Treatment with vitamin K antagonists: patients' quality of life, valuations and adherence
Locadia, M.

Citation for published version (APA):
Chapter 7

A comparison of three valuation methods for temporary health states in patients treated with oral anticoagulants

M. Locadia, P.F.M. Stalmeier, F.J. Oort, M.H. Prins, M.A.G. Sprangers & P.M.M. Bossuyt

Accepted for publication in Medical Decision Making
Abstract

Objectives
To compare a chained time trade-off (TTO) method with a conventional TTO method in the valuation of temporary health states, in terms of consistency and reliability and to compare both TTO methods with direct rating.

Methods
Eighty-four patients treated with oral anticoagulants were interviewed twice. During the first interview values for five temporary health states were obtained with a rank ordering procedure, direct rating and the chained TTO method. During the second interview either the first interview was repeated (N=30), or health state values were obtained with the conventional TTO method (N=54). Consistency was assessed by comparing the three valuation methods with the rank ordering procedure. Generalisability theory was used to assess reliability.

Results
The three methods produced significantly different valuations of health states. Chained TTO values were higher than values obtained with direct rating and the conventional TTO. Consistency and reliability did not differ across the three methods.

Conclusion
We found no evidence for a difference in consistency and reliability between the chained TTO method and the conventional TTO method in the valuation of temporary health states. As direct rating is simpler to administer than both TTO methods, one could consider using direct ratings for the valuation of temporary health states. TTO values are often preferred over direct ratings for the use in formal decision analyses. We therefore discuss biases associated with the conventional and the chained TTO method.
Introduction

Patient valuations of health states are used in decision analyses and cost-utility analyses. So far, methods for health state valuation have focused mainly on the valuation of chronic outcomes of care, with less attention to the evaluation of temporary health states, such as temporary side-effects of treatment and the effects of being tested, screened or treated.

The time trade-off (TTO) is a commonly applied valuation method for chronic health states. In the TTO, the patient is asked to choose between a fixed duration of time in imperfect health and a variable amount of time in perfect health; both health states are followed by death. In contrast, temporary health states are, by definition, followed by another health state. A straightforward application of the TTO to temporary health states may lead to less valid results because an unrealistic scenario is presented to patients. Furthermore, the terminal status of the temporary health state, caused by subsequent death, may influence the evaluation of the health state [1]. For these reasons adaptations of the TTO have been proposed as alternative methods to value temporary health states.

Adaptations of the TTO for the valuation of temporary health states include the sleep trade-off [2,3] and the wait trade-off [4]. The sleep trade-off assesses the amount of time a patient is willing to sleep to avoid a temporary health state. Values are scaled from the temporary sleep state (0) to current or perfect health (1). The sleep trade-off resembles the chained TTO but does not convert the temporary health state values to a scale from death (0) to perfect health (1). We therefore did not use the sleep trade-off in this study. In the wait trade-off patients are asked to trade off an additional amount of time with current symptoms to avoid the short-term side-effects of a diagnostic test. Since the wait trade-off was designed to assess values for temporary health states associated with diagnostic testing, it was not appropriate for the valuation of health states in our study.

Torrance [5] has suggested a chained TTO procedure for the valuation of temporary health states. With this method the temporary health state is not directly compared to perfect health and subsequent death. Instead, an anchor health state is used. The chained TTO consists of two steps. In the first step, the temporary health state is compared to the anchor health state and perfect health. In the second step, the anchor health state is compared to death and perfect health. The valuation of the anchor health state in the second step is used to convert the value of the temporary health state to a scale ranging from death (0) to perfect health (1).
chained TTO has been used in a number of studies to assess values for temporary health states [6-8]. Jansen et al [7] concluded that the feasibility and consistency of the chained TTO for the assessment of temporary health states were satisfactory.

The chained TTO has also been used to assess values for chronic health states [9-11]. In these studies the best end point (perfect health) was replaced by another health state. Values assessed with the chained TTO were lower than values assessed with the conventional TTO. A possible explanation for this finding is that respondents anchor their responses on a master health scale with the endpoints death and perfect health, and then adjust their responses for the specific end-points used [9,11]. This phenomenon is known as 'anchoring and adjustment'.

Although the chained TTO is more realistic for the measurement of temporary health states than the conventional TTO it is also more complex. The chained TTO procedure involves an additional step which might introduce extra noise. In addition, patients are presented with an anchor health state in the chained TTO, which is often irrelevant for the situation of interest. The chained TTO may therefore be more prone to error than the conventional TTO. The conventional TTO has been used to assess values for temporary health states [12], but to our knowledge values for temporary health states obtained with the conventional TTO have never been related to values for temporary health states obtained with the chained TTO. We therefore compared the chained TTO with the conventional TTO for the valuation of temporary health states, in terms of consistency and reliability. We also compared both TTO methods with a more simple method: direct rating. We assessed patient values for health states associated with venous thromboembolism and its treatment with oral anticoagulants.

Methods

Patients

Three groups of eligible patients were recruited between October 2000 and June 2002 in three hospitals in the Netherlands: the Academic Medical Centre in Amsterdam, the University Hospital of Groningen and the University Hospital of Maastricht. The first group consisted of patients with a first or second episode of venous thromboembolism, for whom treatment with oral anticoagulants had been started. The second group consisted of patients who had experienced an episode of major bleeding during oral anticoagulant treatment in the previous year. The third group encompassed patients with a post-thrombotic syndrome, diagnosed at
least one year after an episode of deep-vein thrombosis, who had been treated with oral anticoagulants for at least three months. Excluded were patients under 18, patients with cancer or other dominant co-morbidity, as well as patients with an insufficient command of the Dutch language.

Patients were approached for study entry by their attending physician or a research nurse, who explained the purpose of the study and provided written information. After written informed consent had been obtained, patients were invited for an interview. Since we wanted patients to have experienced treatment with oral anticoagulants at the time of the interview, patients with an episode of venous thromboembolism were interviewed approximately three months after inclusion. Patients with a major bleeding event occurring during treatment with oral anticoagulants, and patients with a post-thrombotic syndrome were contacted a few days after inclusion, to schedule an interview appointment. All patients were interviewed by one of four trained interviewers. For quality control, each interviewer audio-taped three of the first ten interviews and received detailed feedback by one of two investigators (ML, PS). The present study is part of a larger project to determine the optimal duration of treatment with oral anticoagulants, based on a decision model that uses the available evidence and patients’ preferences.

Health states
Eight scenarios with possible outcomes after an episode of venous thromboembolism were developed, based on literature and the experience of medical specialists. Each scenario described an outcome in terms of a physical, psychological, and social dimension. All scenarios were labelled and written in the third person. Separate versions were used for men and women. Of the eight health states, five were temporary: ‘deep-vein thrombosis’, ‘pulmonary embolism’, ‘treatment with oral anticoagulants’, ‘gastrointestinal bleeding’ and ‘muscular bleeding’. An additional health state ‘hospitalisation after a serious accident’ was used as an anchor state in the chained TTO method. In the present study, only data on the five temporary health states are presented. As an example health state descriptions for ‘deep-vein thrombosis’ and ‘hospitalisation after a serious accident’ are provided in appendix A.

Interview
Rank ordering procedure. Health states were randomly presented to the patients by shuffling the cards with health state descriptions. Patients were first asked to rank order the health states from least to most preferred. Next, they were asked to
position perfect health and death in the previously obtained ranking.

**Direct rating.** Patients rated the health states on the vertical Euroqol thermometer with the endpoints death (0) and perfect health (100) [13].

**Conventional TTO.** In the conventional TTO, patients were asked to choose between a fixed duration of time in the temporary state and a variable amount of time in perfect health using a series of choices. Both health states were to be followed by death. The fixed duration of the temporary health states was four weeks, with one exception: for the assessment of treatment with oral anticoagulants, the duration was fixed at three months. In the first two choices, the duration of time in perfect health was set equal to 0 or the fixed duration, in random order. The bisection method was used to vary the time spent in perfect health until the patient reported indifference [14].

**Chained TTO.** The chained TTO consists of two steps. In the first step patients were asked to choose between the fixed duration of time in the temporary health state and a variable amount of time in the anchor health state. In contrast with the conventional TTO, both states were followed by perfect health for the rest of one’s life. In the first two choices, the duration of time in the anchor health state was initially set equal to 0 and to the fixed duration, respectively, in random order. The bisection method was then used to vary the time spent in the anchor health state until the patient reported indifference. In the second step the anchor health state was compared to death and perfect health. The anchor health state was valued twice: the duration of the anchor health state was fixed at four weeks and three months in random order. The choice series described in the conventional TTO were used to determine the patient’s indifference point. Values of the temporary health states were converted to a scale from death (0) to perfect health (1) using the following transformation (see Jansen et al. [7] for a figure clarifying formula 1):

\[ h_q = 1 - (1 - h_a) \cdot x/t \]

where \( h_q \) equals the value of the temporary health state; \( h_a \) equals the value of the anchor health state obtained in the second step; \( x \) equals the period of time for which the patient is indifferent between the temporary health state and the anchor health state in the first step; and \( t \) equals the fixed duration of the temporary health state.
Study procedure

During the first interview health state values for temporary health states were obtained with the rank ordering procedure, direct rating and the chained TTO. At the end of the interview, patients were asked if they were willing to participate in a second interview. The second interview took place approximately six weeks after the first interview. At the University Hospital of Groningen, we repeated the first interview. At the Academic Medical Centre and the University Hospital of Maastricht the temporary health states were evaluated with the conventional TTO.

Analyses

All health state values were transformed to a 0 to 100 scale. To test if health state values differed significantly between the three methods, analyses of variance were performed, specifying the ‘methods’ and ‘health states’ factors as ‘within subject factors’. To test whether the period following a temporary health state influenced the evaluation of that state, we compared values obtained with the conventional TTO with values obtained with the first step of the chained TTO by means of an analysis of variance. Again, the ‘methods’ and ‘health states’ factors were specified as ‘within subject factors’. To judge consistency, the three methods were compared with the simple rank ordering procedure. Across patients, consistency was assessed by calculating Spearman rank correlation coefficients between the median overall ranking of the temporary health states on the one hand and the median values obtained with direct rating, the conventional TTO and the chained TTO on the other. Within individuals, consistency was assessed by calculating Spearman rank correlation coefficients between the rankings and values obtained with direct rating, the conventional TTO and the chained TTO. Wilcoxon signed rank tests were used to test for differences between the median correlation coefficients.

To assess the reliability of the different health state valuation methods, Generalisability Theory was used [15]. Generalisability Theory expresses reliability coefficients as functions of variance components. Variance components represent the contributions of different sources of variance to the health state valuations. In the present study patients, health states, and time of measurement were the source of variance, known as facets in Generalisability Theory. Reliability was assessed in two ways: by the inter-rater method and the test-retest method.

Inter-rater method. The variance components for the facets patients, health states and their interaction were estimated for direct rating, the conventional TTO and the chained TTO using the restricted maximum likelihood method [16]. Next, inter-
rater reliability and inter-rater agreement were calculated for each valuation method from these variance components (appendix B). Inter-rater reliability expresses the extent to which two arbitrary patients agree on the ranking of health states, whereas inter-rater agreement expresses the extent to which two arbitrary patients agree on the absolute values attributed to the health states. Both coefficients vary between 0 (no agreement) and 1 (perfect agreement).

Test-retest method. For direct rating and the chained TTO, the variance components for the facets patients, health states, time of measurement and their interactions were estimated using the restricted maximum likelihood method [16]. Next, test-retest reliability and test-retest agreement were calculated on the basis of the respective variance components (appendix B). Test-retest reliability expresses the agreement between the rank orderings of an arbitrary patient’s valuations of health states at two points in time, whereas test-retest agreement expresses the extent of agreement between the absolute values attributed to the health states at two points in time. Both coefficients vary between 0 (no agreement) and 1 (perfect agreement).

To assess whether the direct ratings and time trade-off values remained stable over time, an analysis of variance was used, with the ‘methods’ and ‘health states’ factors specified as ‘within subject factors’.

Results

Patients

Of the 159 eligible patients, 129 (81%) consented to participate. Five patients died before the first interview could take place. The data of the first three patients interviewed were excluded because the anchor health state was adapted after the first interviews had taken place. Thus, rank ordering, direct rating and chained TTO data were available for 121 patients. Of these patients, 84 (69%) agreed to participate in the second interview. The second interview varied across treatment centre: in 54 patients values were assessed with the conventional time trade off and in 30 patients the first interview was repeated. Thus, 54 patients completed both the direct ratings, the conventional TTO and the chained TTO and in 30 patients the direct ratings and the chained TTO were repeated. Patient characteristics are shown in Table 1. Patients who had experienced a major bleeding event were significantly older than patients who had experienced an episode of venous thromboembolism or patients with the post-thrombotic syndrome (p<0.05). The three groups of patients did not differ significantly in terms of sex, marital status or educational level.
Table 1 Patient characteristics (N=54)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex</td>
<td>23 (43%)</td>
</tr>
<tr>
<td>Mean age (±SD), years</td>
<td>53 (±16.6)</td>
</tr>
<tr>
<td>Married</td>
<td>26 (51%)</td>
</tr>
<tr>
<td>Educational level*</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>16 (32%)</td>
</tr>
<tr>
<td>Medium</td>
<td>22 (44%)</td>
</tr>
<tr>
<td>High</td>
<td>12 (24%)</td>
</tr>
<tr>
<td>Clinical background</td>
<td></td>
</tr>
<tr>
<td>Venous thromboembolism</td>
<td>23 (43%)</td>
</tr>
<tr>
<td>Post-thrombotic syndrome</td>
<td>21 (39%)</td>
</tr>
<tr>
<td>Major bleeding</td>
<td>10 (19%)</td>
</tr>
</tbody>
</table>

*Missing in 3 patients. †Missing in 4 patients.

Health state values

Median health state values obtained with the three methods are shown in Figure 1. Health state values did not differ significantly between the three groups of patients. The analyses of variance indicated that health state values differed significantly between the three methods (p<0.001). Chained TTO values were higher than direct rating and conventional TTO values. For the three most severe health states, values obtained with conventional TTO were lower than direct rating values, whereas conventional TTO values for the other health states were higher than direct rating values.

![Health state values graph](image)

**Figure 1** Median values for the health states as assessed by the three methods. N = 54. TTO = time trade-off. 1: pulmonary embolism, 2: gastro-intestinal bleeding, 3: muscular bleeding, 4: deep-vein thrombosis, 5: treatment with oral anticoagulants.
values. For both four weeks and three months in the anchor health state the median value was 0. The mean value was 12 for both periods of time in the anchor state.

To test whether the period following a temporary health state influenced the evaluation of that state we compared values obtained with the conventional TTO with values obtained with the first step of the chained TTO. Median values are shown in Figure 2. The analysis of variance indicated that values obtained with the conventional TTO were significantly lower than values obtained with the first step of the chained TTO (p<0.001).

**Figure 2** Median values for the health states as assessed by the conventional TTO and the first step of the chained TTO. N=54. TTO = time trade-off. 1: pulmonary embolism, 2: gastrointestinal bleeding, 3: muscular bleeding, 4: deep-vein thrombosis, 5: treatment with oral anticoagulants.

**Consistency**

Across individuals, Spearman rank correlation coefficients between the median rankings and the median values obtained with direct rating, the conventional TTO and the chained TTO were 1.00, 1.00, and 0.98, respectively. Within individuals, Spearman rank correlation coefficients could not be computed whenever the five health states were valued equally with one health state valuation method. None of the patients valued the five health states equally with direct rating. With the conventional TTO health state values were identical in 7 out of 54 patients (13%); 5 patients valued all health states at 0, and 2 patients valued all health states at 100. With the chained TTO health state values were identical in 5 out of 54 patients.
These 5 patients valued all health states at 100. Thus, Spearman rank correlation coefficients between the rankings and the conventional TTO could not be computed in seven patients. In addition, Spearman rank correlation coefficients between the rankings and the chained TTO could not be computed in five patients. Median Spearman rank correlation coefficients between the rankings and values obtained with direct rating, the conventional TTO and the chained TTO were 0.81, 0.71, and 0.73, respectively. Wilcoxon signed rank tests showed that these coefficients did not differ statistically significantly (p>0.10).

Reliability

The results of the inter-rater method are shown in Table 2. The relative contribution of the variance component ‘health states’ was comparable for the three methods, indicating that the differences between the health states were the same across methods. The relative contribution of the variance component ‘patients’ was highest for the conventional TTO indicating that differences between patients were larger for the conventional TTO than for the other methods. For each valuation method, the largest source of variation was the interaction between ‘health states’ and ‘patients’, indicating that the magnitude of the differences between the health states differed across patients. The inter-rater reliability coefficient, indicating the extent to which two arbitrary patients agree on the ranking of the health states, was 0.57 for direct rating, 0.61 for the conventional TTO and 0.53 for the chained TTO. The inter-rater agreement coefficients, indicating the extent to which two arbitrary patients agree on the absolute values attributed to the health states, were 0.45, 0.44,

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Direct rating</th>
<th>Conventional TTO</th>
<th>Chained TTO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health states</td>
<td>29%</td>
<td>29%</td>
<td>27%</td>
</tr>
<tr>
<td>Patients</td>
<td>26%</td>
<td>35%</td>
<td>25%</td>
</tr>
<tr>
<td>Interaction</td>
<td>45%</td>
<td>36%</td>
<td>48%</td>
</tr>
</tbody>
</table>

Inter-rater reliability 0.57 0.61 0.53
Inter-rater agreement 0.45 0.44 0.42

N=54. TTO = time trade-off.
Note: the square root of the generalisibility coefficients can be interpreted as the correlations between the valuations of two arbitrary patients.
A comparison of three valuation methods

and 0.42 respectively. Note that the square root of these generalisibility coefficients can be interpreted as the correlations between the valuations of two arbitrary patients. The confidence intervals of the variance components overlapped widely, suggesting that the generalisability coefficients did not differ statistically significantly. Thus, the three valuation methods did not differ in terms of inter-rater reliability and inter-rater agreement.

Table 3 Variance component estimates for health states, patients and time, and test-retest reliability and agreement

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Direct rating</th>
<th>Chained TTO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health states</td>
<td>27%</td>
<td>20%</td>
</tr>
<tr>
<td>Patients</td>
<td>14%</td>
<td>7%</td>
</tr>
<tr>
<td>Time</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Interactions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health states x Patients</td>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td>Health states x Time</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Patients x Time</td>
<td>11%</td>
<td>28%</td>
</tr>
<tr>
<td>Health states x Patients x Time</td>
<td>40%</td>
<td>39%</td>
</tr>
<tr>
<td>Test-retest reliability</td>
<td>0.63</td>
<td>0.57</td>
</tr>
<tr>
<td>Test-retest agreement</td>
<td>0.57</td>
<td>0.44</td>
</tr>
</tbody>
</table>

N=30. TTO = time trade-off.
Note: the square root of the generalisibility coefficients can be interpreted as the correlations between the valuations of an arbitrary patient at two points in time.

The results of the test-retest evaluation are shown in Table 3. Variability in health state values due to time of measurement alone was 0% for both methods, indicating that overall, health state values remained stable over time for the chained TTO and direct rating. This finding was confirmed by an analysis of variance; values obtained with direct rating and the chained TTO did not differ statistically significantly at the two points in time. However, individual patients did vary in their valuations of health states as indicated by the interaction between 'patients' and 'time' and the interaction between 'health states' and 'patients' and 'time'. Test-retest reliability, indicating the agreement between the rankings of an arbitrary patient's health state valuations at two points in time, was 0.63 for the direct rating and 0.57 for the chained TTO. Test-retest agreement, indicating the agreement between the absolute values an arbitrary patient attributes to the health states at two points in time, was 0.57 for direct rating and 0.44 for the chained TTO. Note that the square root of the
generalisibility coefficients can be interpreted as the correlations between the valuations of an arbitrary patient at two points in time. Here also, the confidence intervals of the variance components overlapped widely, suggesting that the generalisability coefficients did not differ statistically significantly. Thus, the two valuation methods did not differ in terms of test-retest reliability and test-retest agreement.

Discussion
In this study, the consistency and reliability of direct rating, the conventional TTO and the chained TTO were compared for the valuation of temporary health states. Consistency was assessed by comparing the three valuation methods with the rank ordering procedure. We used generalisability theory to assess the reliability of the different methods. Although direct rating, the conventional TTO method and the chained TTO method produced significantly different valuations of health states, consistency and reliability did not differ between the three valuation methods.

This study has some limitations. Firstly, we evaluated a limited number of health states, and the health state descriptions we used did not evenly cover the continuum between death (0) and perfect health (1). As a consequence, a relatively small percentage of variance could be attributed to the variability in health states compared to other studies assessing the reliability of health state evaluation methods using Generalisability theory [17-21]. Secondly, health state values for the chained and the conventional TTO were collected six weeks apart. This could have biased the results. However, values obtained with the chained TTO remained stable between measurement points, suggesting that the observed differences between the conventional and the chained TTO were not caused by a preference shift over time. Finally, Torrance [5] not only introduced a chained TTO for the assessment of temporary health states but also a chained standard gamble and chained direct rating. Since additional data collection was judged to be too time consuming in view of our original research goals, we did not administer these two valuation methods in this study.

Direct rating is simpler to administer than the conventional or the chained TTO. Since we found no evidence for a difference in consistency and reliability between the three methods, one could consider using direct ratings in case of temporary health states. However, values derived using the TTO method are more closely tied to the theoretical foundation of expected utility theory than values obtained by means of direct rating. TTO values are therefore preferred over direct ratings in
A comparison of three valuation methods

decision analyses and cost-utility analyses [22]. In the next paragraphs, we will discuss the potential for bias in the conventional and the chained TTO.

We tested whether the period following a temporary health state influenced the evaluation of that state, by comparing values obtained with the conventional TTO to values obtained with the first step of the chained TTO. Values obtained with the conventional TTO were significantly lower than values obtained with the first step of the chained TTO. This indicates that conventional TTO values are biased downwards, as a consequence of explicitly mentioning that the temporary health state is followed by death. The downward bias, may be explained by the effect of a 'dying mode' type value system in which quality of remaining life, rather than quantity, is a more salient factor in decision-making [23]. In the second step of the chained TTO the anchor health state is also followed by death. Chained TTO values are therefore biased downwards as well. Since the value of the anchor health state \((h_x, \text{ see formula 1})\) is used to convert the value of the temporary state obtained in the first step \((x/t, \text{ see formula 1})\) to a scale from death to perfect health, the overall downward bias in the chained TTO is probably not as large as in the conventional TTO. The fact that the period following a health state influences the valuation of that health state indicates that the judgement of the anchor health state in the first step of the chained TTO is not the same as the judgement of the anchor health state in the second step of the chained TTO. This can be perceived as a drawback of the chained TTO for temporary health states.

Both the conventional TTO and the chained TTO have an upward bias [7]. For several patients we obtained a value of 0 for one of the temporary health states or the anchor health state. It is possible that that the actual value would have been less than 0 for some of these patients if negative values had been allowed. In the chained TTO, the value of the anchor health state is used to convert values of the temporary health states, obtained in the first step, to a scale from death to perfect health (see Formula 1). Therefore, the upward bias is not as large as in the conventional TTO. Since negative health state values are controversial [24,25], we have not considered them in this study.

The chained TTO has a second source of upward bias. When patients preferred four weeks in the anchor health state to four weeks in the temporary health state, the value of the temporary health state was set equal to the value of the anchor state. For gastro-intestinal bleeding this was the case in 33% of patients, for pulmonary embolism in 28% of patients, for muscular bleeding in 15% of patients,
and for deep-vein thrombosis in 7% of patients. Therefore, the median values for these health states are likely to be too high. Johnston et al. [8] avoided this upward bias by using the worst temporary health state, as ranked by the patient, as the anchor health state. This method has drawbacks, as pointed out by the authors. Jansen et al. [7] suggested a less problematic solution for the upward bias in the chained TTO. Instead of shortening the time in the anchor state, it can be prolonged. The same formula (Formula 1) can then be applied to convert values for the temporary health states to a scale from death to perfect health.

A third source of upward bias of the chained TTO may be anchoring and adjustment; respondents might use an internal scale for health evaluations with the end points death and perfect health [9,11]. In our study the median value of the anchor health state was 0, indicating that patients valued the anchor health state similar to death. This makes ‘anchoring and adjustment’ a less plausible explanation for the observed difference between conventional and chained TTO values.

We did not find any differences between the conventional TTO and the chained TTO in terms of consistency and reliability. We therefore discussed the potential for bias in these two methods. Both the conventional and the chained TTO have a downward bias caused by explicitly mentioning the health state to be evaluated is followed by death. In addition, both methods have an upward bias because we did not allow for negative health state values. However, these biases are probably larger for the conventional TTO than for the chained TTO. The chained TTO has two additional upward biases. The upward bias of the chained TTO that can occur when patients prefer the anchor health state over the temporary health state can be easily handled by prolonging the time in the anchor health state whenever the temporary health state is valued worse than the anchor state. The potential for upward bias in the chained TTO caused by anchoring and adjustment does not seem to apply to our study. When comparing biases associated with the conventional TTO and the chained TTO, biases seem larger for the conventional TTO. In addition, the chained TTO provides more realistic scenarios for valuing temporary health states. We therefore feel that, when selecting a TTO method, the chained TTO is a more suitable method in the evaluation of temporary health states than the conventional TTO.
A comparison of three valuation methods

Acknowledgements

The authors thank Mrs M. Laumann, Mrs. P. Mentink and Mrs. M. Voskuilen for their assistance in collecting the data, the physicians for their support in approaching their patients, and the patients for their contribution to this study.
Chapter 7

References

A comparison of three valuation methods


Appendix A

Health state descriptions for ‘deep-vein thrombosis’ and ‘hospitalisation after a serious accident’ (male version).

Deep-vein thrombosis

*Physical functioning.* The patient is admitted to the hospital for one day. His leg hurts and his calf is swollen. He has difficulties walking short distances. He is not able to walk long distances. The patient is partly dependent upon others for his daily care. After a couple of days the patient recovers.

*Social functioning.* Work, family and leisure activities are limited for approximately four weeks, due to pain in the leg.

*Psychological functioning.* The patient is somewhat anxious due to his condition. He is also somewhat worried about a recurrent deep-vein thrombosis in the future.

Hospitalisation after a serious accident

*Physical functioning.* The patient is admitted to the hospital. He is critically ill. The patient is hardly able to move. He is not able to get out of bed, and fully dependent upon others for his daily care.

*Social functioning.* Work, family and leisure activities are not possible.

*Psychological functioning.* The patient is afraid and depressed due to his condition. The patient worries a lot because his future is unclear.
Appendix B

Cardinet et al. [15] give guidelines for expressing reliability coefficients as functions of variance components (V). Formulas used in the present paper are:

Inter-rater reliability = \[ \frac{V_H}{V_H + \frac{1}{2}V_{HP}} \]

Inter-rater agreement = \[ \frac{V_H}{V_H + \frac{1}{2}V_{HP} + \frac{1}{2}V_P} \]

Test-retest reliability = \[ \frac{V_H + V_{HP}}{V_H + V_{HP} + \frac{1}{2}(V_{HT} + V_{HPT})} \]

Test-retest agreement = \[ \frac{V_H + V_{HP}}{V_H + V_{HP} + \frac{1}{2}(V_{HT} + V_{HPT}) + \frac{1}{2}(V_T + V_{PT})} \]

Where subscripts H, P and T refer to the health states, patients and time of measurement.