to ensure that the chosen anchor state is suitable. The most common anchor state in the identified studies, ‘hospitalization after a serious accident,’ was justified by Jansen et al. [12] on the grounds that although this is not likely to be something that has been experienced by the respondents, it is an imaginable scenario. An alternative approach is for respondents to rank all the health states before the valuation task, and the health state considered to be the worst can be used as the anchor health state, as done in several studies [11,14,19]. A potential issue with this approach is that within a single study, more than one anchor health state could be valued in stage 2 (across respondents), which may raise concerns regarding the comparability of utilities across respondents.

In situations in which the anchor state is preferred to a temporary health state being valued, Locadia et al. [17] concluded that this may be a source of upward bias, which may be a concern. Jansen et al. [16] suggested that the length of time spent in the anchor state could be increased beyond the initial time horizon \( T \) to prevent this issue in stage 1 of the chained TTO task. In other words, it would be possible for the indifference point \( X_1 \) to occur when the time spent in the anchor state is longer than the fixed time horizon \( T \) in the temporary state. This is possible because a value of \( X_1/T \) that is greater than 1 in stage 1 simply denotes that the anchor state is preferred to the temporary health state (i.e., it is not better than perfect health). Although unlikely, this has the potential to create negative utilities for health states once the two stages are complete and it could be argued that it would simply be wiser to take the approach of using the lowest ranked health state as the anchor state.

Time horizons

The ability to use a short time horizon (at least in stage 1) is one of the benefits of the chained TTO method, and naturally the time horizon selected will depend directly on the health context of the study. Nevertheless, questions have arisen surrounding the time horizon in stage 2. The studies by Locadia et al. [17] and Cook et al. [11] found no differences in utilities when they used multiple time horizons to value the anchor state in stage 2, providing some support for the assumption of constant proportional TTO. The other studies, however, did not test the use of different time horizons in stage 2, despite the fact that short time horizons provide respondents with a scenario that includes imminent death.

Although selecting an appropriate time horizon is a concern when using the chained TTO method, Jansen et al. [13] argued that, on the whole, this is an advantage given that the assumption of constant proportional TTO is required only for the anchor health state. Furthermore, Jansen et al. [12] discussed how TTO calculations assumed linear utility and risk neutrality (zero discounting) for the life duration. They stated that this leads to underestimation of utilities using conventional TTO, and that this effect is less pronounced when using chained TTO because of the shorter time horizons. As a result, they concluded that chained TTO might be less biased than the conventional method. The choice of time horizon in stage 2 is an important area for further research.

Other issues

A limitation of the chained TTO procedure is that it does not incorporate health states that are considered to be worse than dead. According to Locadia et al. [17], this may be a source of upward bias in chained TTO. This problem, however, would also arise when using the conventional TTO method without prior consideration of health states that are considered worse than dead. It is also possible to use a variant of the conventional TTO designed for health states that are considered worse than dead, should this be the case.

In addition, the chained TTO method is more complex than the conventional TTO method because of the two-stage procedure, and Locadia et al. [17] suggested that it might be more prone to error because of this. Although this is likely to be true given that the nature of the task changes during the interview, no studies reported a significant burden among respondents when conducting chained TTO. Furthermore, difficulty with the task could be minimized by offering practice tasks before each stage and by ensuring that the number of health states valued is not too large. The latter can be determined through pretesting and piloting, which is of clear importance in any preference elicitation study.

Study Limitations

Our review is not without limitations. First, given the search strategy, only articles with a variant of ‘time trade-off’ in the title, abstract, or keywords could be identified. We tried to overcome this limitation by searching for relevant references in reference and citation lists of the identified articles. Nonetheless, it is still possible that some relevant studies could have been missed with this approach. We, however, believe that if TTO was not considered important enough to include within these fields, it is unlikely that the methodology used would have been reported in sufficient detail for this review. Second, the number of studies identified is low \( n = 9 \) and three of them are by the same author. This means that although we have attempted to assess the application of the chained TTO method, our results are based on very few existing applications of the methodology. This, however, also demonstrates the lack of published evidence in the application of the chained TTO technique and supports our case for further studies in the field. Last, it was not possible to assess the quality of the studies identified because of a lack of established criteria for evaluating such studies.

Conclusions

Overall, chained TTO appears to be a feasible, consistent, and responsive method for valuing temporary health states. It may be a suitable alternative to conventional methods when health states are considered temporary, and its application allows the valuation of temporary health states that would improve the efficiency and accuracy of decision making. Nevertheless, the evidence base is limited. Further research is needed to examine the performance and validity of chained TTO compared with conventional TTO in the valuation of temporary health states.

Acknowledgments

We thank two anonymous reviewers and the participants at the Health Economists’ Study Group conference held in June 2016 in Gran Canaria, Spain, for their comments and suggestions.

Source of financial support: The development of this work was funded by the Health Foundation for an internship project.

REFERENCES


