VALUE IN HEALTH 21 (2018) 993–1001

Available online at www.sciencedirect.com

ScienceDirect

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

Preference-Based Assessments

Advocating a Paradigm Shift in Health-State Valuations: The Estimation of Time-Preference Corrected QALY Tariffs

Marcel F. Jonker, PhD[1,2,3,*], Bas Donkers, PhD[2,4], Esther W. de Bekker-Grob, PhD[2,3], Elly A. Stolk, PhD[2,5]

[1]Duke Clinical Research Institute, Duke University, Durham, NC, USA; [2]Erasmus Choice Modelling Centre, Erasmus University Rotterdam, The Netherlands; [3]Erasmus School of Health Policy & Management, Erasmus University Rotterdam, The Netherlands; [4]Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands; [5]EuroQol Research Foundation, Rotterdam, The Netherlands

ABSTRACT

Background: Despite evidence of nonproportional trade-offs in time trade-off exercises and the explicit incorporation of exponential discounting in health technology assessment calculations, quality-adjusted life-year (QALY) tariffs are currently still established under the assumption of linear time preferences. **Objectives:** The aim of this study was to introduce a general method of accommodating for nonlinear time preferences in discrete choice experiment (DCE) duration studies and to evaluate its impact on estimated QALY tariffs. **Methods:** A parsimonious utility function is proposed that accommodates any discounting function and preserves linear time preferences as a special case. Based on an efficient DCE design and 1775 respondents from a nationally representative scientific household panel, preferences and QALY tariffs for the Dutch SF-6D were estimated while accommodating for nonlinear time preferences via exponential and hyperbolic discounting functions. **Results:** When the discount rate was estimated directly, we found strong evidence of nonlinear time preferences (with an exponential and hyperbolic discount rate of 5.7% and 16.5%, respectively). When the discount rate was estimated as a function of health state severity, we found that years lived in better health states are discounted less than years lived in impaired health states. Finally, the best statistical fit was obtained when using a hyperbolic discount function, which resulted in smaller QALY decrements and fewer health states classified as worse than immediate death. **Conclusions:** Our results highlight the relevance and even necessity of a paradigm shift in health valuation studies in favor of time-preference corrected QALY tariffs, with potentially important implications for health technology assessment calculations and regulatory decisions.

Keywords: discrete choice experiment, health state valuation, SF-6D, QALY.

Introduction

Economic evaluations in health care are often based on quality-adjusted life-years (QALYs), providing an intuitive and straightforward approach to combine quality and duration of life into a single index [1,2]. Traditionally, the required preference weights for QALY calculations were derived using time trade-off (TTO) and standard gamble techniques. These valuation techniques required respondents to iterate toward a point of indifference between two health outcomes. Currently, ordinal valuation tasks and, in particular, discrete choice experiments (DCEs) are increasingly considered as reliable alternatives to estimate QALY tariffs [3–8]. DCEs have a solid theoretical foundation [9–11], require less abstract reasoning, more closely resemble real-life decisions, and can typically be administered online (without resulting in different QALY tariffs compared with pen-and-paper surveys [12,13]).

An important issue regarding the use of DCEs in health state valuations is that DCEs only provide information about the relative value of health states, ranging from the best to the worst included health state, and on a latent utility scale that is different in each estimation. In contrast, a fundamental requirement of the QALY model is that estimated health-state values have to reflect the relative desirability of health states on a scale anchored at 0 and 1, with 1 equivalent to perfect health and 0 to immediate death. Accordingly, a procedure is required to transform the latent DCE utility estimates to QALY health-state values [14–16].

The theoretically most appealing scaling method is the DCE duration approach, which is based on the inclusion of a duration-of-life attribute in the choice tasks [14]. This format resembles traditional TTO in the sense that health-state values are derived from trade-offs between quantity and quality of life, yet without

Conflict of interest: The authors have indicated that they have no conflicts of interest with regard to the content of this article.

* Address correspondence to: Marcel F. Jonker, Duke Clinical Research Institute, 2400 Pratt Street, Durham, NC 27705, United States. E-mail: marcel@mfjonker.com.

1098-3015$36.00 – see front matter Copyright © 2018, ISPOR–The Professional Society for Health Economics and Outcomes Research. Published by Elsevier Inc.

https://doi.org/10.1016/j.jval.2018.01.016

requiring respondents to iterate toward a point of indifference. However, DCE duration data have, thus far, typically been analyzed under the assumption of linear time preferences. This makes sense because linear time preferences are congruent with the analysis of traditional TTO data, adhere to the QALY assumption of constant proportional trade-offs, and keep the construction of the experimental design relatively straightforward. However, this assumption ignores the fact that linear time preferences are hardly ever observed in human decision making [17] and the existing evidence on the not-so-constant proportional trade-offs in TTO [18].

Accordingly, this article aims to introduce a general method of accommodating for nonlinear time preferences in DCE duration studies and to evaluate its empirical relevance by using a DCE duration design that is specifically optimized to measure deviations from linear time preferences. Combined with an empirical application in a large, online scientific panel, this allows for the measurement of nonlinear time preferences with substantial statistical power. On the basis of the presented evidence, we intend to establish the presence of nonlinear time preferences and initiate a discussion about the desirability and necessity of time-preference corrected QALY tariffs, which could have serious implications for cost–utility comparisons in health technology assessment (HTA) applications.

Methods

Conceptual Model

In the DCE duration literature, health-state preferences are derived by using choice tasks that require respondents to choose between living longer in a relatively bad health states or living for a shorter period in a relatively good health state. Here, the utility that respondent *i* obtains from alternative *j* in choice task *t* is usually the product of the health state characteristics (X_{ijt}) and accompanying preference parameters (β_i) multiplied by the number of life years, T_{ijt}, that is:

$$U_{ijt} = \left(\beta_{i.} X_{ijt} \right)_{.} T_{ijt} + \varepsilon_{ijt} \tag{1}$$

with the error term, ε_{ijt}, reflecting the impact of unobserved characteristics on respondents' choices [19]. Note that this specification is quite flexible by not imposing any structure on the functional form between brackets (i.e., $\beta_{i.} X_{ijt}$). The multiplicative structure also conveniently anchors immediate death at 0 on the utility scale. However, equation (1) is simultaneously very restrictive by imposing linear time preferences (i.e., each additional life-year receives the same utility). We aim to relax this assumption by introducing a class of utility functions that allows for the accommodation of nonlinear time preferences while preserving the option of linear time preferences as a special case. In our suggested approach, obtained utilities are still defined as the product of quality and quantity of life, but with the latter included as the net present value (NPV) of the number of life years T. This results in the following specification:

$$U_{ijt} = \left(\beta_{i.} X_{ijt} \right)_{.} NPV_{ijt} + \varepsilon_{ijt} \tag{2}$$

where the NPV_{ijt} is defined as:

$$NPV_{ijt} = \sum PV_{(r,s)},\ s = 1 \text{to} T_{ijt} \tag{3}$$

One of the major benefits of the proposed methodology is that any discounting function can be used to calculate the present value (PV) of future life-years *T*; equation (3) only states that the NPV is the sum of the present values of all future years lived and thus depends on the type of discount function, the discount rate, *r*, and, of course, on the number of life years, *T*.

Empirical Application: Deriving a Value Set for the Dutch SF-6D

The impact of accommodating for nonlinear time preferences in health-state valuations is evaluated in an empirical application based on the Dutch SF-6D. The SF-6D is a six-dimensional health-state classification that includes the dimensions physical functioning, role limitations, social functioning, pain, mental health, and vitality. It was developed by Brazier et al. [20] and is directly based on a subset of questions from either the SF-36 or SF-12, which are well-validated and commonly used health measurement instruments. SF-6D health states are defined by selecting one level from each dimension. Depending on the specific version of the SF-6D, each dimension has either four to six levels (the SF-36 version [20]), or three to five levels (the SF-12 version [21]). In this study, the SF-12 version was used (Table 1). This version has the practical advantage of having fewer levels per health dimension, which made it easier for respondents to differentiate between the various levels.

The Discrete Choice Experiment

Health states in the DCE were hence defined in terms of the attributes and levels (e.g., PF1, PAIN2) as described in Table 1 and combined with a duration of life attribute (Fig. 1).

To make it easier for respondents to evaluate SF-6D health states, some of the regularly recurring text in the Role Limitations and Pain attributes was included in the attribute description rather than in the level description. Additionally, to reduce the task complexity of the DCE duration format and to create a DCE design that allowed for the measurement of nonlinear time preferences, three types of constraints were imposed. The first type was aimed at improving the internal consistency of health states, which was done by blocking the occurrence of several combinations of health state attributes in the DCE design. More specifically, PAIN5, MH5, and VIT5 were not allowed to be combined with RL1, and, after evaluation of the comments received from respondents in the pilot study, MH5 and VIT5 were not allowed to be combined with SF1. These constraints avoided evocation of health states that respondents might consider internally conflicting, such as pain that interferes extremely with one's normal work but without any role limitations. It should be noted that the number of constraints was deliberately minimized to avoid too much impact on the statistical efficiency of the DCE design.

The second set of constraints resulted in the "matched pairwise" choice format that was previously introduced by Jonker et al. [13]. An example of matched pairwise choice task is shown in Figure 1. First, the respondents were asked to indicate which health profile (A or B) they preferred. Here, the duration of life was always constrained to be identical for both alternatives, implying that the choice was only determined by the relative attractiveness of the health states. Three of the six attributes were always constrained to have the same value, which further simplified the choice tasks [22–24]. In the subsequent matched choice task, respondents were asked which health profile (B or C) they preferred. Option B here was always the same as in the first task, and option C always represented perfect health with a shorter duration of life. Together with the "traffic-light" color coding system that was used (with green assigned to the best attribute level, orange to the middle attribute level, red to the worst attribute level, and intermediate colors to intermediate levels), these constraints were implemented to reduce the complexity of the DCE duration format.

The third and final set of constraints pertained to the duration levels, which were restricted to integer values (i.e., whole years) for values equal to or larger than 1 year and restricted to 2, 3, 6, or

Table 1 – The SF-6D (SF-12 version).

Physical Functioning		**Pain**	
PF1	Your health does not limit you in moderate activities.	PAIN1	You have pain that does not interfere with your normal work (both outside the home and housework) at all.
PF2	Your health limits you a little in moderate activities.	PAIN2	You have pain that interferes with your normal work (both outside the home and housework) a little bit.
PF3	Your health limits you a lot in moderate activities.	PAIN3	You have pain that interferes with your normal work (both outside the home and housework) moderately.
		PAIN4	You have pain that interferes with your normal work (both outside the home and housework) quite a bit.
		PAIN5	You have pain that interferes with your normal work (both outside the home and housework) extremely.
Role Limitations		**Mental Health**	
RL1	You have no problems with your work or other regular daily activities as a result of your physical health or any emotional problems.	MH1	You feel downhearted and low none of the time.
RL2	You are limited in the kind of work or other activities as a result of your physical health,	MH2	You feel downhearted and low a little of the time,
RL3	You accomplish less than you would like as a result of emotional problems.	MH3	You feel downhearted and low some of the time,
RL4	You are limited in the kind of work or other activities as a result of your physical health and accomplish less than you would like as a result of emotional problems.	MH4	You feel downhearted and low most of the time.
		MH5	You feel downhearted and low all of the time.
Social Functioning		**Vitality**	
SF1	Your health limits your social activities none of the time.	VIT1	You have a lot of energy all of the time.
SF2	Your health limits your social activities a little of the time.	VIT2	You have a lot of energy most of the time.
SF3	Your health limits your social activities some of the time.	VIT3	You have a lot of energy some of the time.
SF4	Your health limits your social activities most of the time.	VIT4	You have a lot of energy a little of the time.
SF5	Your health limits your social activities all of the time.	VIT5	You have a lot of energy none of the time.

From Brazier et al. [21].

9 months otherwise. To ensure sufficient variation in the duration levels, not all choice tasks were optimized by using a standard 10-year time frame. Two choice tasks (in each subdesign) were optimized with durations for options A and B in the range of 3 to 8 years, and two in the range of 16 to 20 years. Option C (i.e., perfect health) necessarily had a smaller duration than option B to avoid strictly dominant choice tasks. Hence, all duration levels were selected as an integral part of the design optimization algorithm (see the subsequent section) in conjunction with the selection of health states. Note that Tables A1 and A2 in eAppendix A contain an overview of the selected duration levels in the DCE designs, with all durations adhering to the described constraints.

Optimization of the DCE Design

To be able to implement the imposed constraints while ensuring sufficient statistical efficiency, the DCE design was constructed by using a Bayesian efficient design algorithm [25]. Given that suitable priors were unavailable, a two-step procedure was used. An initial DCE design was constructed by using prior information from a conditional logistic regression based on SF-6D rank data obtained from UK respondents [26]. These prior data were not on the appropriate scale but did reflect the ordinal structure of the SF-6D attributes. The resulting DCE design was administered to a pilot sample (N = 569 completes) to obtain more reliable and directly relevant estimates. These estimates were subsequently used to improve the main DCE design that was used to collect data for all remaining respondents.

Further robustness and additional variation in the attribute levels was introduced by simultaneously optimizing eight different DCE subdesigns. The implementation of such a heterogeneous DCE design comes at no additional cost to respondents because each respondent is only randomly assigned to one of the subdesigns. However, the use of several subdesigns increases the robustness and efficiency of the overall design [27].

The heterogeneous design optimizations were implemented in Matlab and distributed on the Dutch Life Science Grid [28]. Health states and durations of life were selected to minimize the weighted average Bayesian D-efficiency of the DCE design, with one third of the weight assigned to the combined (i.e., population) D-efficiency and two thirds of the weight assigned to the individual D-efficiencies of the subdesigns. For the pilot study, 250 Latin hypercube draws were used and for the main design 25,000 draws. Both the pilot and main design were optimized for an additive main-effects multinomial logit model, which accommodated both linear and nonlinear time preferences. Each subdesign consisted of 14 matched pairwise choice tasks, resulting in an overall DCE design, with 112 matched pairwise choice tasks and a total of 224 paired comparisons.

Fig. 1 – Visual presentation of the matched pairwise choice tasks. A, **Type I choice task.** B, **Matched type II choice task.**

Survey Administration and Instrument

The DCE was administered in a large, nationally representative, scientific Dutch online household panel: the Longitudinal Internet Studies for the Social Sciences (LISS) [29]. This panel is subsidized by the Dutch government, with the goal of improving the climate for innovation in The Netherlands and providing a platform for researchers to pilot innovative methodologies. The LISS consists of 5000 households, comprising roughly 8000 individuals randomly drawn from the population register of Statistics Netherlands. Panel members are invited to complete online questionnaires every month for about 15 to 30 minutes in total, for which they receive a small financial compensation.

The survey started with a brief introduction, followed by a question in which respondents were asked to rate their current health in terms of the attributes and levels of the SF-6D. This procedure familiarized respondents with the format of the health states that was used in the DCE. Subsequently, two warm-up questions, with a careful explanation of the layout and structure of the choice tasks, were included. Then, the actual set of 14 matched pairwise choice tasks were shown, and at the end of the survey, several cognitive debriefing questions were included (see eAppendix C for an overview of the full instrument). Note that respondents received a random order of the SF-6D health attributes, although for each respondent, the same order of attributes was used in the self-rating, the warm-up questions, and the 14 matched pairwise choice tasks. This minimized potential attribute-order effects without requiring respondents to reorient to a new attribute order after each new choice task [30,31].

Statistical Analyses

Utility decrements and QALY estimates for the SF-6D were obtained by using four different statistical models. All models were based on an additive main effects specification without interactions between attribute levels except for those between the quantity and quality of life.

Specification 1

In the first model specification, linear time preferences were imposed (cf. equation 1).

Specification 2

In the second model specification, nonlinear time preferences were accommodated for using the standard exponential discounting function, defined as PV(r,s) = EXP (–r * s) [32]. After working out the summation, equation (3) combined with an exponential PV simplifies into a standard annuity with exponential discount rate, *r*;

$$\mathrm{NPV}_{itj} = \mathrm{T}_{itj} \qquad \textit{if } \mathrm{r}=0$$

$$= (1-\exp(-\mathrm{rT}_{itj}))/(\exp(\mathrm{r})-1) \qquad \textit{if } \mathrm{r}\neq 0 \tag{4}$$

As shown, when the discount rate equals zero, all future life-years receive unit weight and the NPV equals T; the latter implies that equation (1) indeed forms a special case of the more general model specification based on equations 2 and 4.

Specification 3

In the third model specification, nonlinear time preferences were again accommodated for using the exponential discounting function. However, instead of estimating the discount rate, *r*, directly, the discount rate was decomposed as:

$$\mathrm{r} = \gamma_1 + \gamma_2 * (\mu.\mathrm{X}_{itj}) \tag{5}$$

with ($\mu.\mathrm{X}_{itj}$) reflecting the health state severity and γ denoting the regression coefficients to be estimated. This specification reflected that the discount rate may depend on the severity of the health states under evaluation. For example, living an additional 10 years in perfect health is likely more attractive (and consequently less severely discounted) compared with living an 10 additional years in a relatively poor health state. Note that the DCE design was not specifically optimized to identify this specification; hence, more elaborate models with nonlinear and/or discontinuous discount rates were not investigated.

Specification 4

In the fourth and final model specification, nonlinear time preferences were accommodated for by using the most commonly used alternative to exponential discounting, which is hyperbolic discounting defined as PV(r,s) = 1/(1 + r*s) [33]. Again, working out the summation of 1 to *T* in equation (3), the NPV for the hyperbolic function was implemented as:

$$\mathrm{NPV}_{itj} = \mathrm{T}_{itj} \qquad \textit{if } \mathrm{r}=0$$

$$= (1/r) * [\psi(1+1/r+\mathrm{T}_{itj}) - \psi(1+1/r)] \qquad \textit{if } r>0 \tag{6}$$

with ψ denoting the di-gamma function. As shown, linearity was preserved as a special case, but negative values of the hyperbolic discount rate were not supported. Note that the summation in equation (3) could also be calculated recursively, which would allow for negative hyperbolic discount rates at the expense of additional computation time.

All four model specification assumed IID Gumbel distributed error terms and respondent-specific β-parameters that were multivariate normal distributed with population mean, μ, and covariance matrix, Σ:

$$\beta_i \sim \text{multivariate Normal } (\mu, \Sigma). \tag{7}$$

This implies a standard mixed logit specification. Furthermore, with all SF-6D attributes dummy coded and with the best levels used as reference categories, the models' intercepts captured the utility of perfect health and thereby the highest possible utility. The latter permitted the calculation of standardized health-state decrements on the QALY scale by dividing all elements in the vector μ by the intercept, that is, the first element of μ:

$$\mathrm{QALY}_{tariff} = \mu/\mu_{(1)}. \tag{8}$$

All model specifications were programmed in the BUGS language and fitted using OpenBUGS utilizing Bayesian Markov chain Monte Carlo (MCMC) methods. This involved selecting prior densities for the unknown model parameters and updating those densities with the likelihood of the observed data. Uninformative normal priors (i.e., with means of zero and standard deviations of 10) were assigned to μ, *r*, and γ and a Wishart prior with an identity scale matrix and 21 degrees of freedom to Σ. Standard Gibbs updates were used to update μ and Σ, antithetic Metropolis-within-Gibbs update steps to update β and γ, and slice sampling update steps were used to update *r*. All estimations used 15,000 iterations to let three MCMC chains converge and 45,000 iterations to reliably approximate the posterior distributions. Convergence was evaluated on the basis of a visual inspection of the chains and diagnostics proposed by Geweke [34]. The statistical fit of the models was evaluated based on the Watanabe-Akaike information criterion (WAIC) [35] and individual-level McFadden R^2 statistics [36].

Results

A sample of 2579 panel members ages 18 years or greater was invited to participate in the survey, resulting in 1775 completes (69%), 25 dropouts (<1%), and a non-response rate of 30%. The non-response rate was similar to those typically observed in the LISS panel and the number of dropouts was, given the length of the survey and complexity of the choice tasks, smaller than a priori anticipated. The evaluation questions at the end of the survey confirmed that the choice tasks were manageable for almost all respondents: 6% of the included respondents considered the choice tasks "very unclear" and 10% "unclear." The remaining 84% considered the questions "neutral," "clear," or "very clear." Additionally, only 47 of the 1775 respondents (2.6%) had an individual-level R^2 statistic that was insignificantly different from zero (i.e., indistinguishable from random choice behavior). Because the exclusion of these respondents had a negligible impact on the estimation results, no respondents were dropped from the sample. Table D1 in eAppendix D contains the descriptive statistics of all included respondents.

Table 2 presents the results of the four model specifications on the latent utility scale. In all specifications, the estimates were logically consistent, had the correct sign, and were statistically significant (i.e., have small standard errors and 95% credible intervals that did not contain zero). The pain and mental health domains were considered most important by respondents and had approximately twice the weight of that of the other SF-6D domains. All nonlinear time preference specifications provided an improvement in WAIC compared with the first specification that imposed linearity. In the second specification, clear evidence of nonlinear time preferences was found with an estimate of the exponential discount rate of 5.7%. The third specification additionally confirmed that the value of the discount rate depends on the severity of the health state under evaluation. More specifically, with the estimates of γ_1 of 0.063, γ_2 of –0.005 and with the latent utility values ranging from 1.70 to –0.86 (i.e., the latent utility values of the best and worst possible health states), the discount rate ranged from 5.4% to 6.7%. The fourth specification performed best in terms of WAIC and confirmed that respondents were more likely to make choices based on a hyperbolic than exponential discounting function.

Figure 2 and Table 3 contain the results of the model specifications on the QALY scale, which allows for the obtained parameter estimates to be directly compared across the four specifications. As shown, the QALY tariff of the standard model had stronger (i.e., more negative) decrements than those of the models that take nonlinear time preferences into account. The difference was about 5% to 10%. Whether the discount rate was estimated directly, as in specification 2, or was allowed to depend on the severity of the health states under evaluation, as in specification 3, had a negligible impact on the QALY tariff. In contrast, the difference between exponential and hyperbolic discounting did have an impact on the resulting QALY tariff, with the latter resulting in slightly smaller decrements.

Discussion

This paper has introduced a general method of accommodating for nonlinear time preferences in DCE duration studies and has evaluated its empirical relevance in an empirical example. Based on a nationally representative sample of 1775 participants and an efficient DCE duration design, a QALY tariff for the Dutch SF-6D was estimated under the assumption of linear as well as nonlinear time preferences. Clear evidence of nonlinear time

Table 2 – Estimates of μ, r, and γ on the latent utility scale.*

	1. Linear time preferences (no discounting)	2. Nonlinear time preferences (exponential)	3. Nonlinear time preferences (severity dependent)	4. Nonlinear time preferences (hyperbolic)
Perfect health	1.25 (0.034)	1.74 (0.055)	1.70 (0.061)	2.30 (0.113)
PF2	–0.07 (0.004)	–0.08 (0.005)	–0.08 (0.005)	–0.10 (0.007)
PF3	–0.26 (0.008)	–0.34 (0.010)	–0.33 (0.012)	–0.42 (0.019)
RL2	–0.11 (0.005)	–0.13 (0.006)	–0.13 (0.006)	–0.16 (0.008)
RL3	–0.14 (0.005)	–0.16 (0.006)	–0.16 (0.007)	–0.19 (0.009)
RL4	–0.24 (0.007)	–0.29 (0.009)	–0.29 (0.009)	–0.36 (0.014)
SF2	–0.04 (0.003)	–0.03 (0.004)	–0.03 (0.004)	–0.04 (0.006)
SF3	–0.07 (0.004)	–0.08 (0.005)	–0.07 (0.006)	–0.09 (0.007)
SF4	–0.15 (0.005)	–0.19 (0.007)	–0.19 (0.008)	–0.24 (0.012)
SF5	–0.25 (0.007)	–0.32 (0.009)	–0.32 (0.011)	–0.40 (0.017)
PAIN2	–0.08 (0.004)	–0.09 (0.005)	–0.09 (0.004)	–0.11 (0.006)
PAIN3	–0.22 (0.006)	–0.27 (0.008)	–0.26 (0.008)	–0.34 (0.014)
PAIN4	–0.31 (0.008)	–0.40 (0.011)	–0.39 (0.013)	–0.51 (0.022)
PAIN5	–0.53 (0.013)	–0.69 (0.018)	–0.67 (0.021)	–0.87 (0.036)
MH2	–0.07 (0.004)	–0.07 (0.005)	–0.07 (0.005)	–0.09 (0.006)
MH3	–0.13 (0.005)	–0.16 (0.006)	–0.15 (0.006)	–0.19 (0.009)
MH4	–0.31 (0.008)	–0.40 (0.011)	–0.39 (0.013)	–0.50 (0.021)
MH5	–0.48 (0.012)	–0.63 (0.016)	–0.62 (0.019)	–0.80 (0.032)
VIT2	–0.05 (0.004)	–0.04 (0.005)	–0.04 (0.005)	–0.05 (0.006)
VIT3	–0.10 (0.004)	–0.12 (0.005)	–0.12 (0.005)	–0.15 (0.008)
VIT4	–0.19 (0.006)	–0.23 (0.007)	–0.23 (0.008)	–0.29 (0.013)
VIT5	–0.27 (0.007)	–0.34 (0.010)	–0.34 (0.011)	–0.44 (0.019)
r		0.057 (0.003)		0.165 (0.016)
γ_1			0.063 (0.004)	
γ_2			–0.005 (0.002)	
WAIC *(lower = better)*	47,296	47,098	47,182	47,005
N	1775	1775	1775	1775

* Mean posterior estimates with standard deviations in parentheses.

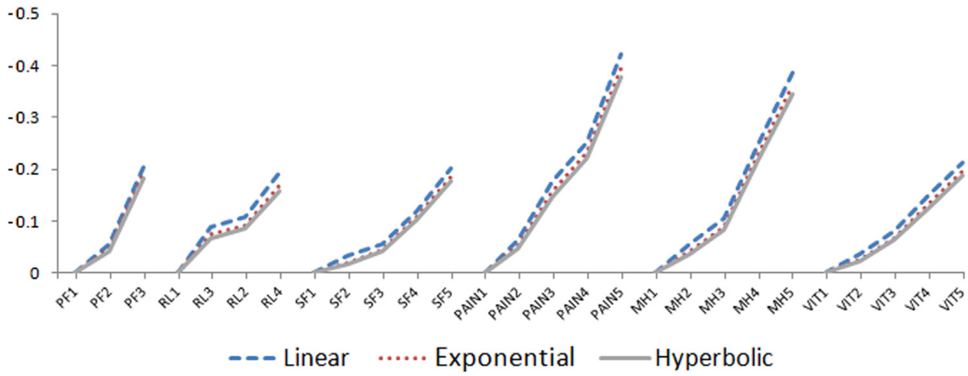

Fig. 2 – SF-6D estimates on the QALY scale.

preferences was established, with estimated exponential and hyperbolic discount rates of 5.7% and 16.5%, respectively. Compared with exponential discounting, the analyses based on hyperbolic discounting provided a better fit to the data and thus resulted in a more reliable separation between time and health-state preferences.

Of course, QALY tariffs are mainly used in HTA, where exponential discounting on health effects is already universally applied. Accordingly, QALY tariffs as used in HTA applications are actually already assumed to be free (or statistically purged) from confounding time preferences, which makes the time-preference corrected QALY tariffs as proposed in this article preferable from a HTA perspective. Furthermore, the discount rates as used in HTA are typically unrelated to empirically observed discount rates [37,38], which means that estimated empirical discount rates (which depend on the discounting function used) are currently less important than the quality of the time-preference correction. Of course, the fact that empirical discount rates are not considered in HTA applications may be related to the historical absence of reliable empirical discount rates. From this perspective, it does make sense to report empirical exponential discount rates. However, the estimation of empirical discount rates will not be of immediate relevance for HTA; discount rates in HTA have to at least partially reflect societal concerns, and, more importantly,

Table 3 – SF-6D estimates on the QALY scale.*

	1. Linear time preferences (no discounting)	2. Nonlinear time preferences (exponential)	3. Nonlinear time preferences (severity dependent)	4. Nonlinear time preferences (hyperbolic)
Perfect health	1.00	1.00	1.00	1.00
PF2	–0.05 (0.003)	–0.05 (0.003)	–0.05 (0.003)	–0.04 (0.003)
PF3	–0.21 (0.006)	–0.19 (0.006)	–0.19 (0.006)	–0.18 (0.006)
RL2	–0.09 (0.003)	–0.07 (0.003)	–0.07 (0.003)	–0.07 (0.003)
RL3	–0.11 (0.004)	–0.09 (0.004)	–0.09 (0.004)	–0.08 (0.004)
RL4	–0.19 (0.005)	–0.17 (0.005)	–0.17 (0.005)	–0.16 (0.005)
SF2	–0.03 (0.003)	–0.02 (0.003)	–0.02 (0.003)	–0.02 (0.003)
SF3	–0.05 (0.003)	–0.04 (0.003)	–0.04 (0.003)	–0.04 (0.003)
SF4	–0.12 (0.004)	–0.11 (0.004)	–0.11 (0.004)	–0.10 (0.004)
SF5	–0.20 (0.006)	–0.18 (0.006)	–0.19 (0.006)	–0.18 (0.006)
PAIN2	–0.06 (0.003)	–0.05 (0.003)	–0.05 (0.003)	–0.05 (0.003)
PAIN3	–0.18 (0.005)	–0.16 (0.005)	–0.16 (0.005)	–0.15 (0.005)
PAIN4	–0.25 (0.007)	–0.23 (0.007)	–0.23 (0.007)	–0.22 (0.007)
PAIN5	–0.42 (0.011)	–0.40 (0.011)	–0.40 (0.011)	–0.38 (0.011)
MH2	–0.06 (0.003)	–0.04 (0.003)	–0.04 (0.003)	–0.04 (0.003)
MH3	–0.11 (0.004)	–0.09 (0.004)	–0.09 (0.004)	–0.08 (0.004)
MH4	–0.25 (0.007)	–0.23 (0.007)	–0.23 (0.007)	–0.22 (0.007)
MH5	–0.39 (0.010)	–0.36 (0.010)	–0.36 (0.010)	–0.35 (0.01)
VIT2	–0.04 (0.003)	–0.02 (0.003)	–0.02 (0.003)	–0.02 (0.003)
VIT3	–0.08 (0.003)	–0.07 (0.003)	–0.07 (0.003)	–0.06 (0.003)
VIT4	–0.15 (0.005)	–0.13 (0.005)	–0.13 (0.005)	–0.13 (0.004)
VIT5	–0.21 (0.006)	–0.20 (0.006)	–0.20 (0.006)	–0.19 (0.006)
% SF-6D health states worse than immediate death	13.1%	7.1%	7.2%	4.7%
Worst possible SF-6D health state (3-4-5-5-5-5)	–0.62	–0.50	–0.51	–0.44

* Mean posterior estimates with standard deviations in parentheses.

they are often simply set equal to the discount rate on costs to ensure time-consistent decisions [38].

At the same time, the disentanglement of health-state and time preferences in DCE duration studies does have a direct impact on the results obtained from HTA studies. As shown in Table 3, when assuming linear time preferences, 13.1% of all SF-6D health states are classified as worse than immediate death, which reduces to approximately 7.1% and 4.7% when assuming exponential and hyperbolic discounting, respectively. These percentages directly correspond to the smaller utility decrements and a narrower QALY range as obtained when accommodating for nonlinear time preferences and, in turn, result in more conservative estimates of costs per QALY in HTA studies.

Interestingly, the assumption of linear time preferences offers at least a partial explanation for the relatively low QALY values that have thus far been reported in DCE duration studies compared with studies based on TTO [39]. Even though accommodating nonlinear time preferences is expected to affect both TTO and DCE, it is likely to have a stronger impact on DCEs. In TTO tasks, it is directly observed whether health states are considered better or worse than immediate death. In contrast, in DCE duration studies, health states worse than immediate death are identified by extrapolating observed trade-offs to the hypothetical case of zero duration of life (i.e., immediate death). The assumptions that constitute those extrapolations have a direct impact on the fraction of health states better or worse than death. The stronger the curvature in the time preferences, the fewer the health states that will be classified as worse than immediate death, resulting in proportionally smaller QALY decrements. Accordingly, the presented results suggest that adequate modeling of time preferences is an essential requirement for the validity of DCE duration results and that caution is warranted when interpreting results obtained in studies that imposed linear preferences.

Taking a somewhat broader perspective, it is clear that DCE duration approaches are expanding our understanding of time preferences in the QALY framework. From a historical point of view, there has been considerable evidence of nonproportional trade-offs in TTO studies, albeit often based on relatively small samples and sometimes with inconclusive results [19]. With DCE duration methods, it has become possible to collect larger data sets at lower costs and to include more choice tasks per respondent than feasible with TTO or standard gamble valuation methods. This has provided the possibility to unequivocally establish the relevance and merit of accommodating for nonlinear time preferences in health-state valuations. Additionally, as shown in this article, the estimation of models that explicitly accommodate nonlinear time preferences can be relatively straightforward and doing so avoids potentially severely downwards biased QALY tariffs. The accommodation of nonlinear time preferences also avoids the current inconsistency between the estimation of QALY tariffs under the assumption of linear time preferences and their application in HTA under the explicit acknowledgement of nonlinear time preferences.

Accordingly, it seems logical, as a field, to transition toward health state valuation estimations that explicitly accommodate nonlinear time preferences. Of course, we acknowledge that further research is required to confirm the empirical relevance of nonlinear time preferences in other data sets and to compare the performance of alternative discounting functions [40]. However, the suggested approach to incorporating nonlinear time preferences is very general; it can handle any discount function and accommodate linear time preferences as a special case, which makes the transition toward incorporating nonlinear time preferences in the estimation of QALY tariffs an important incremental as well as feasible step forward.

Funding

The authors have no other financial relationships to disclose.

Acknowledgements

This work relied on LISS data that were collected by CentERdata (Tilburg University, The Netherlands) through its MESS project, and has made use of supercomputer facilities that were sponsored by the National Computing Facilities Foundation (NCF), both of which are funded by The Netherlands Organization for Scientific Research (NWO). The authors would like to thank SURFsara and Prof. Dr. M.C.J. Bliemer for technical assistance with the design optimizations, and Prof. Dr. W.B. Van den Hout for helpful comments and suggestions.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2018.01.016.

REFERENCES

[1] Brazier J, Ratcliffe J, Saloman J, et al. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford, UK: Oxford university press, 2016.
[2] Drummond MF, Sculpher MJ, Claxton K, et al. Methods for the economic evaluation of health care programmes. Oxford, UK: Oxford university press, 2015.
[3] Bansback N, Brazier J, Tsuchiya A, et al. Using a discrete choice experiment to estimate health state utility values. J Health Econ 2012;31 (1):306–18.
[4] Craig BM, Pickard AS, Stolk E, et al. US valuation of the SF-6D. Med Decis Making 2013;33(6):793–803.
[5] Viney R, Norman R, Brazier J, et al. An Australian discrete choice experiment to value EQ-5D health states. Health Econ 2014;23(6):729–42.
[6] Norman R, Viney R, Brazier J, et al. Valuing SF-6D health states using a discrete choice experiment. Med Decis Making 2014;34(6):773–86.
[7] Craig BM, Reeve BB, Brown PM, et al. US valuation of health outcomes measured using the PROMIS-29. Value Health 2014;17(8): 846–53.
[8] Norman R, Viney R, Aaronson N, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. Qual Life Res 2016;25(3):637–49.
[9] McFadden D. Conditional logit analysis of qualitative choice behavior. 1973. Available from: https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf. [Accessed January 2018].
[10] Manski CF. The structure of random utility models. Dec. Theory 1977;8:229–54.
[11] Chorus CG. Introduction. Random Regret-based Discrete Choice Modeling. Berlin, Germany: Springer, 2012.
[12] Mulhern B, Bansback N, Brazier J, et al. Preparatory study for the revaluation of the EQ-5D tariff: methodology report. Southampton, UK: NIHR Journals Library, 2014: Health Technology Assessment, No. 18.12., Chapter 9, Stage 2 online and CAPI comparison study methods and results.
[13] Jonker MF, Attema AE, Donkers B, et al. Are health state valuations from the general public biased? A test of health state reference dependency using self-assessed health and an efficient discrete choice experiment. Health Econ 2016;26(12):1534–47.
[14] Flynn TN. Using conjoint analysis and choice experiments to estimate QALY values. Pharmacoeconomics 2010;28(9):711–22.
[15] Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. Med Decis Making 2015;35(3):328–40.
[16] Norman R, Mulhern B, Viney R. The impact of different DCE-based approaches when anchoring utility scores. Pharmacoeconomics 2016;34(8):805–14.
[17] Soman D, Ainslie G, Frederick S, et al. The psychology of intertemporal discounting: why are distant events valued differently from proximal ones? Market Lett 2005;16:347–60.

[18] Attema AE, Brouwer WB. On the (not so) constant proportional trade-off in TTO. Qual Life Res 2010;19(4):489–97.
[19] Craig BM, Busschbach JJ. The episodic random utility model unifies time trade-off and discrete choice approaches in health state valuation. Popul Health Metr 2009;7:3.
[20] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21(2):271–92.
[21] Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004;42(9):851–9.
[22] Maddala T, Phillips KA, Johnson FR. An experiment on simplifying conjoint analysis designs for measuring preferences. Health Econ 2003;12(12):1035–47.
[23] Louviere JJ, Islam T, Wasi N, et al. Designing discrete choice experiments: do optimal designs come at a price? J Consumer Res 2008;35:360–75.
[24] Jonker MF, Donkers B, De Bekker-Grob, et al. The impact of level overlap and color coding on attribute non-attendance in discrete choice experiments. Value Health, in press, http://www.valueinhealthjournal.com/article/S1098-3015(17)33609-4/fulltext.
[25] Sandor Z, Wedel M. Designing conjoint choice experiments using managers' prior beliefs. J Market Res 2001;38:430–44.
[26] McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. J Health Econo 2006;25(3):418–31.
[27] Sandor Z, Wedel M. Heterogeneous conjoint choice designs. J Market Res 2005;42:210–8.
[28] Life science grid, Surf/SARA. Available from: https://www.surf.nl/binaries/content/assets/surf/en/2016/life-science-grid.pdf.
[29] Scherpenzeel A, Das J, Ester P, et al. "True" longitudinal and probability-based Internet panels: evidence from The Netherlands. Social and behavioral research and the Internet 2010:77–103.
[30] Mulhern B, Shah K, Janssen MFB, et al. Valuing health using time trade-off and discrete choice experiment methods: does dimension order impact on health state values? Value Health 2016;19(2):210–7.
[31] Mulhern B, Norman R, Lorgelly P, et al. Is dimension order important when valuing health states using discrete choice experiments including duration? Pharmacoeconomics 2017;35(4):439–51.
[32] Samuelson PA. A note on measurement of utility. Rev Econ Studies 1937;4:155–61.
[33] Mazur JE. An adjusting procedure for studying delayed reinforcement. In: Commons ML, Mazur JE, Nevin JA,eds. Quantitative Analyses of Behavior, Vol. 5: The effect of delay and of intervening events on reinforcement value. Hove, UK: Erlbaum, 1987. p. 55–73.
[34] Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. MN, USA: Federal Reserve Bank of Minneapolis, Research Department Minneapolis, 1991.
[35] Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. Stat Comput 2014;24:997–1016.
[36] McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P,ed. Conditional Logit Analysis of Qualitative Choice Behavior. New York, NY: Academic Press, 1973. p. 105–42.
[37] Mathes T, Jacobs E, Morfeld J-C, et al. Methods of international health technology assessment agencies for economic evaluations-a comparative analysis. BMC Health Serv Res 2013;13:371.
[38] Brouwer WB, Niessen LW, Postma MJ, et al. Need for differential discounting of costs and health effects in cost effectiveness analyses. BMJ 2005;331(7514):446.
[39] Robinson A, Spencer AE, Pinto-Prades JL, et al. Exploring differences between TTO and DCE in the valuation of health states. Med Decis Making 2017;37(3):273–84.
[40] Doyle JR. Survey of time preference, delay discounting models. Judg Decis Making 2013;8(2):116–35.