

Medical Decision Making

<http://mdm.sagepub.com/>

A Direct Method for Measuring Discounting and QALYs More Easily and Reliably

Arthur E. Attema, Han Bleichrodt and Peter P. Wakker

Med Decis Making 2012 32: 583 originally published online 15 June 2012

DOI: 10.1177/0272989X12451654

The online version of this article can be found at:

<http://mdm.sagepub.com/content/32/4/583>

Published by:



<http://www.sagepublications.com>

On behalf of:



<http://www.smdm.org>

Additional services and information for *Medical Decision Making* can be found at:

Email Alerts: <http://mdm.sagepub.com/cgi/alerts>

Subscriptions: <http://mdm.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - Jul 20, 2012

[OnlineFirst Version of Record](#) - Jun 15, 2012

[What is This?](#)

A Direct Method for Measuring Discounting and QALYs More Easily and Reliably

Arthur E. Attema, PhD, Han Bleichrodt, PhD, Peter P. Wakker, PhD

Time discounting and quality of life are two important factors in evaluations of medical interventions. The measurement of these two factors is complicated because they interact. Existing methods either simply assume one factor given, based on heuristic assumptions, or invoke complicating extraneous factors, such as risk, that generate extra biases. The authors introduce a method for measuring discounting (and then quality of life) that involves no extraneous factors and

that avoids distorting interactions. Their method is considerably simpler and more realistic for subjects than existing methods. It is entirely choice based and thus can be founded on economic rationality requirements. An experiment demonstrates the feasibility of this method and its advantages over classical methods.
Key words: utility measurement; discounting; QALY; utility of life duration; time tradeoff (*Med Decis Making* 2012;32:583–593)

Quality-adjusted life year (QALY) evaluations integrate two components, quality of life and discounting (utility of life duration). These components interact, and it is hard to measure one without knowing the other.¹ Most measurements of one component assume the other component known, based on heuristic assumptions. Thus, in time tradeoff (TTO) measurements of quality of life, utility of life duration is usually assumed to be linear (no discounting). In measurements of discounting, outcomes are usually monetary, with the utility of money assumed to be linear. Whether discounting of health can be equated with discounting of money remains a point of debate.^{2–5}

Only few attempts have been made to avoid the interaction between discounting and quality of life,

and these invoke extraneous factors such as risky or interpersonal (utilitarian) aggregations.^{1,6–10} Then attitudes toward risk and welfare intervene and generate additional biases.^{11–16}

This article presents a method to measure discounting within the QALY model that avoids the aforementioned problems: Our method (a) needs no extraneous factors, (b) is not affected by the interactions between discounting and quality of life, and (c) uses stimuli that can be simpler and more realistic than those for existing methods. We can measure any general discount function, constant or not. Because we avoid all extraneous factors and interactions, we call our method the direct method (DM). With utility of life duration (discounting) measured, we can also measure quality of life by correcting traditional TTO measurements. We can thus measure the whole QALY model.

Unlike classical methods for measuring the utility of life duration and discounting (the standard gamble method and the certainty equivalence [CE] method), which are based on risky decisions, we need not invoke the outcome of immediate death. This outcome is very aversive and is known to arouse negative and distorting emotions.* Especially problematic for classical measurements is that, besides immediate death, they also need scenarios of sure death at

Received 26 February 2011 from iBMG/iMTA (AEA, HB) and the Department of Economics (HB, PPW), Erasmus University, Rotterdam, the Netherlands. Bas Karreman, Tom Van Ourti, Kirsten Rohde, and Stefan Trautmann made useful suggestions for the experiment. Guido Attema helped to develop the computer program. A first version of this article circulated as Attema, Bleichrodt, and Wakker (2007), "Measuring the Utility of Life Duration in a Risk-Free versus a Risky Situation." Arthur Attema's and Han Bleichrodt's research was made possible through grants from the Netherlands Organization for Scientific Research (NWO). Revision accepted for publication 11 September 2011.

Address correspondence to Han Bleichrodt, PhD, Department of Economics, Erasmus University, PO Box 1738, 3000 DR Rotterdam, the Netherlands; e-mail: bleichrodt@ese.eur.nl.

DOI: 10.1177/0272989X12451654

*Sometimes an outcome of death within a week or month is used instead, mitigating the aversiveness at the price of a small inaccuracy in measurement.^{8,59}

some precisely fixed future time point already specified and known at present. Subjects have great difficulties imagining such unrealistic scenarios, leading to misunderstandings and distortions. Our method for measuring the utility of life duration and discounting avoids such scenarios and leaves the time of death unspecified, as it is in reality.[†] This enhances realism and applicability. Furthermore, our method can be entirely based on observable decisions (revealed preferences) and does not require introspective data (stated preferences). Hence, it is completely grounded on the normative rationality requirements of economics.¹⁵

METHODS

Existing QALY Measurements and Their Restrictions

To prepare for our method, we present a number of classical evaluation models in increasing order of generality.

The linear QALY model (no discounting). This is the first refinement of life duration as outcome measure. Life years are adjusted for quality of life. Discounting is not yet incorporated. Figure 1 illustrates this evaluation for a health profile of 5 years at various levels of quality of life, followed by death.

The QALY value is

$$1 \times 1 + 1 \times 1/2 + 1 \times 3/4 + 1 \times 1/2 + 1 \times 1/2 + 0 = 3.25. \quad (1)$$

Here the quality of life of a health state can easily be measured using the well-known TTO method:

If 10 years in health state *Q* is equally preferred as 8 years in perfect health, then (2)
the quality of life in *Q* is the ratio of the life durations, 8/10, which is 80%.

The constant-discounted QALY model. Now the weight of each future year *j* is reduced by a discount factor

$$r^{j-1}, \quad (3)$$

[†]If we also measure quality of life and want to include (scalings relative to) death, then we obviously cannot avoid using this health state. See the comment in the elaborated example presented later.

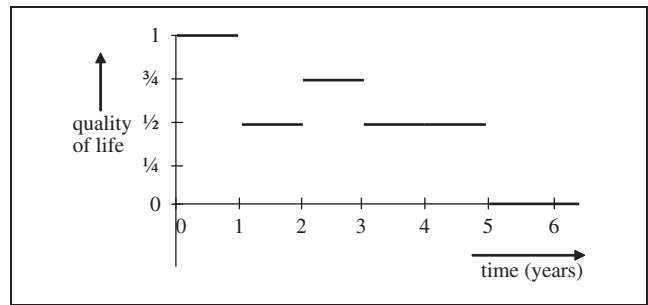


Figure 1. Example of a health profile.

with $1 - r$ the discount rate. In standard cost-effectiveness studies, the discount rate $1 - r$ is often assumed to be 0.03. Then the resulting QALY in Figure 1 is

$$1 \times 1 + 0.97 \times 1/2 + 0.97^2 \times 3/4 + 0.97^3 \times 1/2 + 0.97^4 \times 1/2 = 3.09. \quad (4)$$

The *utility of a period of life* is the QALY value of this period when spent in perfect health. That is, it is the discounted number of years in question. Periods are denoted by their beginning and their end, as in [5,10] for years 6 to 10. Here 5 refers to the time point after 5 years, which is the start of year 6, and the difference $10 - 5 = 5$ is the duration of the period. We chose this notation to be consistent with interval notation for continuous time. Writing $U[5,10]$ for the utility of this period, we have, with $r = 0.97$,

$$U[5, 10] = \sum_{j=6}^{10} r^{j-1} = \sum_{j=6}^{10} 0.97^{j-1} = 4.04.$$

The utility of the first 10 years to come, $U[0,10]$, then is

$$\sum_{j=1}^{10} r^{j-1} = \sum_{j=1}^{10} 0.97^{j-1} = 8.75.$$

Conversely, if the utility U of life duration is given, then the discount factor for year j can be obtained as $U[j - 1, j] (= r^{j-1})$, being the incremental utility of prolonging $j - 1$ life years to j life years. Utility of life duration and discounting are two different but equivalent ways of expressing time preference. We often suppress the beginning of a period if it is 0, writing $U(10)$ as shorthand for $U[0,10]$.

The (general) QALY model (nonconstant discounting). In general, discounting need not be constant, and year j may have general utility $U[j - 1, j]$ that is different from r^{j-1} . Then the (general) QALY value in Figure 1 is

$$U[0, 1] \times 1 + U[1, 2] \times 1/2 + U[2, 3] \times 3/4 + U[3, 4] \times 1/2 + U[4, 5] \times 1/2. \tag{5}$$

We still have the following relations between utility of life years and discounting:

The discount factor for a year j is its utility $U[j - 1, j]$; the utility U of a period is equal to the sum of the discounted life years in that period. $\tag{6}$

In the continuous case, the discount factor is the derivative of the utility of life duration and, vice versa, the utility of life duration is the integral of the discount factor.

Not only discounting (i.e., U) but also quality of life is unknown beforehand and has to be measured. The TTO observation of equation (2) now implies a quality of life of

$$U(8)/U(10). \tag{7}$$

With U unknown, we cannot easily know what the quality of life of health states is. This demonstrates how the interaction of discounting and quality of life complicates their measurement. It is not readily clear how one can be measured if we do not know the other.

Because of the complication just explained, the standard gamble method, which is subject to distortion by risk attitude, and the visual analog scale (VAS), which has no economic (revealed-preference) foundation,¹⁷ are sometimes used as alternatives. The main result of this article will show how U can be measured in general under the assumptions of the QALY model. Then, with U available, we can readily measure quality of life using equation (7). We can then measure the complete QALY model without needing any additional assumption.

More general (non-QALY) models. More general models can be considered, with interactions between different time periods or with nonmultiplicative interactions between discounting and quality of life. Such general models are not commonly used in health because it is not clear how they can be measured or implemented, and we do not consider them here.

The Direct Method for Measuring Utility of Life Duration

We assume the general QALY model throughout. Suppose that an improvement in quality of life from

$1/2$ to $3/4$ is possible in Figure 1, and that it is possible either in period [1,2] (year 2) or in period [3,5] (years 4 and 5). Write X for the quality-of-life difference $3/4 - 1/2 = 1/4$. Assume further that these two improvements are equally preferred, implying that their QALY gains are the same:

$$U[1, 2]X = U[3, 5]X. \tag{8}$$

Dropping the common factor X gives

$$U[1, 2] = U[3, 5]. \tag{9}$$

A convenient feature of the measurement just described is that we need not know quality-of-life X because it drops out anyhow. To see this point in general, imagine that some health profile yields a health state β (β for bad) in two different periods, $P = [P_1, P_2]$ and $Q = [Q_1, Q_2]$. Imagine that improving β into a health state γ (γ for good) is equally preferred for period P as for period Q . Then we have the following equality for the total QALY gains, where X denotes the quality-of-life difference between health states β and γ :

$$U[P_1, P_2]X = U[Q_1, Q_2]X. \tag{10}$$

This implies

$$U[P_1, P_2] = U[Q_1, Q_2] \tag{11}$$

because we can drop X irrespective of what it is (as long as it is not zero).

Another desirable feature of the measurement just proposed is that we need not know or specify the health states outside the two periods considered, as long as these are kept constant. In the above calculation based on QALY gains, we simply did not need such information. Whatever the other periods contribute to the total QALY evaluations is immaterial for the QALY gains considered. In Figure 1, with the improvement in year 2 equally preferred as the improvement in years 4 and 5, the health states in years 1 and 3 are immaterial for the conclusion of our QALY analysis. Importantly, after 5 years, no immediate death has to follow, but any realistic health profile may be assumed. All of this does not affect our inference of $U[1,2] = U[3,5]$.

The observations just made allow measuring the utility of life duration to any desired degree of precision.[‡] If $[0, D]$ is the total period of interest, then

[‡]In what follows, we will use bisection techniques from the psychological literature⁶⁰ within a revealed-preference setup.

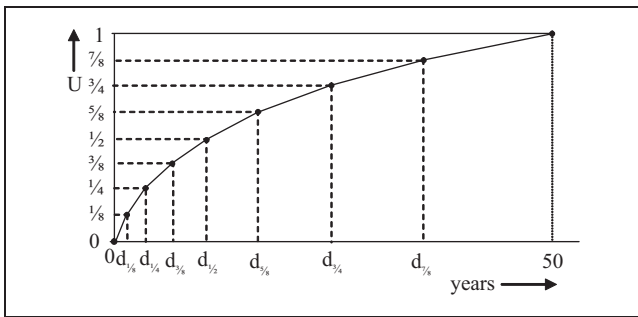


Figure 2. Utility graph of life duration.

we can normalize $U(D) = 1$. Obviously, $U(0) = 0$. We first find $d^{1/2}$ such that the period $[0, d^{1/2}]$ has the same utility as $[d^{1/2}, D]$. Then $U(d^{1/2}) = 1/2$. We next find $d^{1/4}$ and $d^{3/4}$ such that $U[0, d^{1/4}] = U[d^{1/4}, d^{1/2}]$ and $U[d^{1/2}, d^{3/4}] = U[d^{3/4}, D]$. Then $U(d^{1/4}) = 1/4$ and $U(d^{3/4}) = 3/4$. We can continue this bisection procedure to any desired degree of precision and obtain the entire graph of U this way. Figure 2 depicts such a graph of U . The method just described is called the *direct method (DM)*. As explained by equation (6), the graph also captures discounting.

ELABORATED EXAMPLE

This section presents a numerical example to demonstrate the simplicity and generality of the DM and the way it can also be used to measure not only discounting but also quality of life. Assume health states P (poor health), M (mediocre health), and F (fair health). $([i, j]: Q)$ denotes health state Q in period $[i, j]$ (in days), good health for the rest of the coming 2 years, and a regular health profile thereafter (which need not be specified). Assume the indifferences

$$([0, 100]: P) \sim ([100, 260]: P) \sim ([260, 510]: P).$$

Then, $U[0, 100] = U[100, 260] = U[260, 510]$. Scaling these values to be 0.25, linear interpolation gives the bold dashed line in Figure 3.

Assume that we further observe

$$([0, 100]: P) \sim ([0, 260]: M) \sim ([0, 510]: F).$$

Because we have just measured that the utility of period $[0, 260]$ is twice the utility of period $[0, 100]$, we infer that the loss of quality of life due to P is twice that due to M . Similarly, it is 3 times the loss of quality of life due to F . Thus, curing P is worth 3 times more than curing F when it is over the same period.

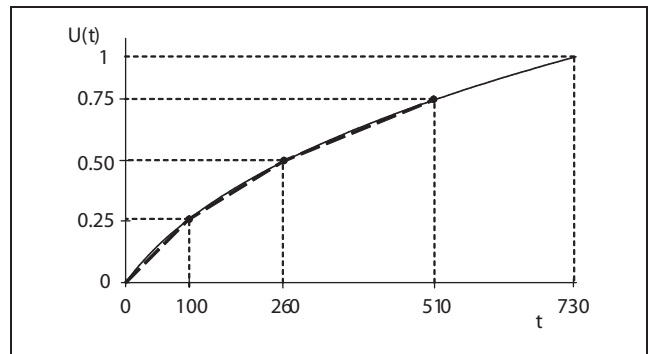


Figure 3. Utility of life duration.

If we want to scale quality of life on the 0 to 1 death-life scale, then we obviously have to include the death outcome. An indifference

$$([0, 100]; \text{perfect health; death after}) \sim ([0, 260]; M; \text{death after})$$

then reveals $U(M) = 1/2$,
implying $U(P) = 1/4$ and $U(F) = 3/4$.

As the example shows, including the death outcome implies that we have to specify the precise time of death, losing one advantage of our method. To retain that advantage, we can use, instead of death, a health state equivalent to death that can be combined with life-as-usual after, such as being unconscious.^{18,19}

The measurements and derivations in this example were all elementary, fully preference based under the normative principles of economics, and valid under all QALY models described before. They completely identify utility of life duration (i.e., time discounting) and quality of life, where these components have been completely disentangled. It can be seen that our method is mathematically analogous to the measurement of subjective *probability* (rather than utility) under expected utility.¹⁵

EXPERIMENT

We implement the DM in an experiment to demonstrate its feasibility and to compare it with the CE (certainty equivalent) method, the classical method for measuring the utility of life duration in health research.

Subjects

Seventy students (30 female) from various departments of Erasmus University Rotterdam participated. They were recruited using e-mail, poster

advertisements, and flyers distributed on the university campus.

Procedure

We tested our design in several pilot sessions. The experiment was computer run and administered in sessions with no more than 2 subjects. An experimenter was present during each session. All subjects finished the experimental session within 45 minutes. They were paid a fixed amount of €12.50 for participating. To avoid order effects, we randomized the order of the DM and the CE across sessions. The 2 methods were administered successively. Both methods were preceded by 2 practice questions.

All indifferences were elicited using sequences of at most 5 binary choices rather than eliciting them through a single matching question. Choice-based elicitation is more time-consuming but causes fewer inconsistencies.²⁰ The indifferences were elicited iteratively. After each choice, the subject was asked to confirm it. At the end of the iteration process, the first choice of the process was repeated. If the respondent changed this choice, then the iteration process recommenced. To further check for consistency, the elicitation of the first indifference value was repeated at the end of each method.

Stimuli of the DM

For the bad health state, we took regular back pain (β = bad back) because it is well known. We described this health state using the EQ-5D questionnaire that has been widely used and validated²¹ and denoted it by EQ-5D state 11221. For the good health state γ , we took full health (i.e., EQ-5D state 11111). It was explained that this health state meant being able to function perfectly well on all five Euro-Qol dimensions, irrespective of age. The descriptions of β and γ were printed on cards and handed to the subjects.

We investigated the utility function over the next 50 years, that is, over the time interval [0,50]. We normalized $U(50) = 1$. The reference health profile was β throughout all 50 years. We told the subjects that after 50 years, all options gave the same health profile without further specifying it. Subjects could choose between periods during which β would be improved to γ . In the first question, we determined $d^{1/2}$ such that $U(d^{1/2}) = 1/2 * U(50) = 1/2$. That is, improving β to γ during the period [0, $d^{1/2}$] is equally preferred as doing so during the period [$d^{1/2}$, 50]. We further elicited $d^{1/8}$, $d^{1/4}$, $d^{3/4}$, and $d^{7/8}$, with utilities $1/8$, $1/4$, $3/4$, and $7/8$.

Stimuli of the CE Method (Risk Involved)

In the CE part of the experiment, we assumed full health throughout and considered outcomes in terms of years, with, for instance, b denoting b years in full health followed by immediate death. In general, subjects had to determine a risk-free option b (i.e., b for sure) that was equivalent to a risky option $a_{1/2}c$ (denoting probability $1/2$ at a years in full health and probability $1/2$ at c years in full health). We assume $a > b > c$. We thus successively obtained the following right-hand side $c^{\text{superscript}}$ from indifferences:

$$50_{1/2}0 \text{ equally preferred as } c^{1/2};$$

$$c^{1/2}_{1/2}0 \text{ equally preferred as } c^{1/4};$$

$$50_{1/2}c^{1/2} \text{ equally preferred as } c^{3/4};$$

$$c^{1/4}_{1/2}0 \text{ equally preferred as } c^{1/8};$$

$$50_{1/2}c^{3/4} \text{ equally preferred as } c^{7/8}.$$

CEs were analyzed in two ways. The first was the classical approach, based on expected utility. The utility function is denoted by the same symbol U as in the QALY model, as they are commonly assumed to be the same in QALY analyses, despite concerns about this assumption.²² Avoiding reliance on this controversial assumption is one of the advantages of the DM.

For an observation that b is equally preferred as $a_{1/2}c$, expected utility implies that $U(b)$ is the midpoint of $U(a)$ and $U(c)$ —that is,²³

$$U(b) = \frac{U(a) + U(c)}{2}. \tag{12}$$

Under the normalization $U(50) = 1$ and $U(0) = 0$, superscripts give utility levels—that is, $U(c^{1/2}) = 1/2$, $U(c^{1/4}) = 1/4$, and so on. We call this classical way to derive utilities from CE data the *CEE method*, where the final E refers to expected utility.

Our second approach to analyzing CEs employs the empirically more realistic prospect theory^{24,25} instead of expected utility, following the method of Wakker and Stiggelbout.²⁶ With immediate death as the reference point, b being equally preferred as $a_{1/2}c$ now implies

$$U(b)=w(1/2) \times U(a) + (1 - w(1/2)) \times U(c). \quad (13)$$

Here $w(1/2)$ is the weight of probability $1/2$. Tversky and Kahneman's estimate²⁵ $w(1/2) = 0.42$ has been found to perform well at the aggregate level. It implies that the worst outcome c is relatively overweighted. This overweighting captures part of the risk aversion that is empirically observed but that cannot be properly captured by the U function. From equation (13), we can calculate all utilities $U(c^{\text{superscript}})$. They will be lower than the utilities under expected utility. We call this method the *CEP method*, where the letter P refers to prospect theory.

Convenience of the Methods

At the end of the experiment, the subjects were asked to rate both the DM and the CE on a scale from 1 (worst) to 7 (best), in terms of understandability and cognitive burden.

Statistical Analyses of DM v. CEE (Assuming the Same Utility Function U)

We first test the null hypothesis H_0 that expected utility holds for the risky CE questions and that utility U from expected utility is the same as the utility to be used in QALY evaluations. Under H_0 , $c^{\text{superscript}} = d^{\text{superscript}}$ should hold for every superscript. Because both the c -values and the d -values are chained, they are not independent, and direct tests of the above equalities would not be independent. We follow Miyamoto and Eraker's solution²⁷ to this problem by testing proportional matches. We test the following 5 equalities predicted by H_0 , using Wilcoxon signed ranks tests:

$$c^{1/2}=d^{1/2}; \quad (14a)$$

$$c^{1/4}/c^{1/2}=d^{1/4}/d^{1/2}; \quad (14b)$$

$$(c^{3/4} - c^{1/2})/(50 - c^{1/2})=(d^{3/4} - d^{1/2})/(50 - d^{1/2}); \quad (14c)$$

$$c^{1/8}/c^{1/4}=d^{1/8}/d^{1/4}; \quad (14d)$$

$$(c^{7/8} - c^{3/4})/(50 - c^{3/4})=(d^{7/8} - d^{3/4})/(50 - d^{3/4}). \quad (14e)$$

We also determined the shape of the utility function for life duration for each subject and for each

method. The degree of concavity was measured by computing the area under the normalized utility function that results from linear interpolation:

$$50 - 50 \times \left(1/8 d^{1/8} + 3d^{1/4}/16 + 1/4 d^{1/2} + 3d^{3/4}/16 + 1/8 d^{7/8}\right).$$

Finally, to smooth response errors, we fitted exponential utility

$$U(x)=(1 - e^{-rx})/(1 - e^{-r}) \text{ (with } U(x)=x \text{ for } r=0)$$

for both methods, minimizing nonlinear squared distances. We did this both for each individual and for the median data. Exponential utility is widely used and generally gives a good fit.²⁸ An additional advantage is that the estimated exponential coefficient equals the discount rate. Utility is concave if $r > 0$, convex if $r < 0$, and linear if $r = 0$. All tests reported below are 2-sided.

Statistical Analyses of DM v. CEP (Assuming the Same Utility Function U)

Under prospect theory, there is no easy way to compare the c -values of the CE method directly to the d -values of the DM method. Hence, we did not test proportional matches (equation (14)) for prospect theory. The tests of curvature through area under the utility curve and through fitted exponential utility could easily be adapted to prospect theory and were carried out accordingly.

RESULTS

We excluded the data of 3 subjects who did not understand the task or who were not willing to make risky choices about life duration for religious reasons. This left 67 subjects. The consistency tests revealed satisfactory test-retest reliability. The correlations between original and repeated indifference values were 0.75 for CE and 0.74 for DM ($P < 0.05$ in both cases).

Utility Curvature

Figure 4 shows the utility functions based on the (pointwise) median data. All functions were clearly concave, the CEE curve most so. Table 1 shows the implied discount rates based on the median data. The implied rates for the DM and the CEP were lower than for the CEE. Discount rates are commonly found to decline over time.^{29,30} This is indeed what we

Table 1. Implied Discount Rates

	$U^{-1}(1/8)$	$U^{-1}(1/4)$	$U^{-1}(1/2)$	$U^{-1}(3/4)$	$U^{-1}(7/8)$
CEE, %	8.5	7.8	5.3	5.0	4.5
DM, %	2.8	3.5	3.5	3.5	3.9
CEP, %	4.5	5.0	3.8	3.4	2.7

Each entry gives the constant discount rate on $[0, U^{-1}(x)]$ predicting $U^{-1}(x)$. CEE, certainty equivalent with expected utility; CEP, certainty equivalent with prospect theory; DM, direct method.

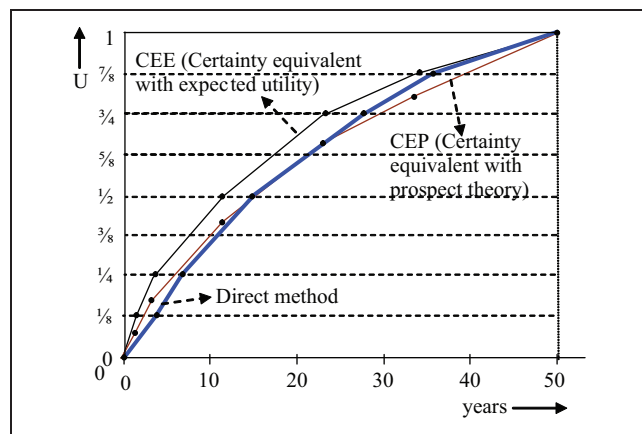


Figure 4. Median utility curves.

observed for the CEE but not for the DM. The discount rates implied by the DM were approximately constant and close to the 3% that is widely used in cost-effectiveness analyses (CEAs).

The estimated exponential coefficients based on the median data (best fitting the whole utility curve) were 0.056 for CEE, 0.036 for the DM, and 0.036 for CEP. The corresponding discount rates are 5.6%, 3.6%, and 3.6%. They differed significantly between DM and CEE (z test, $P < 0.001$) and between CEE and CEP (z test, $P < 0.001$), but not between DM and CEP.

The area under the normalized utility function was highest for the CEE (mean area = 37.03). This area was significantly lower for the DM (mean area = 35.03, $z = -2.296$, $P = 0.02$). For CEP, the area was not significantly different from the DM (mean area = 35.91, $z = -1.399$, $P = 0.16$), but it was obviously lower than the CEE ($z = -7.115$, $P < 0.001$). All areas differed significantly from 25, the case corresponding to linear utility ($P < 0.001$ in all tests).

The median individual discount rates were 6.2% for CEE, 3.5% for DM, and 3.6% for CEP. All estimates differed significantly from zero discounting ($P \leq 0.001$). The CEE estimate differed significantly from the DM estimate ($t = 2.58$, $P = 0.01$) and from

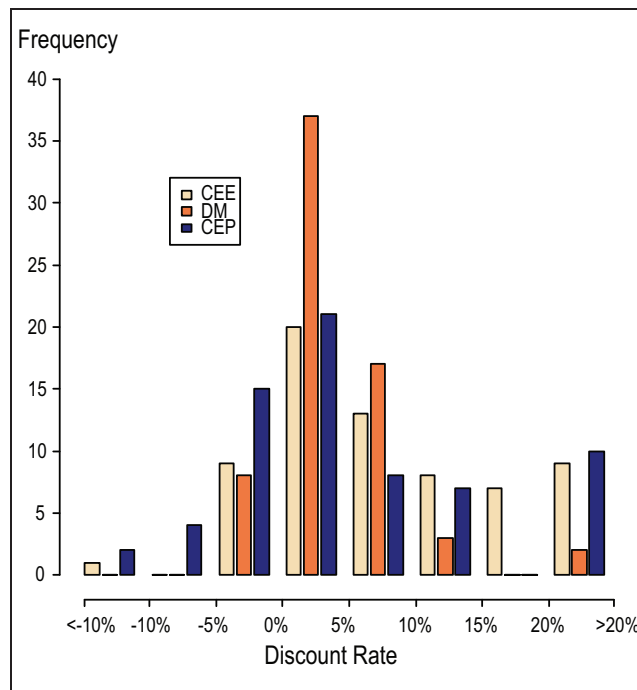


Figure 5. Distribution individual discount rates. CEE, certainty equivalent with expected utility; CEP, certainty equivalent with prospect theory; DM, direct method.

the CEP estimate ($t = 3.81$, $P < 0.001$). The CEP and DM estimates did not differ significantly ($t = 0.63$, $P = 0.53$). Thus, CEP and DM may measure the same utility, but CEE measures something different.

Figure 5 shows the distribution of the individual discount rates. The individual discount rates are more centered for the DM than for the CEE and the CEP. The variance for DM measurements was substantially lower than for the traditional CE method. For example, the variance for $d^{1/2}$ was 43.1, whereas it was 122.8 for $c^{1/2}$. Although the increased variance of the CE method could be interpreted as capturing more individual heterogeneity, we believe that the large increase is primarily due to extra noise. This is confirmed by the better fit of the exponential model

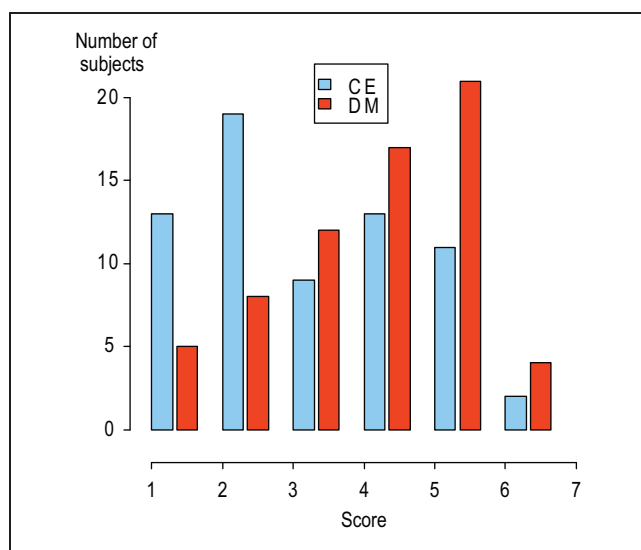


Figure 6. Distribution of understandability scores. CE, certainty equivalent; DM, direct method.

for the DM. The median square root of the estimated variance of the random error was 0.024 for the DM and was significantly lower than the 0.056 for the CEE ($t = 4.87, P < 0.001$) and the 0.061 for the CEP ($t = 5.25, P < 0.001$). The fit of the CEE and of the CEP did not differ significantly ($t = 1.11, P = 0.27$). The relatively poor fit of the CEP indicates that individual prospect theory parameters vary substantially, implying that the method is only reasonable at the group level and does not fit well at the individual level.

Our subjects also considered the DM to be easier than the CE method. Figure 6 shows that the distribution of individual scores for the DM was clearly to the right of the distribution for the CE. The mean scores on our understandability scale (from 1 = worst to 7 = best) were 4.76 for the DM and 3.88 for the CE method. The mode was 6 for the DM and 3 for the CE. The scores differed significantly ($P < 0.001$).

Directly Testing DM v. CEE (Equations 14a–14e)

Equations (14d) (proportional match of $c^{1/8}$ v. $d^{1/8}$) and (14e) (proportional match of $c^{7/8}$ v. $d^{7/8}$) were rejected, with the c -values smaller than the d -values (14d: $z = -2.361, P = 0.02$; 14e: $z = -2.780, P = 0.01$). The other 3 equations (14a, 14b, and 14c) were not rejected (14a: $z = 0.534, P = 0.59$; 14b: $z = -1.511, P = 0.13$; 14c: $z = -1.631, P = 0.10$). We also compared CEE and DM over the whole domain by taking the differences between the proportional matches

for each question and performing a Friedman test. This yielded a significant difference, indicating more concavity for CEE than for DM ($\chi^2 = 11.570; P = 0.04$).

DISCUSSION

DM v. CE utility. Traditional measurements result in more concave (higher) utility than the DM. This discrepancy leaves open the question which of the utilities is more valid. One argument against traditional CEEs is that they are based on expected utility, and there is much empirical evidence against this theory.^{24,31–34} When the CE data are analyzed using the empirically more realistic prospect theory (CEP), they are adjusted downwards and become statistically the same as DM utilities. This suggests that the traditionally analyzed CE measurements (CEE) overestimate utility and, hence, lead to discount rates that are too high and that the DM measurements are more valid.

When choosing between DM and CEP utility, we prefer the former. Although the empirical violations of expected utility have been corrected for by CEP for group averages, individual variations in those violations still generate errors at the individual level, which was reflected in the better fit of the DM for the individual level data. Furthermore, CEP retains the other drawbacks of risky choice.

The absence of significant differences between CEP and DM group average utilities supports the transferability of utility across different domains (risk v. intertemporal). It corroborates similar transfers found between risky utility and other forms of utility if risky utility is analyzed using prospect theory.³⁵

Extraneous devices other than risk. Alternative extraneous devices have been used to measure discounting. The person tradeoff method replaces probabilities by proportions of affected people.^{36–38} Then equity considerations generate distortions much as risk aversion does for risk.^{39,40} The Rawls-Harsanyi veil of ignorance⁴¹ demonstrates the close relationship between the person tradeoff method and the risk approach.

Another approach to measuring discount rates without invoking risky choice is based on inconsistencies in traditional TTO measurements.^{6,42} Such inconsistencies can result if different durations are used and linear utility is erroneously assumed. The resulting inconsistencies have been used to estimate a discount factor.^{43–45} These approaches assume 1-

parameter discounting, usually constant discounting, and do not provide general discount functions. Cairns³ is closest to us. He did not use extraneous devices, and he used similar stimuli. However, he used parametric fitting to measure QALYs.

Violations of the QALY model. The DM was developed for the general QALY model and is valid only to the extent that the QALY model is valid. The central condition underlying the QALY model is an independence condition (Miyamoto and others,⁴⁶ who generalize Pliskin and others⁴⁷), which ensures that the utility of life duration can be measured independently of health quality. Empirical evidence on this condition is limited and mixed, with some violations documented^{48–50} but also some support.^{27,51,52} Other objections against the QALY model have been raised as well—for instance, that it may lead to discrimination of people who have a limited capacity to benefit from health care, such as the disabled. However, no tractable alternative model is available yet. Kahneman and others⁵³ showed that QALY evaluations can be restored if quality of life is sufficiently comprehensive in the sense of incorporating pleasures and pains now felt because of past and future events.

Applications. Our method is applicable not only at the individual level but also in societal CEAs. For the latter, several authors have argued that the social rate of time preference should be based on the diminishing marginal utility for life years.^{54,55} Our method readily provides this information.

Existing algorithms, used to measure health quality in CEAs, are usually based on the TTO (EQ-5D) and the standard gamble (SF6D, HUI). These methods are known to be systematically biased because of discounting, violations of expected utility, and other distortions.⁵⁶ Our method avoids these biases. Our method, as well as the data from our experiment, has been applied in follow-up studies by Attema and Brouwer.^{57,58} These studies used the approach to correct TTO measurements and examine loss aversion and violations of procedural invariance.

CONCLUSION

To evaluate medical interventions, we have to correct the number of resulting life years for 2 factors: 1) quality of life and 2) discounting (utility of life duration). It has traditionally been thought that measuring either of these factors is difficult given that the other factor is also unknown. Complex and distorting extraneous devices have been invoked such as risk or

welfare, or assumptions have been made heuristically. The direct method (DM) resolves these problems. It is surprisingly simple both for subjects and for data analyses. It directly measures the utility of life duration irrespective of quality of life (which may be unknown). With discounting and utility of life duration available, quality of life can readily be measured in a second stage (equation (7)). The whole QALY model can thus be measured in full generality, without using any extraneous device.

Our experiment, implementing the DM, confirms prior expectations. The DM is considerably easier to administer and to understand for subjects than are classical methods. It avoids the aversive and implausible scenarios needed in traditional measurements. DM utilities agree with risky utilities (if the latter are analyzed using prospect theory, a proper descriptive risk theory) on average, but they fit better at the individual level. We conclude that the DM measures the utility of life duration in a more tractable, reliable, and valid manner than methods used before.

REFERENCES

1. van der Pol M, Roux L. Time preference bias in time trade-off. *Eur J Health Econ.* 2005;6:107–11.
2. Bradford WD. The association between individual time preferences and health maintenance habits. *Med Decis Making.* 2010;30:99–112.
3. Cairns J. Health, wealth and time preference. *Project Appraisal.* 1992;7:31–40.
4. Chapman GB, Elstein AS. Valuing the future: temporal discounting of health and money. *Med Decis Making* 1995;15:373–86.
5. Ganiats TG, Carson RT, Hamm RM, et al. Population-based time preferences for future health outcomes. *Med Decis Making.* 2000;20:263–70.
6. Dolan P, Jones-Lee M. The time trade-off: a note on the effect of lifetime reallocation of consumption and discounting. *J Health Econ.* 1997;16:731–9.
7. Martin AJ, Glasziou PP, Simes RJ, Lumley T. A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *Int J Technol Assess Health Care.* 2000;16:137–47.
8. Stalmeier PF, Bezembinder TG, Unic JJ. Proportional heuristics in time tradeoff and conjoint measurement. *Med Decis Making.* 1996;16:36–44.
9. Stiggelbout AM, Kiebert GM, Kievit J, Leer JW, Stoter G, de Haes JC. Utility assessment in cancer patients: adjustment of time trade-off scores for the utility of life years and comparison with standard gamble scores. *Med Decis Making.* 1994;14:82–90.
10. Andersen S, Harrison GW, Lau MI, Rutstrom EE. Eliciting risk and time preferences. *Econometrica.* 2008;76 3:583–618.
11. Richardson J. Cost utility analysis: what should be measured? *Soc Sci Med.* 1994;39:7–21.

12. Samuelson P. A note on the measurement of utility. *Rev Econ Stud.* 1937;4:155–61.
13. Arrow KJ. Alternative approaches to the theory of choice in risk-taking situations. *Econometrica.* 1951;19:404–37.
14. Kahneman D, Snell JS. Predicting utility. In: Hogarth RM ed. *A Tribute to Hillel J. Einhorn.* Chicago: University of Chicago Press; 1990.
15. Savage LJ. *The Foundations of Statistics.* New York: John Wiley; 1954.
16. Varian HR. *Intermediate Microeconomics.* New York: Norton; 1993.
17. Dolan P. The measurement of health-related quality of life for use in resource allocation decisions in health care. In: Culyer AJ, Newhouse JP, eds. *Handbook of Health Economics.* North Holland: Elsevier; 2000.
18. Guerrero AM, Herrero C. A semi-separable utility function for health profiles. *J Health Econ.* 2005;24:33–54.
19. Dolan P. Modelling valuations for health states: the effect of duration. *Health Policy.* 1996;38:189–203.
20. Bostic R, Herrnstein RJ, Luce RD. The effect on the preference-reversal phenomenon of using choice indifference. *J Econ Behav Organ.* 1990;13:193–212.
21. Dolan P. Modeling valuations for EuroQol health states. *Med Care.* 1997;35:1095–108.
22. Schoemaker PJH. The expected utility model: its variants, purposes, evidence and limitations. *J Econ Lit.* 1982;20:529–63.
23. Farquhar PH. Utility assessment methods. *Manage Sci.* 1984;30:1283–300.
24. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica.* 1979;47:263–91.
25. Tversky A, Kahneman D. Advances in prospect theory: cumulative representation of uncertainty. *J Risk Uncertainty.* 1992;5:297–323.
26. Wakker P, Stiggelbout A. Explaining distortions in utility elicitation through the rank-dependent model for risky choices. *Med Decis Making.* 1995;15:180–6.
27. Miyamoto JM, Eraker SA. A multiplicative model of the utility of survival duration and health quality. *J Exp Psychol Gen.* 1988;117:3–20.
28. Wakker PP. Explaining the characteristics of the power (CRRA) utility family. *Health Econ.* 2008;17:1329–44.
29. Cairns J. Discounting in economic evaluation. In: Drummond MF, McGuire A, eds. *Economic Evaluation in Health Care: Merging Theory with Practice.* Oxford, UK: Oxford University Press; 2001.
30. Frederick S, Loewenstein G, O'Donoghue T. Time discounting and time preference: a critical review. *J Econ Lit.* 2002;40:351–401.
31. Oliver A. The internal consistency of the standard gamble: tests after adjusting for prospect theory. *J Health Econ.* 2003;22:659–74.
32. Starmer C. Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *J Econ Lit.* 2000;38:332–82.
33. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. The measurement of patients' values in medicine. *Med Decis Making.* 1982;2:449–62.
34. Rutten-van Molken MPMH, Bakker CH, Van Doorslaer EKA, van der Linden S. Methodological issues of patient utility measurement: experience from two clinical trials. *Med Care.* 1995;33:922–37.
35. Abdellaoui M, Barrios C, Wakker PP. Reconciling introspective utility with revealed preference: experimental arguments based on prospect theory. *J Econometrics.* 2007;138:356–78.
36. Baron J, Wu Z, Brennan DJ, Weeks C, Ubel PA. Analog scale, magnitude estimation, and person trade-off as measures of health utility: biases and their correction. *J Behav Decis Making.* 2001;14:17–34.
37. Nord E. The person-trade-off approach to valuing health care programs. *Med Decis Making.* 1995;15:201–8.
38. Pinto-Prades JL. Is the person trade-off a valid method for allocating health care resources? *Health Econ.* 1997;6:71–81.
39. Abasolo I, Tsuchiya A. Exploring social welfare functions and violation of monotonicity: an example from inequalities in health. *J Health Econ.* 2004;23:313–29.
40. van Doorslaer E, Wagstaff A, Bleichrodt H, et al. Income-related inequalities in health: some international comparisons. *J Health Econ.* 1997;16:93–112.
41. Andersson F, Lyttkens CH. Preferences for equity in health behind a veil of ignorance. *Health Econ.* 1999;8:369–78.
42. Hansen KS, Osterdal LP. Models of quality-adjusted life years when health varies over time: survey and analysis. *J Econ Surveys.* 2006;20:229–55.
43. Dolan P, Gudex C. Time preference, duration and health state valuations. *Health Econ.* 1995;4:289–99.
44. Johannesson M, Pliskin JS, Weinstein MC. A note on QALYs, time tradeoff, and discounting. *Med Decis Making.* 1994;14:188–93.
45. Olsen JA. Persons vs years: Two ways of eliciting implicit weights. *Health Econ.* 1994;3:39–46.
46. Miyamoto JM, Wakker PP, Bleichrodt H, Peters HJM. The zero-condition: a simplifying assumption in QALY measurement and multiattribute utility. *Manage Sci.* 1998;44:839–49.
47. Pliskin JS, Shepard D, Weinstein MC. Utility functions for life years and health status. *Oper Res.* 1980;28:206–24.
48. McNeil BJ, Weichselbaum R, Pauker SG. Speech and survival: Tradeoffs between quality and quantity of life in laryngeal cancer. *N Engl J Med.* 1981;305:982–7.
49. Sutherland HJ, Llewellyn-Thomas H, Boyd NF, Till JE. Attitudes toward quality of survival. *Med Decis Making.* 1982;2:299–309.
50. Dolan P, Stalmeier PFM. The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. *J Health Econ.* 2003;22:445–58.
51. Miyamoto J, Eraker SA. Parameter estimates for a QALY utility model. *Med Decis Making.* 1985;5:191–213.
52. Bleichrodt H, Pinto JL. The validity of QALYs under non-expected utility. *Econ J.* 2005;115:533–50.
53. Kahneman D, Wakker PP, Sarin R. Back to Bentham? Explorations of experienced utility. *Q J Econ.* 1997;112:375–405.
54. van Hout BA. Discounting costs and effects: a reconsideration. *Health Econ.* 1998;7:581–94.

55. Gravelle H, Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Health Econ.* 2001;10:587–99.
56. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ.* 2002;11:447–56.
57. Attema AE, Brouwer WBF. Can we fix it? Yes we can! But what? A new test of procedural invariance in TTO-measurement. *Health Econ.* 2008;17:877–85.
58. Attema AE, Brouwer WBF. The correction of TTO-scores for utility curvature using a risk-free utility elicitation method. *J Health Econ.* 2009;28:234–43.
59. van Osch SM, Wakker PP, van den Hout WB, Stiggelbout AM. Correcting biases in standard gamble and time tradeoff utilities. *Med Decis Making.* 2004;24:511–7.
60. Stevens SS. A scale for the measurement of a psychological magnitude: loudness. *Psychol Rev.* 1936;43:405–16.