# Deriving time discounting correction factors for TTO tariffs

Attema, Arthur and Brouwer, Werner
Erasmus University Rotterdam

02. February 2012

# Deriving Time Discounting Correction Factors for TTO Tariffs[1]

Arthur E. Attema[a] and Werner B.F. Brouwer[b]

[a] (Corresponding author) iBMG/iMTA, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands. E-mail: attema@bmg.eur.nl, --31-10.408.91.29 (O); --31-10.408.90.81 (F)

[b] iBMG/iMTA, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands. E-mail: brouwer@bmg.eur.nl.

February, 2012

ABSTRACT. The Time Tradeoff (TTO) method is a popular method for valuing health state utilities and is frequently used in economic evaluations. However, this method produces utilities that are distorted by several biases. One important bias entails the failure to incorporate time discounting. This paper aims to measure time discounting for health outcomes in a sample representative for the general population. In particular, we estimate TTO scores alongside time discounting in order to derive a set of correction factors that can be employed to correct raw TTO scores for the downward bias caused by time discounting. We find substantial positive correction factors, which are increasing with the severity of the health state. Furthermore, higher discounting is found when using more severe health states in the discounting elicitation task. More research is needed to further develop discount rate elicitation procedures and test their validity, especially in general public samples. Moreover, future research should investigate the correction of TTO score for other biases as well, such as loss aversion, and to develop a criterion to test the external validity of TTO scores.

**Key Words**: Discounting, QALY model, Time Tradeoff, Utility Measurement

[1] This research was made possible through a grant from The Netherlands Organization for Health Research and Development (ZonMW), project number 152002010.

## 1. Introduction

Economic evaluations of health technologies often and ideally express outcomes in terms of quality-adjusted life years (QALYs). In order to find appropriate QALY weights for different health states — a crucial matter obviously in coming to accurate estimates of cost-utility ratios — often the time trade-off (TTO) method is used to elicit preferences for health states (e.g., Dolan, 2000). This can be either done within the context of one particular economic evaluation, or more systematically in order to derive national quality of life 'tariffs', such as those corresponding to the frequently used EuroQol-5D descriptive system (Dolan *et al.*, 1996). The latter strategy was also adopted in the Netherlands, where a national Dutch tariff exists, which is based on the TTO method (Lamers *et al.*, 2006). The accuracy of this national tariff therefore depends to a large extent on the accuracy of the TTO method to correctly elicit preferences for health states. The same holds for many other national tariffs.

In a TTO, individuals need to make a tradeoff between quality of life and duration of life. A typical TTO exercise involves a tradeoff between living in some imperfect health state $\beta$ for 10 years and living in full health for a period less than 10 (say X) years. The amount of time people are willing to sacrifice in order to regain full health then indicates the value of the health state under consideration and can subsequently be used to calculate the QALY weight of that health state. Normally, using the linear QALY model as a theoretical underpinning, this is done by simply dividing X by 10. The linear QALY model underlying this calculation assumes that individuals attach an equal weight to each future year. Under that assumption, the amount of time people are willing to sacrifice in order to regain full health immediately indicates the value of the health state $\beta$ and can subsequently be used to calculate its QALY weight (which equals X/10). If a person for instance is indifferent between 7 years in full health and 10 years in health state β, then β is assumed to have a utility value of 0.7.

However, it is well-known by now that the traditional TTO method and the linear QALY model are not without methodological problems, which can lead to systematic bias in resulting health state valuations (and through that in cost-utility analysis using the outcomes). Using TTO responses in the way described above requires strong assumptions, of which linear utility of life duration is an important one (Bleichrodt, 2002). Linear utility of life duration simply refers to the fact that the conventional (linear) QALY model assumes that each added life year in a certain health condition is of equal value regardless of its timing and to what health stock it is added. This assumption is, however, hard to maintain. People weight future years differently (normally lower) than present ones, importantly due to discounting (while diminishing marginal utility may also play a role). This implies that the utility increase from having the projected tenth year in the TTO is lower than that from the first year. Discounting is problematic in the context of using the responses from a TTO, as the common calculation method does not take into account utility curvature, leading to a downward bias in QALY weights (Bleichrodt, 2002).

Consequently, the QALY-scores elicited by the conventional TTO procedure, also those used in national tariffs, are biased. The influence of this bias (which we will label simply as discounting from this point onwards) can be substantial. A typical respondent having to trade-off future life years in order to regain full health is likely to discount future life years (Stiggelbout *et al.*, 1994; Stalmeier *et al.*, 1996; Wakker and Deneffe, 1996; Martin *et al.*, 2000; Bleichrodt and Pinto, 2005; van der Pol and Roux, 2005; Abellán-Perpinán *et al.*, 2006; Attema *et al.*, forthcoming b). Simply using the number of future life years that individuals are willing to trade-off in calculating QALY weights, thus leads to a misrepresentation of the utility attached to a current imperfect health state, which can have a substantial impact, also on cost-effectiveness outcomes (Attema and Brouwer, 2010). Yet, most current valuations and national tariffs (including the Dutch tariff) are based on this method.

In order to have a better estimate of the true QALY weight of a health state, a correction for utility curvature is required (Attema and Brouwer, 2009). This is especially true for discounting given the way that resulting health state valuations are normally used in economic evaluations, i.e., they are discounted to calculate a net present value of QALYs (e.g., Gravelle *et al.*, 2007). If uncorrected TTO scores are used to calculate QALYs and these are subsequently discounted using some discount rate for health effects, this would amount to double discounting and an underestimation of the utility derived from some health state (MacKeigan *et al.*, 2003).

A number of alternatives exists to measure (and correct TTO scores for) discounting of future life years. These include the Certainty Equivalence (CE) method for a risky setting (Miyamoto and Eraker, 1985), and, for riskless settings, the parametric 'delay of ill health' (DOI) method (Cairns, 1992) and the nonparametric Direct Method (DM) (Attema *et al.*, forthcoming b). However, the elicitation of discounting can be a burdensome task, obviating the need for a toolkit that can be easier implemented. This study has as its first aim to develop such a toolkit by presenting a first attempt for a general correction set for national tariffs that can be used to correct 'ordinary' TTO tariffs for discounting.

It is important to recognize that different biases in the TTO methods work in different directions and, hence, may cancel each other out (Bleichrodt, 2002). In theory, a correction for one of these biases may then deteriorate the accuracy of TTO tariffs in describing health preferences. A second aim of this study is therefore to investigate whether a correction for discounting improves the predictive validity in a ranking task of health profiles that are composed of periods in bad health and full health. To this end, we elicit discounting of future life years in a large representative sample of the Dutch population. We estimate discounting by means of two different methods and explore their impact on TTO scores for health states of different degrees of severity. Moreover, we assess the validity of the corrected TTO scores compared to each other and to uncorrected TTO scores, Standard Gamble (SG) scores and Visual Analogue Scale (VAS) scores by employing a 'ranking of health profiles' task (Bleichrodt and Johannesson, 1997).

The organization of this paper is as follows. We describe the methodological background in Section 2, followed by the design of the experiment in Section 3. The results are presented in Section 4. Finally, Section 5 presents the discussion.

## 2. Method

The general QALY model evaluates chronic health profiles (Q,T) by the function U(Q,T)=V(Q)W(T), with U(Q,T) the total utility of a period T in the chronic health state Q, V(Q) the QALY weight of Q, W(T) the discounted utility of duration T. Assuming this model, the QALY weight V(Q) can be estimated by several methods. Because all methods are prone to different biases, it is not a priori clear which method is best to use to elicit QALY weights. As indicated by Bleichrodt (2002), each bias is expected to work in a particular direction, causing either an overall upward bias (SG) or an ambiguous bias (TTO). This study measures the size of the bias in TTO caused by discounting. Moreover, we test the predictive validity of the SG, TTO, and VAS by using an intertemporal ranking task, and, in particular, investigate whether correcting TTO scores for discounting increases or decreases TTO's predictive validity. The applied methods are described below.

### *2.1. Time tradeoff method*

The TTO method elicits preferences for health states by letting a subject imagine living T more years in an imperfect health state. The subject then has to indicate the number remaining life time $x<T$ in full health such that he is indifferent between living T years in the imperfect health state and living x years in full health. According to the QALY model, the resulting indifference can be evaluated by:

$$V(Q)W(T) = V(FH)W(x). \quad (1)$$

Normalizing V(Q) such that V(FH)=1, leaves us with:

$$V(Q) = W(x)/W(T). \quad (2)$$

Investigators using TTO often assume the linear QALY model, i.e., $W(t)=t/T$, which implies a simplification of Eq. 2 to:

$$V(Q) = x/T. \quad (3)$$

However, since the aforementioned empirical literature suggests this assumption is not valid, we do not make it and first measure the shape of W(T). Furthermore, we estimate a set of correction factors (CFs, such that V(Q)=x/T+CF) to directly correct TTO scores for discounting (Attema and Brouwer, 2010). The CFs are regressed on several explanatory variables, including gender, age, health status, and time horizon, so that different CFs can be applied depending on the specifics of the situation.

In addition to discounting, TTO is subject to distortions caused by loss aversion and scale compatibility (Bleichrodt, 2002). *Loss aversion* occurs if individuals adopt a reference point and consider outcomes as deviations from this reference point. Higher outcomes are seen as gains, and lower outcomes as losses, with losses looming larger than gains, and, hence, receiving more weight than commensurate gains (Kahneman and Tversky, 1979; McNeil *et al.*, 1982; Tversky and Kahneman, 1991; Tversky and Kahneman, 1992; Stalmeier and Bezembinder, 1999; Bleichrodt and Pinto, 2002). The effect of loss aversion on TTO scores depends on the elicitation procedure. The most common procedure is to fix the duration in imperfect health and ask for the number of years in full health that makes the respondent indifferent. In this case, loss aversion will cause an upward bias in TTO scores. If instead the duration in full health is fixed and the duration in imperfect health that is considered equivalent is asked for, a downward bias results (Bleichrodt, 2002).

*Scale compatibility* means that an individual assigns more weight to an attribute the higher its compatibility with the response scale used (Bleichrodt and Pinto, 2002; Bleichrodt, 2002). The response scale in TTO is life duration, so scale compatibility predicts the respondent to give more weight to life duration than to health status. As a consequence, the TTO scores will be inflated (Bleichrodt, 2002).

*2.2. Standard gamble method*

The SG method typically asks a subject to imagine being in an imperfect health state and to consider two alternatives. One is a risky treatment with a probability p that the subject returns to full health and will live for T additional years, and a complementary probability 1-p of immediate death. The other alternative involves the certainty that the current health state will persist for the rest of his life (T years again). The probability p is then varied until the subject is indifferent between these alternatives. Using the QALY model under expected utility and normalization, we get $V(Q)W(T) = pW(T)$, so $V(Q) = p$ and p represents the utility of the considered imperfect health state.

Two major biases distort SG utilities. These biases are consequences of EU often being descriptively falsified (Llewellyn-Thomas *et al.*, 1982; Rutten-van Molken *et al.*, 1995). First, people tend to attach nonlinear decision weights to probabilities instead of handling them linearly (Wakker and Stiggelbout, 1995; Bleichrodt *et al.*, 1999; Stalmeier and Bezembinder, 1999; Bleichrodt and Pinto, 2000; Bleichrodt, 2001). Second, loss aversion causes individuals to have a reference point (e.g., the certain option) and may consider the worst outcome of the gamble a loss, which they give more weight. Both biases are predicted to produce an upward bias on SG utilities (Bleichrodt, 2002).

*2.3. Visual analogue scale*

A third way to elicit V(Q) is by means of the VAS. Such a rating scale simply asks a respondent to put the health state to be valued on a thermometer, mostly scaled between 0 (worst imaginable health state) and 100 (best imaginable health state). The VAS does not involve any duration. Furthermore, the VAS is often seen as a choiceless method with no clear theoretical foundation (Dolan, 2000), although its use has been defended (Parkin and Devlin, 2006). The estimate of V(Q) is directly given by the number provided by the respondent (or a transformation of it, see S.3.4.1).

*2.4. Intertemporal ranking task*

Given the fact that the aforementioned distorting factors work in different directions, they may (partly) cancel each other out. Therefore, correcting for one of them need not necessarily improve the predictive validity of the resulting QALY weights. Therefore, we incorporated an intertemporal ranking task in order to test whether correcting TTO scores for discounting would increase the predictive accuracy of choices among health profiles.

This task was proposed by Bleichrodt and Johannesson (1997) as a means to compare the ability of utilities elicited by different methods (e.g., SG, TTO, or VAS) to predict people's choices among different health profiles.[2] It enables a comparison of the revealed ranking and the predicted ranking according to the elicited health state values. In case of substantial differences between results obtained with different methods, the test might detect whether one method predicts revealed rankings better than others. A reduction in predictive accuracy after correcting for discounting would suggest other factors being at work as well, with influences in the opposite directions.

As long as correction mechanisms for the other biases are lacking, it would, then, not be worthwhile to correct TTO scores for discounting (at least not for the health state and time horizon under consideration). If, on the other hand, a correction for discounting would *increase* the predictive ability, it suggests that the absolute influence of discounting is higher than that of the other biases and, hence, it becomes worthwhile to correct for discounting (according to this standard, at least). Finally, our intertemporal ranking task allowed a replication of the test performed by Bleichrodt and Johannesson (1997), i.e., providing a standard against which to judge the performance of SG, TTO, and VAS.

## 3. Experiment

### *3.1. Subjects*

A total of 520 subjects (version A: 262; version B: 258) representative for the Dutch general population participated in the experiment.

### *3.2. Procedure*

The experiment was conducted by a professional internet sampling company (Survey Sampling International). This company has much experience with internet surveys and a large representative database of subjects. The subjects were rewarded with a monetary amount to be given to a charity fund of their choice.

---

[2] The task has also been applied in follow-up studies (Abellán-Perpinán *et al.*, 2009; Attema *et al.*, forthcoming a).

The experiment started with some questions regarding background characteristics. After that, the subjects had to answer the ranking and VAS tasks. These tasks were asked at the beginning in order to let subjects familiarize with the valuation tasks. The survey continued with a random draw of either the SG or the TTO task. Finally, discounting was elicited by means of the DM, and the experiment ended with the DOI task.

Indifferences were elicited by using sequences of binary choices, because indifference by choice tends to cause fewer inconsistencies than indifference by matching (Bostic *et al.*, 1990; Hey *et al.*, 2009; Attema and Brouwer, 2012b). We used a bisection procedure that adjusted the value of X upwards or downwards depending on the chosen option. The size of the change was always half the size of the change in the previous question. We randomized the labeling of the options as "A" (left) or "B" (right).

### *3.2.1. Discounting*

Because of time constraints, we could only use two discounting methods. We chose to use the DOI method and the DM for this purpose. The CE method was not used because it heavily relies on the validity of EU, which has been shown to be descriptively flawed (Starmer, 2000; Bleichrodt *et al.*, 2007) and because it uses a risky context, unlike the TTO method. Finally, Attema *et al.* (forthcoming a) found no difference between utility elicited with the DM and utility elicited with the CE after correcting for probability weighting in the CE.

The DOI method aims to elicit intertemporal preferences for non-fatal changes in health. It identifies an indifferent point between two durations of ill-health which occur at two different points in time. This is accomplished by one or more open-ended questions that ask the subject to imagine being ill at some point in the future and offers an opportunity for this spell of ill-health to be delayed due to a one-off treatment. The subject then has to identify a maximum number of days of future ill-health at which it would still be worthwhile to receive this treatment (van der Pol and Cairns, 2001; van der Pol and Cairns, 2008)[3]. Subsequently, one has to specify a parametric discounting function (e.g., the constant discounting function or some kind of hyperbolic function) and estimate its parameter(s) that best fit the subject's answer(s). This method has often been used to measure discounting future health benefits (Chapman and Elstein, 1995; Chapman, 1996; van der Pol and Cairns, 2002; van der Pol and Roux, 2005; van der Pol and Cairns, 2008).

The DM lets a subject compare two simple health profiles with horizon T, which are both combinations of two health states, e.g., $\gamma$ and $\beta$, with $\gamma \succ \beta$. The difference between the profiles is that one starts with the better health state $\gamma$ and ends with the worse state $\beta$: $(\gamma_1,\ldots,\gamma_t,\beta_{t+1},\ldots,\beta_T)$; whereas the other starts with $\beta$, followed by an improvement toward $\gamma$: $(\beta_1,\ldots,\beta_t,\gamma_{t+1},\ldots,\gamma_T)$. Now, the purpose is to elicit the point $t=d^{1/2}$ such that an individual is indifferent between the two profiles, i.e., $(\gamma_1,\ldots,\gamma_t,\beta_{t+1},\ldots,\beta_T) \sim (\beta_1,\ldots,\beta_t,\gamma_{t+1},\ldots,\gamma_T)$. The period $[0,d^{1/2}]$ then has the same utility as $[d^{1/2},T]$:

$$W(d^{1/2})V(\gamma) + W[d^{1/2},T]V(\beta) = W(d^{1/2})V(\beta) + W[d^{1/2},T]V(\gamma). \quad (4)$$

If we denote the health improvement from $\beta$ to $\gamma$ as $X=V(\gamma)-V(\beta)$, we get:

[3] Of course, the indifference point can also be estimated by making use of multiple closed-ended

$$W(T)V(\beta) + W(d^{1/2})X = W(T)V(\beta) + W[d^{1/2},T]X. \quad (5)$$

After simplification, we obtain:

$$W(d^{1/2}) = W[d^{1/2},T] \quad (6)$$

And because $W(d^{1/2})+W[d^{1/2},T]=W(T)=1$, we have:

$$W(d^{1/2}) = 1/2 \quad (7)$$

Hence, we need not know the quality of life weights of the health states involved. After sufficient elicitations, this method allows for a measurement of the complete utility function for life duration. For example, we can next find $d^{1/4}$ such that $W[0,d^{1/4}]=W[d^{1/4},d^{1/2}]$ and, hence, $W(d^{1/4})=1/4$, etc. If an individual would not discount the future, his value of $d^{1/2}$ would be $d^{1/2}=1/2T$ and, accordingly, his utility function would be linear.

The situation changes if the utility of a health state and the utility of life duration are not mutually independent. The utility of life duration may for example be different for different health states. Therefore, we used three health states of differing severity in DM task. The DM has been applied in Attema and Brouwer (2008, 2009, 2010).

*3.2.1. Visual analogue scale*

questions, for example in a discrete choice experiment (van der Pol and Cairns, 2008).

We started the experiment with a VAS with endpoints “worst imaginable health state” and “best imaginable health state” to familiarize subjects with the tasks. First, they were asked to rate their own health state on this scale, followed by three EQ-5D states (see S.3.3), as well as full health and death.

### *3.2.2. Standard gamble and time tradeoff methods*

The indifference sequence was specified as follows. We always started the sequence with two sorting questions, which aimed to establish whether the subject preferred living T years in FH to T years in the impaired health state, and if so, whether it was valued better than dead (BTD) or worse than dead (WTD). To this end, if a subject indicated a preference for immediate death over the certain option in an impaired health state, a WTD procedure was started (Torrance, 1986). That is, the subject then had to evaluate the options “immediate death” and “% chance of full health and % chance of impaired health state” (SG), and the options “immediate death” and “X years in full health and 10-X years in an impaired health state” (TTO). Each sequence comprised five iterations.

## *3.3. Stimuli*

We classified health states according to the EQ-5D system. This system classifies health states using five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), each consisting of three levels, with level 1 meaning no problems on a dimension and level 3 meaning severe problems). We valued the following five EQ-5D health states: 21111, 22222, 32211, 32313 and

33333[4]. These were all included in the study by Lamers *et al.* (2006), from which national EQ-5D tariffs were derived, allowing for a direct comparison with the tariffs. A time frame of T=120 months (10 years) was chosen in the SG and TTO tasks, which facilitates comparability across tasks. The DOI and DM tasks both had a time frame of T=240 months (20 years)[5].

In order to reduce response burden, we divided our sample into two subsamples. Both subsamples received all elicitation tasks (i.e., ranking of health states, ranking of health profiles, SG, TTO, VAS and discounting), but they had to value only three health states each (A: 21111, 32211, 33333; B: 22222, 32313, 33333). Therefore, it was still possible to perform a within-subjects comparison of the utilities resulting from the different methods, and to individually correct TTO scores for discounting.

The DOI method was employed using two different EQ-5D health states, i.e., state 21111 for sample A and state 22222 for sample B. The subjects had to imagine becoming ill (i.e., move from full health to 21111 or 22222) during 30 days after exactly 5 years from now. However, they could take a one-off treatment that would delay the illness by more 5 years, so that it would start 10 years from now. They were then asked whether they preferred treatment, no treatment, or whether they were indifferent. If they were not indifferent, they were asked to state the number of days that did make them indifferent (which had to be a number of at least 30 if they preferred the treatment at first, and no more than 30 if they preferred no treatment at first).

[4] For example, state 32211 stands for: Confined to bed, some problems washing or dressing oneself, some problems with performing one's usual activities, no pain or other discomfort, not anxious or depressed.

[5] This time frame had to be at least 10 years to capture all discounting affecting the TTO task.

The DM also used EQ-5D health states 21111 and 22222 for samples A and B, respectively, but it also included state 33333 in both samples. We deliberately used states that were also included in the health state valuation task, in order to prevent a distortion in the correction of health state utilities for discounting that would arise if subjects would differentially discount different health states. Attema and Brouwer (2012a) reported empirical evidence for such a pattern. The TTO scores for state 32211 [32313] were corrected with the discounting results using both state 21111 [22222] and state 33333. The questions were posed in terms of years and months.

Subjects ranked seven different health profiles in the intertemporal ranking task. Each profile consisted of the same two health states: full health and an impaired EQ-5D health state. The latter was state 21111 for Sample A and 22222 for Sample B. Table I shows the included health profiles for the two samples.

<TABLE I HERE>

*3.4. Analysis*

*3.4.1. Visual analogue scale*

A number of subjects did not place death on the lower end of the scale, and, hence, indicated they did not consider immediate death to be the worst imaginable health state. Therefore, we normalized the VAS scores:

$$VASN = \frac{VAS - D}{100 - D} \tag{8}$$

Consequently, VASN was on a similar scale as SG and TTO [U(D)=0 and U(FH)=1], facilitating comparisons between the methods.

### *3.4.2. Discounting*

The TTO scores were adjusted for the DM discounting estimates by means of the procedure described by Attema and Brouwer (2009). The adjustment for the DOI measure was performed by estimating a discounting parameter under the assumption of exponential discounting and solving the resulting equation representing the TTO indifference. This procedure has been described by Johannesson *et al.* (1994) and van der Pol and Roux (2005).

### *3.4.3. Intertemporal ranking task*

We computed the number of QALYs of the health profiles in the intertemporal ranking exercise by applying the estimated values of health state 21111 [22222] in version A [B] and the linear QALY model to the profiles for each valuation task. For example, the number of QALYs of profile 5 in Table I, using the data for TTO, is equal to $5.5 + 4v_{UTTO}(21111)$, where the subscript UTTO indicates the unadjusted estimate obtained by means of the TTO method. This exercise was repeated for discounted QALYs, using the discounting estimates obtained with the DOI method and the DM, respectively.

We performed three tests also used by Bleichrodt and Johannesson (1997), i.e., we compared consistency with direct ranking using the mean Spearman rank correlation coefficients and two social choice rules: the method of majority voting and the Borda rule. These three tests generated similar results, and, hence, we only report the results of the Spearman tests. The Spearman coefficients were compared using Wilcoxon signed ranks tests. We also performed t-tests which, yielding similar results, are not reported.

## 4. Results

Table II shows summary statistics of the subjects.

<TABLE II HERE>

In Sample A [B], 25 [15] subjects chose a dominated option at least once in the TTO task. We excluded these subjects from the analysis. We did the same in the case of the SG task, which resulted in the exclusion of 11 [11] additional subjects. Furthermore, we removed subjects who did not value FH as the best possible health state in the VAS. This resulted in the exclusion of 39 [40] additional subjects, leaving 187 [192] subjects for the analysis.

Tables IIIa presents the median and mean health utility estimates and compares them with the TTO scores elicited by Lamers *et al.* (2006). Table IIIb does the same, while excluding subjects who chose the same option for all questions of at least one of the two DM tasks (A: 75; B: 74). Although it may be possible that such a choice pattern reflects a subject’s true preferences, this is highly unlikely, since it would for example imply the subject would prefer never being in perfect health to being in perfect health for at least some months (and being in the same health state for the remaining time). Furthermore, Table IIIb excludes 30 [35] more subjects violating dominance in the ranking task at least one time. We performed the analyses both with and without subjects violating dominance in the ranking task.

<TABLES IIIA AND IIIB HERE>

These results show that the TTO scores produced here are fairly comparable to those in the Dutch national tariffs (especially in Table IIIa). Moreover, the TTO scores are fairly similar to the SG scores (but still significantly different for states 21111, 22222, and 3333; Wilcoxon signed ranks test, $p<0.05$). The normalized VAS scores, on the other hand, are substantially lower for milder states ($p<0.06$), and higher for the most severe state ($p<0.01$). This may be the result of endpoint bias. The finding of a positive median VAS value for state 33333 is in accordance with previous studies, which suggest that people tend to take the outcome 'death' to be a natural lower end on a rating scale (Gudex *et al.*, 1996; Robinson *et al.*, 1997); whereas, this kind of behavior is much less obvious for SG and TTO. Moreover, the VAS produced significantly fewer WTD states than the other methods ($p<0.01$, see Table IV).

<TABLE IV HERE>

Tables Va and Vb shows the median corrected TTO scores and the implied CFs[6]. All TTO scores corrected by means of the DM are significantly higher than the

[6] Two CFs are reported for states 32211 and 32313, because we computed a CF using the estimates of both discounting tasks (i.e., discounting elicited both with 21111 and 33333 for 32211, and discounting elicited both with 22222 and 33333 for 32313) for these states (as they are somewhere in-between).

uncorrected ones (p<0.03). When corrected with the DOI method, however, no significant differences result.[7]

<TABLES Va AND Vb HERE>

The CFs make clear that correcting for discounting may have a strong effect on TTO scores, and that this effect increases with the health state's severity. Indeed, the CF is particularly high for the WTD state 33333. This can be partly explained by the higher range for WTD values (minimum -29, so range of 29) as compared to BTD values (0–1), allowing for a higher correction potential for WTD states. We regressed the CFs on several other background characteristics, but none of them turned out to be significantly related to CF.

The CFs estimated by means of the DOI method are considerably lower than those estimated by means of the DM. This finding is in agreement with earlier studies employing the DOI method (Dolan and Gudex, 1995, van der Pol and Roux, 2005). These tend to find a discount rate around 0, as we do for the DOI method.

### *4.1. Intertemporal ranking task*

Table VI shows the results of the intertemporal ranking analysis for Sample A. They provide mixed evidence regarding the effect of correcting for discounting upon the ability to explain ranking of health profiles. Some correlations are quite low, raising the question whether the intertemporal ranking task is able to provide a sound criterion against which to evaluate the validity of the different methods to generate

[7] Note that in Sample A [B], 5 [1] more subjects were excluded from the analysis for the DOI method,

accurate utilities. In Sample A, the VAS scores clearly predict the rankings best (p<0.01). This would indicate that our TTO and SG elicitations are too high. The same holds for the median score of Lamers *et al.* (2006). Furthermore, correcting for discounting does not increase predicted ranking here: instead, it decreases correlation with direct ranking even further, although the difference is not significant (p=0.198) in the sample including ranking violators and significant at the 5% level only when excluding them (p=0.04). Given that the uncorrected TTO scores were already too high, according to the ranking standard at least, this could be expected. One explanation would be that loss aversion causes an upward bias in the TTO and SG biases. If, for the moment, we assume the intertemporal ranking task is a good benchmark, our results suggest that the magnitude of this bias is higher than the magnitude of the bias caused by discounting. Hence, correcting for discounting removes a countervailing force against the upward tendency of loss aversion, changing TTO scores that are too high into TTO scores that are even higher.

<TABLE VI HERE>

The results for Sample B are quite different (Table VII). The VAS is now predicting the rankings the worst (p<0.01), whereas the TTO estimates perform best (although the difference with SG is not significant, p>0.12). The VAS scores therefore seem to be too low in this sample. Correcting for discounting causes no significant change in the accuracy of the prediction generated by the TTO scores.

<TABLE VII HERE>

---

because these gave an answer of 0, implying a discount rate of minus infinity.

## 5. Discussion

Given the clear and potentially large influence discounting may have on health state utilities derived with the TTO method, we set out to find a set of correction factors with which commonly used national tariffs could be corrected. Our results suggest that correcting TTO scores for discounting indeed can have a substantial impact, especially for severe (worse than dead) health states. This holds even though, overall, in our sample the discount rates for future health turned out to be relatively low. Our results, moreover, indicate that the choice for a particular elicitation method for discount rates needs to be well justified, since they produce different estimates regarding discount rates. The results presented in this paper therefore can only serve as a first indication of potential correction factors for TTO tariffs.

Given the different biases present in TTO scores, which work in opposite directions, an important question was whether correcting TTO scores for discounting provide a better indication of health state utilities. This question is not easily answered. We used an intertemporal ranking task to study this. The results from this task provided rather mixed evidence regarding the potential of the correction of TTO scores for discounting to increase TTO's predictive ability. Broadly speaking, corrected TTO scores performed, at best, not worse than uncorrected scores. Whether this is due to the other biases present in TTO or whether such the ranking task may not be fully capable of properly testing face validity of a particular elicitation method, remains to be further investigated. Also given the differences in results between our two subsamples in this respect, we propose that the validity of the intertemporal ranking task used for this purpose needs to be better established. Therefore, more research into developing criteria against which one can compare different estimates is definitely worthwhile.

Our research has several limitations. First, the relatively high amount of dominated and random choices warrants caution. It may highlight the drawbacks of internet surveys, since it is hard to enforce effort among subjects and to ascertain their motivation. Moreover, lay people may have difficulty with the measurement tasks, which are cognitively quite demanding. Discounting elicitation tasks, in particular, seem to be hard to answer for many subjects. This suggests there is a tradeoff between predictive accuracy on the one hand, and cognitive limitations on the other (Dave *et al.*, 2010).

Finding a convenient method to accurately measure discounting of future health outcomes has been an issue for many years, and still is. The DOI method is already quite challenging, needs particular parametric assumptions, and generally elicits discounting estimates close to zero. The latter does not seem to be in line with discounting estimates in other domains. The DM, on the other hand, does not need parametric assumptions, but also seems to be burdensome for general public samples. As such, it generates (very) low discounting estimates as well, contrary to applications of the DM in student samples, where much higher discount rates were reported (Attema and Brouwer, 2012a; Attema *et al.*, forthcoming b). Hence, we recommend the development of a toolkit to elicit discounting of future health outcomes that is easier to grasp for the general population. At the same time, methods capable of investigating the validity of estimated discount rates are warranted, given the differences between the elicitation methods. Guidance regarding which method produces 'better' estimates seems required.

Concluding, we have reported a study deriving correction factors applicable to national tariffs for health state valuations based on the popular TTO method. While it seems pivotal to correct TTO scores for the (several) biases currently distorting them, it is unclear at present which methods are best suited to do so and whether correcting TTO scores for only one bias (discounting), but not for others, results in more accurate health state valuations. Moreover, it seems that sound ways of testing whether corrected scores perform ‘better’ than uncorrected scores are currently lacking. It seems therefore, that much research in this important area is required.

## References


Abellán-Perpinán JM, Bleichrodt H, Pinto-Prades JL. 2009. The predictive validity of prospect theory versus expected utility in health utility measurement. *Journal of Health Economics* **28**(6) : 1039-1047.

Abellán-Perpinán JM, Pinto JL, Méndez-Martinez I, Badia-Llach X. 2006. Towards a better QALY model. *Health Economics* **15**(7) : 665-676.

Attema AE, Brouwer WBF. 2008. Can we fix it? Yes we can! But what? A new test of procedural invariance in TTO-measurement. *Health Economics* **17**(7) : 877-885.

Attema AE, Brouwer WBF. 2009. The correction of TTO-scores for utility curvature using a risk-free utility elicitation method. *Journal of Health Economics* **28**(1) : 234-243.

Attema AE, Brouwer WBF. 2010. The value of correcting values: Influence and importance of correcting TTO scores for time preference. *Value in Health* **13**(8) : 879-884.

Attema AE, Brouwer WBF. 2012a. Constantly Proving The Opposite? A test of CPTO using a broad horizon and correcting for discounting. *Quality of Life Research* **21**(1) : 25-34.

Attema AE, Brouwer WBF. 2012b. In search of a preferred preference elicitation method: A test of the internal consistency of choice and matching procedures.

Working Paper, Erasmus University Rotterdam. Available from http://www.bmg.eur.nl/personal/attema/PrefRev_2012.pdf.

Attema AE, Bleichrodt H, Wakker PP. forthcoming b. A direct method for measuring discounting and QALYs more easily and reliably. *Medical Decision Making*.

Attema AE, Versteegh MM, Oppe M, Brouwer WBF, Stolk EA. forthcoming a. Lead time TTO: Leading to better health state valuations? *Health Economics.*

Bleichrodt H. 2001. Probability Weighting in Choice under Risk: An Empirical Test. *Journal of Risk and Uncertainty* **23**(2) : 185-198.

Bleichrodt H. 2002. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics* **11**(5) : 447-456.

Bleichrodt H, Abellan-Perpiñan JM, Pinto-Prades JL, Mendez-Martinez I. 2007. Resolving Inconsistencies in Utility Measurement Under Risk: Tests of Generalizations of Expected Utility. *Management Science* **53**(3) : 469-482.

Bleichrodt H, Johannesson M. 1997. Standard gamble, time trade-off and rating scale: Experimental results on the ranking properties of QALYs. *Journal of Health Economics* **16**(2) : 155-175.

Bleichrodt H, Pinto JL. 2000. A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis. *Management Science* **46**(11) : 1485-1496.

Bleichrodt H, Pinto JL. 2002. Loss aversion and scale compatibility in two-attribute trade-offs. *Journal of Mathematical Psychology* **46**(3) : 315-337.

Bleichrodt H, Pinto JL. 2005. The validity of QALYs under non-expected utility. *The Economic Journal* **115**(503) : 533-550.

Bleichrodt H, van Rijn J, Johannesson M. 1999. Probability weighting and utility curvature in QALY-based decision making. *Journal of Mathematical Psychology* **43**(2) : 238-260.

Bostic R, Herrnstein RJ, Luce RD. 1990. The effect on the preference-reversal phenomenon of using choice indifferences. *Journal of Economic Behavior & Organization* **13**(2) : 193-212.

Cairns J. 1992. Health, wealth and time preference. *Project Appraisal* **7**(1) : 31-40.

Chapman GB. 1996. Temporal discounting and utility for health and money. *Journal of Experimental Psychology: Learning, Memory & Cognition* **22**(3) : 771-791.

Chapman GB, Elstein AS. 1995. Valuing the future: temporal discounting of health and money. *Medical Decision Making* **15**(4) : 373-386.

Dave C, Eckel C, Johnson C, Rojas C. 2010. Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty* **41**(3) : 219-243.

Dolan P. 2000. The measurement of health-related quality of life for use in resource allocation decisions in health care. In: Culyer AJ and Newhouse JP, eds, *Handbook of Health Economics.* North Holland: Elsevier, pp. 1723-1760.

Dolan P, Gudex C, Kind P, Williams A. 1996. The time trade-off method: Results from a general population study. *Health Economics* **5**(2) : 141-154.

Gravelle H, Brouwer W, Niessen L, Postma M, Rutten F. 2007. Discounting in Economic Evaluations: Stepping forward Towards Optimal Decision Rules. *Health Economics* **16 3** 307-317.

Gudex C, Dolan P, Kind P, Williams A. 1996. Health State Valuations from the General Public Using the Visual Analogue Scale. *Quality of Life Research* **5**(6) : pp. 521-531.

Hey JJD, Morone A, Schmidt U. 2009. Noise and bias in eliciting preferences. *Journal of Risk and Uncertainty* **39**(3) : 213-235.

Johannesson M, Pliskin JS, Weinstein MC. 1994. A note on QALYs, time tradeoff, and discounting. *Medical Decision Making* **14**(2) : 188-193.

Kahneman D, Tversky A. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47**(2) : 263-291.

Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJV. 2006. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Economics* **15**(10) : 1121-1132.

Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. 1982. The measurement of patients' values in medicine. *Medical Decision Making* **2**(4) : 449-462.

MacKeigan LD, Gafni A, O'Brien BJ. 2003. Double discounting of QALYs. *Health Economics* **12**(2) : 165-169.

Martin AJ, Glasziou PP, Simes RJ, Lumley T. 2000. A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *International Journal of Technology Assessment in Health Care* **16**(1) : 137-147.

McNeil BJ, Pauker SG, Sox HC, Tversky A. 1982. On the Elicitation of Preferences for Alternative Therapies. *New England Journal of Medicine* **306**(21) : 1259-1262.

Miyamoto J, Eraker SA. 1985. Parameter estimates for a QALY utility model. *Medical Decision Making* **5**(2) : 191-213.

Parkin D, Devlin N. 2006. Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Economics* **15**(7) : 653-664.

Robinson A, Dolan P, Williams A. 1997. Valuing health status using VAS and TTO: what lies behind the numbers? *Social Science & Medicine* **45**(8) : 1289-1297.

Rutten-van Molken MPMH, Bakker CH, Van Doorslaer EKA, van der Linden S. 1995. Methodological Issues of Patient Utility Measurement: Experience from Two Clinical Trials. *Medical Care* **33**(9) : 922-937.

Stalmeier PF, Bezembinder TG. 1999. The discrepancy between risky and riskless utilities: a matter of framing? *Medical Decision Making* **19**(4) : 435-447.

Stalmeier PF, Bezembinder TG, Unic IJ. 1996. Proportional heuristics in time tradeoff and conjoint measurement. *Medical Decision Making* **16**(1) : 36-44.

Starmer C. 2000. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* **38**(2) : 332-382.

Stiggelbout AM, Kiebert GM, Kievit J, Leer JW, Stoter G, de Haes JC. 1994. Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. *Medical Decision Making* **14**(1) : 82-90.

Torrance GW. 1986. Measurement of health state utilities for economic appraisal. *Journal of Health Economics* **5**(1) : 1-30.

Tversky A, Kahneman D. 1991. Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics* **106**(4) : 1039-1061.

Tversky A, Kahneman D. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* **5**(4) : 297-323.

van der Pol M, Cairns J. 2001. Estimating time preferences for health using discrete choice experiments. *Social Science & Medicine* **52**(9) : 1459-1470.

van der Pol M, Cairns J. 2002. A comparison of the discounted utility model and hyperbolic discounting models in the case of social and private intertemporal preferences for health. *Journal of Economic Behavior and Organization* **49**(1) : 79-96.

van der Pol M, Cairns J. 2008. Comparison of two methods of eliciting time preference for future health states. *Social Science & Medicine* **67**(5) : 883-889.

van der Pol M, Roux L. 2005. Time preference bias in time trade-off. *European Journal of Health Economics* **6**(2) : 107-111.

Wakker P, Deneffe D. 1996. Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown. *Management Science* **42**(8) : 1131-1150.

Wakker P, Stiggelbout A. 1995. Explaining Distortions in Utility Elicitation through the Rank-dependent Model for Risky Choices. *Medical Decision Making* **15**(2) : 180-186.

## TABLES

**Table I. Stimuli intertemporal ranking task**

| |
|---|
| Sample A |
| Profile 1) 10 years in 21111, then die |
| Profile 2) 9.5 years in 11111, then die |
| Profile 3) 9 years in 11111, then die |
| Profile 4) 8.5 years in 11111, then 1 year in 21111, then die |
| Profile 5) 5.5 years in 11111, then 4 years in 21111, then die |
| Profile 6) 3.5 years in 11111, then 6 years in 21111, then die |
| Profile 7) 3 years in 11111, then 7 years in 21111, then die |
| Sample B |
| Profile 1) 10 years in 22222, then die |
| Profile 2) 6.5 years in 11111, then die |
| Profile 3) 6 years in 11111, then die |
| Profile 4) 5.5 years in 11111, then 1 year in 22222, then die |
| Profile 5) 4 years in 11111, then 4 years in 22222, then die |
| Profile 6) 2.5 years in 11111, then 6 years in 22222, then die |
| Profile 7) 2 years in 11111, then 7 years in 22222, then die |

**Table II. Summary statistics**

SAMPLE A (N=262)

| Variable | Percentage | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|---|
| Age | | 41.5 | 12.88 | 18 | 65 |
| Gender (% male) | 47.7 | | | | |
| Children (%yes) | 56.9 | | | | |
| Number of children (among people with children, n=149) | | 2.15 | | 1 | 9 |
| Income groups: | | | | | |
| <€1000 | 11.1 | | | | |
| €1000-<€2000 | 37.0 | | | | |
| €2000-<€3000 | 30.2 | | | | |
| €3000-<€4000 | 14.9 | | | | |
| >€3999 | 6.9 | | | | |
| Education: | | | | | |
| Lower | 24.0 | | | | |
| Middle | 42.4 | | | | |
| Higher | 33.6 | | | | |
| Health status | | | | | |
| EQ-5D (Dutch tariff) | | 0.88 | 0.18 | 0.009 | 1 |
| VAS | | 77.46 | 16.46 | 4 | 100 |
| Completion time (mins.) | | 19.1 | 12.0 | 6.1 | 91.1 |

SAMPLE B (N=258)

| Variable | Percentage | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|---|
| Age | | 42.1 | 12.13 | 18 | 65 |
| Gender (% male) | 46.9 | | | | |
| Children (%yes) | 63.6 | | | | |
| Number of children (among people with children, n=164) | | 2.09 | | 1 | 8 |
| Income groups: | | | | | |
| <€1000 | 5.8 | | | | |
| €1000-<€2000 | 33.7 | | | | |
| €2000-<€3000 | 37.2 | | | | |
| €3000-<€4000 | 15.9 | | | | |
| >€3999 | 7.4 | | | | |
| Education: | | | | | |
| Lower | 20.5 | | | | |
| Middle | 45.7 | | | | |
| Higher | 33.7 | | | | |
| Health status | | | | | |
| EQ-5D (Dutch tariff) | | 0.86 | 0.22 | 0.09 | 1 |
| VAS | | 76.33 | 17.39 | 10 | 100 |
| Completion time (mins.) | | 22.3 | 10.5 | 5.2 | 78.9 |

**Table IIIa. Median (mean) health state utilities** (excluding subjects with dominated choices)

| EQ-5D Health state | VASN | SG | TTO | TTO Lamers *et al.* (2006) | SAMPLE |
|---|---|---|---|---|---|
| 21111 | 0.78 (0.68) | 0.91 (0.75) | 0.97 (0.84) | 0.99 (0.91) | A (n=187) |
| 22222 | 0.44 (0.16) | 0.59 (-0.37) | 0.72 (0.17) | 0.68 (0.54) | B (n=192) |
| 32211 | 0.39 (0.28) | 0.59 (-0.68) | 0.60 (-0.22) | 0.55 (0.42) | A (n=187) |
| 32313 | 0.23 (-0.05) | 0.03 (-1.81) | 0.03 (-2.19) | 0.03 (0.04) | B (n=192) |
| 33333 | 0.09 (-0.25) | -0.28 (2.96) | -0.67 (-4.17) | -0.38 (-0.30) | A+B (n=379) |

**Table IIIb. Median (mean) health state utilities** (excluding subjects with dominated choices, subjects excluded from discounting task [either for mild one or for severe one] and subjects violating ranking)

| EQ-5D Health state | VASN | SG | TTO | TTO Lamers *et al.* (2006) | SAMPLE |
|---|---|---|---|---|---|
| 21111 | 0.78 (0.72) | 0.91 (0.80) | 0.93 (0.87) | 0.99 (0.91) | A (n=82) |
| 22222 | 0.44 (0.27) | 0.65 (0.16) | 0.72 (0.20) | 0.68 (0.54) | B (n=83) |
| 32211 | 0.42 (0.30) | 0.59 (0.38) | 0.47 (0.06) | 0.55 (0.42) | A (n=82) |
| 32313 | 0.21 (-0.03) | -0.03 (-0.98) | -0.03 (-2.19) | 0.03 (0.04) | B (n=83) |
| 33333 | 0.08 (-0.15) | -0.28 (-1.86) | -0.67 (-3.87) | -0.38 (-0.30) | A+B (n=165) |

**Table IV. Percentage WTD responses**

| EQ-5D Health state | VASN | SG | TTO | TTO Lamers *et al.* (2006) | SAMPLE |
|---|---|---|---|---|---|
| 21111 | 3.7 | 3.7 | 2.1 | 1 | A (n=187) |
| 22222 | 10.4 | 11.5 | 9.9 | 12 | B (n=192) |
| 32211 | 8.0 | 14.4 | 15.0 | 16 | A (n=187) |
| 32313 | 17.2 | 45.8 | 49.0 | 39 | B (n=192) |
| 33333 | 20.6 | 71.0 | 72.3 | 62 | A+B (n=379) |

**Table Va. Median corrected TTO scores and implied CFs**

| EQ-5D Health state | CTTO DM | CF DM | CTTO DOI* | CF DOI | SAMPLE |
|---|---|---|---|---|---|
| 21111 | 0.963 | -0.003 | 0.94 (n=107) | -0.03 | A (n=112) |
| 22222 | 0.780 | 0.06 | 0.72 (n=117) | 0 | B (n=118) |
| 32211 | 0.610 (21111) 0.654 (33333) | 0.11 (21111) 0.154 (33333) | 0.52 (n=107) | 0.05 | A (n=112) |
| 32313 | -0.03 (22222) -0.02 (33333) | 0 (22222) 0.01 (33333) | -0.03 (n=117) | 0 | B (n=118) |
| 33333 | -0.32 | 0.35 | -0.54 (n=224) | 0.06 | A+B (n=230) |

**Table Vb. Median corrected TTO scores and implied CFs without ranking violators**

| EQ-5D Health state | CTTO DM | CF DM | CTTO DOI | CF DOI | SAMPLE |
|---|---|---|---|---|---|
| 21111 | 0.97 | 0.04 | 0.92 n=78 | 0.02 | A (n=82) |
| 22222 | 0.78 | 0.06 | 0.72 n=82 | 0 | B (n=83) |
| 32211 | 0.61 (21111) 0.63 (33333) | 0.14 (21111) 0.16 (33333) | 0.51 n=78 | 0.04 | A (n=82) |
| 32313 | -0.02 (22222 and 33333) | 0.01 (22222 and 33333) | -0.03 n=82 | 0 | B (n=83) |
| 33333 | -0.40 | 0.27 | -0.67 n=160 | 0 | A+B (n=165) |

**Table VI. Spearman rank correlations coefficients of tasks with direct rankings (Sample A)**

| **Method** | **Median (Mean) [n=112]** | **Excluding ranking violators [n=82]** |
|---|---|---|
| Corrected mean TTO Lamers | 0.21 (0.19) | 0.18 (0.16) |
| Corrected TTO (DM) | 0.19 (0.17) | 0.21 (0.18) |
| Corrected TTO (profile discounted DM) | 0.22 (0.18) | 0.29 (0.22) |
| Corrected TTO (DOI) | 0.29 (0.20) n=107 | 0.32 (0.24) n=78 |
| Corrected TTO (profile discounted DOI) | 0.32 (0.21) n=107 | 0.38 (0.24) n=78 |
| Mean TTO Lamers | 0.39 (0.38) | 0.39 (0.40) |
| SG | 0.39 (0.35) | 0.39 (0.34) |
| SG (profile discounted DM) | 0.36 (0.31) | 0.39 (0.35) |
| SG (profile discounted DOI) | 0.43 (0.35) n=107 | 0.43 (0.33) n=78 |
| TTO | 0.30 (0.21) | 0.38 (0.25) |
| TTO (profile discounted DM) | 0.29 (0.23) | 0.34 (0.30) |
| TTO (profile discounted DOI) | 0.29 (0.21) n=107 | 0.29 (0.23) n=78 |
| VAS | 0.71 (0.65) | 0.79 (0.68) |
| VAS (profile discounted DM) | 0.73 (0.59) | 0.79 (0.66) |
| VAS (profile discounted DOI) | 0.71 (0.65) n=107 | 0.75 (0.67) n=78 |

**Table VII. Spearman rank correlations coefficients of tasks with direct rankings (Sample B)**

| **Method** | **Median (Mean) [n=118]** | **Excluding ranking violators [n=83]** |
|---|---|---|
| Corrected TTO (DM) | 0.47 (0.32) | 0.68 (0.36) |
| Corrected TTO (profile discounted DM) | 0.5 (0.31) | 0.5 (0.34) |
| Corrected median TTO Lamers (DM) | -0.14 (-0.12) | -0.11 (-0.07) |
| Corrected TTO (DOI) | 0.43 (0.31) n=117 | 0.61 (0.36) n=82 |
| Corrected TTO (profile discounted DOI) | 0.5 (0.31) n=117 | 0.51 (0.34) n=82 |
| Median TTO Lamers | 0.48 (0.27) | 0.39 (0.24) |
| SG | 0.36 (0.24) | 0.46 (0.30) |
| SG (profile discounted DM) | 0.32 (0.23) | 0.5 (0.29) |
| SG (profile discounted DOI) | 0.36 (0.24) n=117 | 0.46 (0.29) n=82 |
| TTO | 0.59 (0.34) | 0.68 (0.38) |
| TTO (profile discounted DM) | 0.46 (0.27) | 0.5 (0.33) |
| TTO (profile discounted DOI) | 0.54 (0.31) n=117 | 0.54 (0.33) n=82 |
| VAS | 0.04 (0.00) | 0.07 (0.03) |
| VAS (profile discounted DM) | 0.05 (0.01) | 0.07 (0.05) |
| VAS (profile discounted DOI) | 0.11 (0.05) n=117 | 0.16 (0.07) n=82 |