



Contents lists available at ScienceDirect

Journal of Health Economics

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



Living up to expectations: Experimental tests of subjective life expectancy as reference point in time trade-off and standard gamble

Stefan A. Lipman^{a,*}, Werner B.F. Brouwer^{a,b}, Arthur E. Attema^a

^a Erasmus School of Health Policy & Management, Erasmus University Rotterdam, the Netherlands

^b Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands

ARTICLE INFO

Article history:

Received 1 March 2019
Received in revised form 13 March 2020
Accepted 14 March 2020
Available online xxx

JEL classification:

I1
D9

Keywords:

Time trade-off
Standard gamble
QALY model
Prospect theory
Subjective life expectancy
Reference point

ABSTRACT

Earlier work suggested that subjective life expectancy (SLE) functions as reference point in time trade-off (TTO), but has not tested or modelled this explicitly. In this paper we construct a model based on prospect theory to investigate these predictions more thoroughly. We report the first experimental test of reference-dependence with respect to SLE for TTO and extend this approach to standard gamble (SG). In two experiments, subjects' SLEs were used to construct different versions of 10-year TTO and SG tasks, with the gauge duration either described as occurring above or below life expectation. Our analyses suggest that both TTO and SG weights were affected by SLE as predicted by prospect theory with SLE as reference point. Subjects gave up fewer years in TTO and were less risk-tolerant in SG below SLE, implying that weights derived from these health state valuation methods for durations below SLE will be biased upwards.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Time trade-off (TTO) and standard gamble (SG) are two popular methods to value health states, i.e. to obtain utility weights relevant for determining quality adjusted life-years (QALYs).¹ Although the methods share a similar purpose, their framing and outcomes differ substantially (Bleichrodt and Johannesson, 1997; Bleichrodt, 2002), with SG weights typically being higher than TTO weights

(e.g. Read et al., 1984; Torrance, 1976). Bleichrodt (2002) proposed that these differences could be explained by differences in the theoretical assumptions underlying TTO and SG. Both methods' QALY weights are typically calculated using theoretical models that have been shown to be empirically invalid, i.e. expected utility theory (for violations, see: Llewellyn-Thomas et al., 1982; Starmer, 2000) and the linear QALY model (for violations, see: Bleichrodt and Pinto, 2005; Abellan-Perpinan et al., 2006). More specifically, TTO and SG weights are biased, according to Bleichrodt (2002), because individuals show several empirical deviations from these simplified models, including loss aversion, probability weighting, utility curvature and scale compatibility. The first three of these deviations can be modelled through prospect theory, and Bleichrodt (2002) proposed that such modelling could reduce the dif-

* Corresponding author at: P.O. Box 1738, 3000 DR Rotterdam, the Netherlands.

E-mail address: lipman@eshpm.eur.nl (S.A. Lipman).

¹ These weights are sometimes referred to as 'utilities'. We will use the term QALY weights, and TTO or SG weights to refer to QALY weights elicited by TTO and SG respectively.

<https://doi.org/10.1016/j.jhealeco.2020.102318>

0167-6296/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Please cite this article in press as: S.A. Lipman, W.B.F. Brouwer and A.E. Attema, Living up to expectations: Experimental tests of subjective life expectancy as reference point in time trade-off and standard gamble. J. Health Econ. (xxxx), <https://doi.org/10.1016/j.jhealeco.2020.102318>

ference between TTO and SG, for which some empirical support was found by Lipman et al. (2019b).

Prospect theory was originally developed as an alternative to expected utility (EU) theory for decision making under risk and uncertainty (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). Most importantly, prospect theory assumes reference-dependence, i.e. outcomes are not evaluated in final terms, but as changes relative to a reference point (RP). The RP is a neutral outcome, such as the status quo (i.e. current health), but many alternative comparators have been argued to be able to serve as RP, such as the lowest possible outcome (Attema et al., 2012; Bleichrodt et al., 2001), the guaranteed outcome in TTO or SG (van Osch and Stiggelbout, 2008; van Osch et al., 2004), or the best outcome available (van Osch et al., 2006). Furthermore, the RP may be influenced or formed by aspirations, expectations, norms, and social comparisons (Tversky and Kahneman, 1991). It is, however, paramount to determine the exact location of the RP, as this 'neutral outcome' will divide all other outcomes into gains and losses. Within prospect theory, this is especially relevant, as it assumes loss aversion, i.e. losses (relative to the RP) carry more weight than gains of the same size. Furthermore, in prospect theory probabilities can be transformed non-linearly by means of a probability weighting function, which may also differ between gains and losses.

Often it remains unclear exactly how RPs are selected, and how RP selection should be modelled within prospect theory (Wakker, 2010). Two different streams of literature have produced insights on the role of RPs in health-related decision making, which we try to unify in this paper. First, in applications of prospect theory to health outcomes, typically some plausible assumption is made about which outcome could serve as RP. This approach, where typically the RP is selected from the outcomes available *within* the scenarios presented to respondents, allows for tractable modelling and the formation of empirical predictions based on these assumptions. For example, earlier work on RP location for TTO and SG has suggested that the certain outcome, i.e. the impaired health state, will likely serve as RP in these health state valuation exercises (van Osch et al., 2006). Using such an approach to RP selection, prospect theory has been successfully applied in the health domain, for example, earlier work showed that the main tenets of prospect theory (e.g. loss aversion and probability weighting) apply to decisions about human lives (Kemal and Paraschiv, 2018), length of life (Attema et al., 2013; Lipman et al., 2019b; Verhoef et al., 1994; Treadwell and Lenert, 1999; Lipman et al., 2018), and quality of life (Attema et al., 2016). Second, a literature exists suggesting that RPs that originate *outside* the specific decision task at hand may also be selected by respondents. Such studies typically observe some effect of these reference-outcomes on decision-making or well-being, and conjecture that this effect may be due to reference-dependence. Examples of such suggested reference-dependence are RPs based on expectations for length (van Nooten and Brouwer, 2004; Van Nooten et al., 2009) and quality of life (Brouwer et al., 2005; Wouters et al., 2015), or social comparisons (Wouters, 2016).

In this paper, we focus on the effects of individuals' subjective life expectancy (SLE), i.e. self-reported anticipated length of life, which could serve as an RP as defined within prospect theory in health state valuations. It is well-known that many individuals expect to live longer than actuarial life expectancy (Brouwer and van Exel, 2005; Péntek et al., 2014; Rappange et al., 2016). Gauge durations in TTO typically do not coincide with these expectations; frequently, projected life span in TTO is considerably shorter. Although individuals may not be fully aware of this reduction during health state valuation (van Nooten et al., 2014), earlier work on SLE has consistently found that individuals with higher SLE gave up fewer years in TTO, and thus associated QALY weights were higher. We will refer to these changes in QALY weights for durations further away from expectations about length of life as the 'SLE effect'. This SLE effect was found for 10-year TTOs (van Nooten et al., 2009), for patients valuing their own health state (Heintz et al., 2013), and for TTOs using a lifetime time-horizon (van Nooten and Brouwer, 2004). Considering that in most cases life years traded off in TTO fall short of SLE, it may seem plausible to assume that these life years are perceived as being in the loss domain already, and thus given up reluctantly (van Nooten and Brouwer, 2004; Van Nooten et al., 2009). This earlier work postulated that the SLE effect may occur as a result of loss aversion, yielding unwillingness in TTO exercises to further reduce lifetime compared to individuals' SLE which serves as RP.

However, this explanation of the SLE effect has never been modelled adequately or tested directly, as earlier work on the SLE effect has relied on investigation of heterogeneity in SLE by means of an observational between-subjects approach, i.e. explaining differences in TTO responses by differences in SLE. Furthermore, if the SLE effect applies to TTO weights as a result of reference-dependence, one could also expect such effects on SG, as this method may also be affected by loss aversion (Bleichrodt, 2002). This has not yet been tested to our knowledge. Therefore, in this paper we extend earlier work on SLE effects in health state valuation by:

- (i) Constructing a model based on prospect theory with reference-dependence with respect to SLE.
- (ii) Reporting an elaborate test of the SLE effect by experimentally varying the tradable life years above and below SLE for both TTO and SG.

More specifically, by developing a model based on prospect theory we are able to construct tractable predictions about the SLE effect for TTO and SG responses, if it indeed serves as RP. We test these predictions by means of within-subjects experimental methodology in which we construct different versions of TTO and SG, to directly compare QALY weights for life years both under and above SLE for each individual. Through this procedure we test if QALY weights differ for durations that can be perceived as either gains or losses compared to SLE and, hence, whether SLE functions as a formal reference point. This approach is applied in two experiments (labelled Study 1 and Study 2).

The remainder of this paper is structured as follows. In Section 2 we define our theoretical model based on

prospect theory and in Section 3 we derive predictions. Section 4 (Study 1) and 5 (Study 2) report the experiments used to test these predictions using different versions of TTO and SG. Study 1 applied the experimental methodology with a convenience sample of students. The results of this study suggest that SLE indeed serves as RP for TTO and SG. In Study 2, the external validity of these findings is tested by recruiting a sample of individuals aged 60 years and older, largely confirming the results from Study 1. In the final sections we discuss these results and conclude.

2. Theoretical framework

2.1. Notation

TTO and SG are denoted as health profiles described as (Q, t) , where Q represents health status and t denotes the age at which the profile ends (e.g. living in a wheelchair until age 85), with D and FH denoting the states Dead and Full Health, respectively. Subscripts (e.g. a, r, x, y) are used to indicate chronic health profiles faced by a decision-maker with age t_a , where duration is defined as $T_x = t_x - t_a$. Importantly, t_a can, but need not, be the decision maker's current age (it could be any $t_a > 0$). Risky prospects are defined as $(Q_x, T_x)_p(Q_y, T_y)$, i.e. health profile (Q_x, T_x) with probability p , and health profile (Q_y, T_y) with probability $1 - p$. Preference relations are defined as usual, i.e. they are weak-ordered (complete and transitive), and denoted by $>$ (strict preference), \succsim (weak preference), and \sim (indifference).

The TTO method asks for a time equivalent in perfect health which yields indifference between T_x years in health state Q and T_y years in FH. The number of years in T_y is varied until the respondent is indifferent between the two options, i.e. $(Q_x, T_x) \sim (FH, T_y)$. The SG method involves a choice between a number of years (T_x) in health state Q_x for certain and a gamble with two outcomes, which are FH during the same time period (T_x) and D . Probability p is varied until the respondent is indifferent between the two alternatives, i.e. $(Q_x, T_x) \sim (FH, T_x)_p(D)$. Typically, preferences in TTO and SG are modelled within the general QALY model (Miyamoto and Eraker, 1989), which assumes that chronic health profiles (Q_x, T_x) can be evaluated by the utility function $V(\cdot)$:

$$V(Q_x, T_x) = U(Q_x) * L(T_x), \tag{1}$$

with $U(Q)$ denoting utility of health status and $L(T)$ denoting the utility of T life years.

Assuming $L(T) = T$ (i.e. the linear QALY model²), with the common normalisation such that $U(FH) = 1$, TTO indifference can be evaluated by:

$$U(Q_x) = \frac{T_y}{T_x}. \tag{2}$$

² Note that this framework assumes no discounting of future life years, i.e. linear utility. This framework has been generalized to include non-linear utility by Miyamoto and Eraker (1989).

SG indifference, on the other hand, additionally assuming EU and $V(D) = 0$, can be evaluated by:

$$U(Q_x) = p. \tag{3}$$

Although Eq. (2) and Eq. (3) are only valid under these strict assumptions (and more general derivations for TTO and SG are available, see for example: Lipman et al., 2019b), these equations are often used in large scale health state valuations (Versteegh et al., 2016; Brazier et al., 2002).

2.2. Reference-dependence model for SG and TTO

Reference points play no role within the frameworks of EU and the general QALY model. Thus, in order to test whether SLE serves as RP, we will supplement the generalized QALY model with prospect theory, following closely the model developed in Lipman et al. (2019b). This means that we assume that the general QALY model holds with the additional assumptions outlined below included.

We assume separate evaluations of gains and losses in life duration compared to an RP, denoted T_r . This RP is an expected health profile, which is taken to last for T_r years, starting from the age (t_a) of the decision maker until their SLE (t_r), i.e. $T_r = [t_a, t_r] = t_r - t_a$. Throughout we will denote durations of health profiles (Q_x, T_x) as deviations with respect to this RP as follows: we will write (Q_x, T_x^*) with $T_x^* = T_x - T_r$. For example, imagine a 50-year old subject with SLE of living until 80. The health profile of living in a wheelchair until age 70 will be denoted as (*living in wheelchair*, T_x^*) with $T_x^* = (t_x - t_a) - (t_r - t_a) = (70 - 50) - (80 - 50) = -10$ (for more examples, see Online Supplements). We restrict our prospect theory model to life duration, even though it has been suggested that reference-dependence may also exist for health status (Wouters et al., 2015; Brouwer et al., 2005). However, both from a theoretical and from an empirical point of view such reference-dependence for Q is hard to approximate. That is, prospect theory is typically applied to single-attribute outcomes, such as money, while health profiles consist of both life duration and health status. Multi-attribute characterizations of prospect theory exist, but because health status is a qualitative measure, loss aversion is not theoretically meaningful for this attribute (Bleichrodt and Miyamoto, 2003).

As a solution, we apply an attribute-specific evaluation (Bleichrodt et al., 2009) by making three modifications to the general QALY model, to allow testing for reference-dependence with SLE as RP. First, we modify $L(T)$ in the general QALY model to $L^i(T^*)$, which is a standard ratio scale utility function, that can differ between gain outcomes (i.e. (Q_x, T_x^*) with $T_x^* \geq 0, i = +$) and loss outcomes (i.e. (Q_x, T_x^*) with $T_x^* < 0, i = -$), and is strictly increasing and real-valued. Second, loss aversion is incorporated into our model by taking $L^-(T^*) = \lambda L^i(T^*)$ for $T^* < 0$, where λ denotes a loss aversion index, with $\lambda > 1$ [$\lambda = 1, \lambda < 1$] indicating loss aversion [loss neutrality, gain seeking]. Third, we incorporate probability weighting, by evaluating probabilities in risky prospects by probability weighing functions $w^i, i = +, -$, that assign a number to each probability, with $w^i(0) = 0$ and $w^i(1) = 1$. These probability weighting functions can be different for gains and losses.

We do not modify $U(Q)$ of the general QALY model, but we attempt to control for possible effects of reference-dependence of health status by applying our model only to health profiles where health status is better than what is considered acceptable at the ages under consideration. If, as Wouters et al. (2015) suggested, such acceptability serves as RP for health status, this restriction to acceptable health states may avoid confounding effects as losses will only occur in terms of duration while health status will always be above expectation.

Thus, as in Lipman et al. (2019b), references over risky prospects with both gain and loss outcomes, i.e. $(Q_x, T_x^*)_p(Q_y, T_y^*)$, with $T_x^* \geq 0 > T_y^*$ are evaluated by:

$$w^+(p)U(Q_x)L^+(T_x^*) + w^-(1-p)U(Q_y)L^-(T_y^*), \tag{4}$$

while preferences over risky prospects $(Q_x, T_x^*)_p(Q_y, T_y^*)$ for either gains or losses are evaluated by:

$$w^i(p)U(Q_x)L^i(T_x^*) + (1-w^i(p))U(Q_y)L^i(T_y^*), \quad i = +, - \tag{5}$$

where $i = + [-]$ when $T_x^*, T_y^* > [<] 0$, i.e. both outcomes are gains or losses. Lipman et al. (2019b) show that when $w^i(p) = p$, $\lambda = 1$, and no distinction is made between gains and losses (i.e. no reference-dependence), this model reduces to the general QALY model.

3. Predictions

In this paper we consider two versions of TTO and SG. Typically, TTO and SG involve 10-year durations that start at current age. Instead, in this paper, we let the 10-year period in a reduced health state, which occurs in both TTO and SG, a) start at SLE, i.e. $t_a = t_r$ or b) end at each individual's SLE, i.e. $t_a = t_r - 10$. If SLE functions as RP, for a) the gauge duration occurs completely above SLE and thus always involves considerations in the gain domain (because $t_a = t_r$ gives $T_x^*, T_y^* > 0$). Similarly, for b) the gauge duration occurs completely below SLE and thus involves trade-offs in the loss domain (because $t_a = t_r - 10$ gives $T_x^*, T_y^* \leq 0$). Therefore, we label versions with gauge durations completely above SLE as gain versions (i.e. TTO-gains and SG-gains), while those versions with life years occurring completely below SLE are labelled as loss versions (i.e. TTO-losses and SG-losses). To distinguish between the starting ages in these versions for gains and losses, we add superscripts g and l , i.e. $t_a^g = t_r$ and $t_a^l = t_r - 10$. As a final notational convention, given that both versions have the same durations T_x (10 years starting at different ages), for clarity, we will add superscripts to health status for health profiles (Q_x, T_x^*) , such that (Q_x^g, T_x^g) and (Q_x^l, T_x^l) refer to profiles in gain (starting at t_a^g) or loss versions (starting at t_a^l), respectively. For example, consider a subject expecting to live until age 80 ($t_r = 80$). She would receive gain versions with $t_a^g = 80$ and loss versions with $t_a^l = 70$. If SLE indeed serves as RP, this shift from t_a^g to t_a^l allows us to test the SLE effect, as it changes the perception of life years with respect to the RP.

In the remainder of this section, we will employ our theoretical model based on prospect theory with SLE as RP to derive predictions about the SLE effect on TTO and SG. We will obtain these predictions by illustrating the

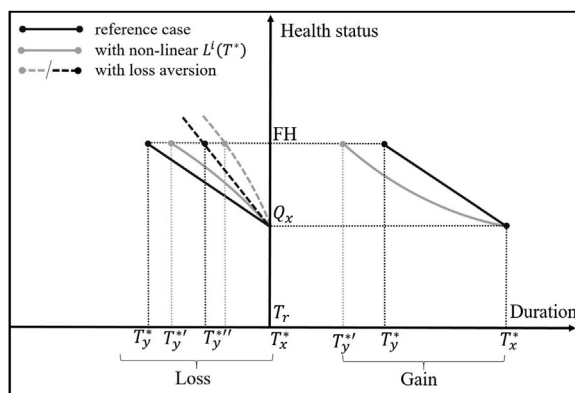


Fig. 1. Indifference curves for time trade-off above and below SLE.

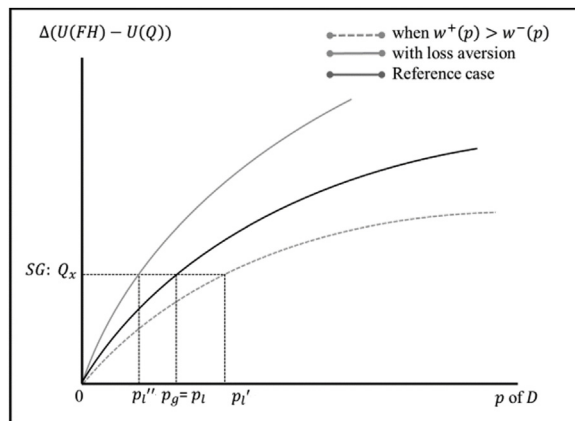


Fig. 2. Indifference curves for standard gambles above and below SLE (superscripts refer to gains and losses).

implications of our prospect theory model as opposed to a reference case, in which linear QALYs and EU hold (i.e. Eq. (2) and Eq. (3) can be applied). For the sake of brevity and clarity, we focus on providing graphical illustrations of these predictions in Fig. 1 and Fig. 2. A complete and formal proof of these predictions can be found in Online Supplements.

3.1. SLE effects for TTO

For TTO, consider as reference case, a subject willing to give up 2 years with reduced health status (Q_x) to obtain full health for 8 more years in the gain version. Using our notation with SLE as RP, this yields the following indifference: $(Q_x^g, 10) \sim (FH^g, 8)$. That is, in the gain version the subject is indifferent between gaining 10 years beyond SLE in health state Q_x and gaining 8 years in full health. We will derive predictions from our model as to what this indifference implies for the years given up in loss versions, i.e. predict T_y^* in $(Q_x^l, 0) \sim (FH^l, T_y^*)$. We first consider the reference case, with linear utility (i.e. $L^-(T^*) = L^+(T^*) = T^*$) and no loss aversion ($\lambda = 1$), which yields the following indifferences: $(Q_x^g, 10) \sim (FH^g, 8)$ and $(Q_x^l, 0) \sim (FH^l, -2)$ for gain and loss versions, respectively (as each year has the same value). In Fig. 1, we have represented such a combination of indif-

ferences more generally. Initially for the reference case, we observe symmetric indifferences: $(Q_x^g, T_x^*) \sim (FH^g, T_y^*)$ and $(Q_x^l, T_x^*) \sim (FH^l, T_y^*)$. That is, shifting t_a^g to t_a^l , which in our experiments with 10 year durations gives $T_r = T_x^*$ for losses, does not affect preferences, as $(T_x^* - T_y^*)$ is equal between gains and losses. These indifferences indicate that in both scenarios each year given up in Q_x (i.e. $T_x^* - T_y^*$) exactly offsets an equal part of the value of the quality of life gained $\Delta(U(FH) - U(Q_x))$. However, such a combination of indifferences does not take into account any discrepancies between gains and losses. In Fig. 1 we provide two illustrations of how the SLE effect for TTO responses due to: a) non-linear utility curvature, and b) loss aversion.

First, whereas TTO typically is derived assuming that utility of life duration is linear, i.e. $L^-(T^*) = L^+(T^*) = T^*$, earlier work on prospect theory has shown that this assumption is likely to be invalid for health outcomes (e.g. Attema et al., 2013; Kemel and Paraschiv, 2018; Lipman et al., 2019b, 2018) and monetary outcomes (e.g. Abdellaoui, 2000; Abdellaoui et al., 2008, 2016; Bruhin et al., 2010). Instead, in prospect theory utility for gains is typically concave, and utility for losses is convex – i.e. utility for life duration is S-shaped. This inflection point in the utility curve may affect years given up in TTO-gains and TTO-losses versions, as it implies diminishing marginal sensitivity for additional life years gained or lost further away from T_r , as opposed to the linearity assumed in the reference case. Hence, it becomes important to consider where the life years given up in Q_x , and the years in which improved quality of life $\Delta(U(FH) - U(Q_x))$ is realized, fall along this S-shaped curve (we illustrate these effects in Fig. 1). For TTO-gains, the years given up in Q_x (e.g. between 8 and 10) are further away from T_r than the years in which improved quality of life $\Delta(U(FH) - U(Q_x))$ is realized (e.g. between 0 and 8). Given that utility for gains is concave, in contrast to the reference case where each year is valued equally, we should find that each year given up in Q_x gets less weight than each year in which improved quality of life $(\Delta(U(FH) - U(Q_x)))$ is experienced. Compared to the linear reference-case, this yields a convex indifference curve, and the respondent will give up more life years to offset the improvement in quality of life $(\Delta(U(FH) - U(Q_x)))$ and restore indifference. Hence, we obtain $((Q_x^g, T_x^*) \sim (FH^g, T_y^*))$, with $T_y^* < T_x^*$. For TTO-losses, however, the years given up in Q_x (e.g. between 0 and -2) occur closer to T_r than the years in which $\Delta(U(FH) - U(Q_x))$ is realized (between -10 and -2). As such, when utility for losses in life duration is convex, each year in which the improvement in quality of life is obtained gets less weight than each year given up. As a result, as compared to the reference case, this yields a concave indifference curve, and the respondent should give up fewer years to offset the improvement in quality of life $\Delta(U(FH) - U(Q_x))$ and restore indifference. Hence, we obtain $(Q_x^l, T_x^*) \sim (FH^l, T_y^*)$, with $T_y^* > T_x^*$.

Second, we take into account loss aversion, i.e. increased sensitivity to losses relative to T_r . Loss aversion yields reluctance to give up life years, and to account for this effect each year given up in Q_x should offset a larger part of

the quality of life gained $\Delta(U(FH) - U(Q_x))$. This yields the steeper indifference curve in Fig. 1, compared to the reference case where people are equally sensitive to gains and losses. As a result, if one is loss averse and durations in TTO occur below T_r , fewer years (T_y^*) should be given up to restore indifference, yielding $(Q_x^l, T_x^*) \sim (FH^l, T_y^*)$. Thus, we predict that loss aversion with respect to SLE will decrease the years given up for TTO-losses versions as compared to gain versions. This conclusion also holds when taking into account non-linearity in the utility curve for life duration (see Fig. 1).

3.2. SLE effects for SG

For SG, consider a subject willing to accept at most a 20% risk of immediate death for SG-gains. In our notation, this yields the following indifferences for gains: $(Q_x^g, 10) \sim (FH^g, 10)_{0.8}(D)$. We will derive predictions from our model as to what this indifference implies for probability of death accepted in loss versions. In the reference case, linear QALYs and EU hold, i.e. the subject will also accept at most a risk of 20% of immediate death for the loss version, i.e. $(Q_x^l, 0) \sim (FH^l, 0)_{0.8}(D)$. In Fig. 2, we have represented such a combination of preferences more generally. Initially we observe the same indifference, i.e. $(Q_x^g, T_x^*) \sim (FH^g, T_x^*)_{p_g}(D)$ and $(Q_x^l, T_x^*) \sim (FH^l, T_x^*)_{p_l}(D)$ with $p_g = p_l$. That is, shifting t_a^g to t_a^l does not affect preferences, i.e. people are willing to risk the same probability of (D) for both SG-gains and SG-losses. This combination of indifferences in the reference case indicates that in both scenarios the possibility of an improvement in quality of life $\Delta(U(FH) - U(Q_x))$ for T_x exactly offsets the generally small chance of dying immediately. In case the difference in quality of life increases, i.e. when $(\Delta(U(FH) - U(Q_x)))$ increases, a larger chance of dying immediately will be accepted. However, just as for TTO, such a combination of indifferences does not take into account any discrepancies in the evaluation of gains and losses.

In Fig. 2 we provide two illustrations of how SG responses are affected when SLE serves as RP: a) probability weighting (which may be different between gain and loss versions), and b) loss aversion. First, whereas in the reference case, probabilities are treated linearly (and thus also identically between gains and losses), our model based on prospect theory allows non-linear probability weighting. Importantly, it is typically observed that probability weighting is less pronounced for losses compared to gains, that is probability weighting is less inverse-S shaped, which has been found for health outcomes (e.g. Attema et al., 2013, Attema et al., 2016) and monetary outcomes (e.g. Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Abdellaoui, 2000). This implies that if SLE serves as RP, the same (small) probability of an extreme outcome receives more decision weight for SG-gains version compared to SG-losses versions. Inversely, when we observe a gain version indifference $(Q_x^g, T_x^*) \sim (FH^g, T_x^*)_{p_g}(D)$, then a higher probability of the extreme outcome (D) may be accepted in the equivalent loss version $(Q_x^l, T_x^*) \sim (FH^l, T_x^*)_{p_l}(D)$. For example, imagine a subject with $p_g = 20\%$, which implies that immediate death

Table 1
Health states used in experiment.

Dimension (EQ-5D)	Description	Best (Q1): 21211	Middle (Q2): 31221	Worst (Q3): 32331
Mobility	You have ... problems with walking	Slight	Moderate	Moderate
Self-care	You have ... problems with washing and dressing yourself	No	No	Slight
Usual activities	You have ... problems with your usual activities	Slight	Slight	Moderate
Pain/discomfort	You have ... pain or discomfort	No	Slight	Moderate
Anxiety/depression	You are ... anxious or depressed.	Not	Not	Not

with decision weight $w^+(0.20)$ offsets³ the possible gain of quality of life for $\Delta(U(FH) - U(Q_x))$ with duration T_x^* . When we have $w^+(p_g) > w^-(p_l)$ for $p_g = p_l$, an increase in p_l to p_l' is required to restore indifference, i.e. for the disutility of (D) to offset $\Delta(U(FH) - U(Q_x))$.

Second, we take into account loss aversion by again assuming increased sensitivity to the possibility of losing compared to SLE, i.e. to durations below T_r . Importantly, the consequence of immediate death (D) differs between gain and loss versions; in the SG-gain version, entails a 20% chance of living up to SLE, while for loss versions dying immediately means living 10 years shorter than expected (i.e. a loss). Hence, SG-gain versions, in our experiment, involved no losses compared to T_r , and were not affected by loss aversion. Hence, if losses are incurred more reluctantly, smaller probabilities ($p_l' < p_g$) of a loss are accepted for the same difference in quality of life ($\Delta(U(FH) - U(Q_x))$). In Fig. 2 we illustrate this by a steeper indifference curve.

Summarizing, for TTO our model predicts two SLE effects, both decreasing the life years given up for losses, while for SG our model predicts SLE effects in opposite directions, where the net direction is determined by the degree of loss aversion and differences in probability weighting for gains and losses. Given that these predictions differ between TTO and SG, shifting gauge duration from above to below SLE (i.e. moving from t_a^g to t_a^l) may yield different SLE effects between these two methods. We can derive no predictions about differences in magnitude of these SLE effects for TTO and SG, as they are affected by different components of prospect theory.

4. Study 1

In this first experiment, we tested our predicted SLE effects for TTO and SG in a lab experiment with a convenience sample of students.

4.1. Methods

This lab experiment started with several questions regarding expectations about length and quality of life followed by an elicitation of TTO and SG. Example instructions and screenshots can be found in Online Supplements. The experiment used a 2 by 2 (method: TTO vs. SG, version:

losses vs. gains) within-subjects design, with randomization by method. The experiment was completed by 102 Business Administration students,⁴ recruited in the Erasmus Behavioral Lab in Rotterdam. A total of 71 males participated, and mean age for our sample was 20.25 (SD=1.22). All students were rewarded course credit for participation in this 30-minute study.

4.1.1. Measures of expectations about length and quality of life

We measured students' expectations about length of life with the following questions (in this order): a) 'What is the minimum age you would hope to become?', b) 'What is the maximum age you would want to become?', and c) 'How old do you expect to become?'. The first two measures were obtained to explore how the typical estimates for SLE fall in between individuals' aspired minimum and maximum age, while question c) measures SLE, using a similar phrasing as van Nooten et al. (2009). Students answered all three questions using a drop-down menu with answers in full years ranging from 30 to 120. To check if health states were considered acceptable, we also explored expectations about quality of life by obtaining a measure of acceptability for the health states that were used to apply TTO and SG (see Table 1). These questions were included as a manipulation check, to determine whether our model, which pertained to acceptable health states, can be applied. To introduce this concept, we used the following instruction (adapted from Wouters et al., 2015): 'In what follows you will receive questions regarding health at different ages. Generally, health deteriorates when we get older. Consider for example an 80 year-old person who is not able to walk further than 1 km. You might find this an acceptable condition for someone of 80, but less acceptable for 20 year old persons.' Next we asked them to rate all three health states using an identical drop-down menu ranging from 30 to 120, using the following question: 'Could you please indicate from which age onwards you find the following three health states acceptable?' Students could also answer 'Never', if they felt that a deteriorated health state was not acceptable at any age.

4.1.2. Operationalization of TTO and SG

All versions of TTO and SG featured a gauge duration of 10 years followed by immediate death, which is typical for this type of valuation exercises (Van Nooten et al., 2009).

³ Typically, in applications of prospect theory outcomes are rank-ordered, where in binary gambles such as SG, probability p is taken to reflect the probability of the extreme outcome in that domain. For the sake of clarity, in these illustrations we deviated from these conventions by taking p_g and p_l to refer to the chance of immediate death in both versions. The Online Supplements show how the conventional notation can be applied without loss of generality.

⁴ Power analysis for paired t-tests with $n=102$, the recommended power of 0.8 (Cohen, 1988) and a significance level of 0.05 indicated that it is adequately powered to detect differences with effect sizes as low as Cohen's $d=0.28$, indicating medium to small effect sizes (Cohen, 1988).

Table 2

Medians, inter-quartile range (first quartile, third quartile), means and standard deviations for measured health outcomes for full sample ($n = 102$).

Outcomes	Median	IQR	Mean	SD
SLE	85.0	(80.00, 90.00)	84.68	9.56
SLE-min	80.0	(70.00, 85.00)	77.20	11.42
SLE-max	100.0	(93.00, 105.00)	99.91	11.80
Acceptable age Q1	60.0	(55.00, 67.00)	59.55	11.23
Acceptable age Q2	70.0	(63.50, 75.00)	67.92	10.08
Acceptable age Q3	79.5	(70.00, 82.75)	76.35	9.71

These four valuation exercises (TTO-gains, TTO-losses, SG-gains and SG-losses) were all completed for three health states described by means of the EuroQol EQ-5D-5L classification system (see Table 1 for the selected health states). TTO and SG were operationalised by using two-stage choice lists (see Online Supplements), which were computerized via Qualtrics to prohibit multiple switching and violations of (stochastic) dominance within each choice list. For TTO, a first choice list identified indifference in years, and afterwards in months in a second choice list. For SG, choice lists elicited indifference with a first choice list identifying indifference at probability intervals of 10%, and afterwards specifying this in percentage points in a second choice list.

4.2. Results

Table 2 reports descriptive statistics for our measures of expectations about length and quality of life. On average, students expected to become close to 85 years old, while wishing to become at least around 77 and at most close to 100 years. As we restricted our theoretical analyses to health states considered acceptable, we determined if students deemed health states Q1, Q2 and Q3 acceptable at all ages used in implemented TTO or SG versions. Overall, health states Q1 and Q2 were considered acceptable by most students, for all ages considered in this experiment, with 84% (Q1) and 72% (Q2) of our sample indicating that such a health status is acceptable from a lower age than the ages considered in our experiment. The most severe health state (Q3) was not considered acceptable at the lowest age considered (i.e. t_h^l), with only 34% of our sample considering such health problems acceptable at the ages presented in the loss versions of TTO and SG. For gain versions, this percentage was considerably higher, at 80%. This

indicates that if reference-dependence exists for health status (as proposed by Wouters et al., 2015), this RP may fall in between the ages considered in the gain and loss versions of TTO and SG for health state Q3. However, we find relatively little non-trading (i.e. QALY weights of 1), with rates of non-trading from as low as 2% to 18% of the sample. Hence, to see if acceptability affected our main results, we ran several tests to explore whether this violation of the simplifying assumptions as described in our theoretical model affects TTO and SG responses (see Online Supplements). We did not observe such an effect of acceptability of health status on TTO and SG responses. As such, we report our main results without excluding respondents from the sample.

4.2.1. Testing predicted SLE effects for TTO

First, we tested our predictions about SLE effects in the two versions of TTO (i.e. TTO-gains and TTO-losses). Table 3 shows aggregate results for TTO responses in both versions. In accordance with our predictions, fewer life years were given up in loss versions of TTO compared to gain versions for all health states (Wilcoxon tests, all p 's < 0.001). According to our model, this suggests that students would either be loss averse or showed less pronounced utility curvature for losses in life duration. Inversely, giving up fewer life years for loss versions will yield higher TTO weights, i.e. higher QALY weights assigned to the same health state. When we analysed our data at within-subjects, we observed that for Q1, Q2 and Q3 respectively, 61, 65, 68% of sample gave up fewer life years in loss-versions. For all three health states, these proportions were significantly larger than the part of our sample that gave up equal life years for both versions, or more life years for loss-versions (all χ^2 's (2, $N = 102$) > 39.71, all p 's < 0.001).

4.2.2. Testing SLE effects for SG

Next, we compared the probabilities of immediate death risked in SG between the two versions (i.e. SG-gain and SG-losses). As can be seen from Table 3, lower probabilities of immediate death were risked for loss versions of SG compared to gain versions (Wilcoxon tests, all p 's < 0.001). According to our theoretical model, this implies that the effect of loss aversion was more pronounced than that of differences in probability weighting. Inversely, this leads to the conclusion that SG with durations below SLE will yield higher QALY weights for the same health state. When

Table 3

Median years given up in TTO and probability of death risk in SG, including within-subject differences between gain and loss versions.

	Gains		Losses			Diff.	
	Median	IQR	Median	IQR	Median	IQR	
TTO							
Years given up							
Best (Q1)	4.00	(2.00, 5.50)	2.13	(0.17, 4.23)	1.08***	(0.00, 3.48)	
Middle (Q2)	5.00	(2.94, 6.50)	3.17	(1.04, 5.00)	1.00***	(0.00, 3.00)	
Worst (Q3)	6.00	(4.50, 8.00)	5.00	(3.00, 6.67)	1.23***	(0.00, 3.00)	
SG							
Probability of D							
Best (Q1)	20.5	(5.75, 35.00)	14.50	(2.25, 30.00)	1.00***	(0.00, 9.75)	
Middle (Q2)	25.00	(20.00, 40.00)	21.00	(10.00, 35.00)	0.00**	(0.00, 10.00)	
Worst (Q3)	38.00	(24.00, 50.00)	31.50	(20.00, 45.75)	1.00***	(0.00, 11.00)	

Note: *, **, *** indicate that differences between gain and loss version were significant at $p < 0.05$, $p < 0.01$, and $p < 0.001$, for the Wilcoxon signed-rank test.

Please cite this article in press as: S.A. Lipman, W.B.F. Brouwer and A.E. Attema, Living up to expectations: Experimental tests of subjective life expectancy as reference point in time trade-off and standard gamble. J. Health Econ. (xxxx), <https://doi.org/10.1016/j.jhealeco.2020.102318>

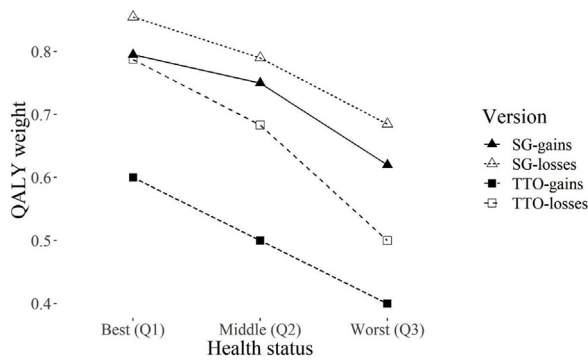


Fig. 3. Median QALY weights for different versions and health states.

we analysed our data within-subjects, we observed that for Q1, Q2 and Q3 respectively, 51, 49, 51% of our sample was willing to take a smaller risk of immediate death in loss versions. For all three health states, these proportions were significantly larger than the part of our sample that assigned equal probabilities to both versions, or was willing to risk a higher chance of immediate death for loss versions (all χ^2 's (2, $N = 102$) > 10.65, all p 's < 0.005).

4.2.3. Comparing SLE effect between TTO and SG weights

In order to compare SLE effects between TTO and SG we needed to normalise weights obtained by these two health state valuation methods to fit on the same scale. We will achieve this normalisation by applying the derivation of TTO and SG weights under EU and the linear QALY framework (i.e. Eq. (2) and Eq. (3)).⁵ Although this is inconsistent with our theoretical model based on prospect theory, it is in line with how TTO and SG responses are typically transformed into QALY weights (see for example: Versteegh et al., 2016; Brazier et al., 2002). Hence, these comparisons may also illustrate the direction and magnitude of reference-dependence with respect to SLE when TTO and SG weights are obtained when this is not accounted for.

Fig. 3 illustrates the aggregate results for our sample. Within versions (i.e. gains or losses), SG weights were significantly higher than TTO weights (Wilcoxon tests, all p 's < 0.037). When comparing within valuation methods (i.e. TTO or SG), QALY weights for health state valuation exercises involving losses produced significantly higher QALY weights, both for TTO and SG (Wilcoxon tests, all p 's < 0.002). For both methods, the differences between gain and loss versions were of similar magnitude for Q1, Q2 and Q3 (not significantly different, Wilcoxon tests, p 's > 0.52). These findings indicate that shifting gauge duration below SLE resulted in an average increase in TTO weights of between 0.15 and 0.23. For SG, a similar pattern was observed, with significant differences between gains and losses visible, where moving life years below SLE increased SG weights on average by 0.02 to 0.12. These SLE effects were significantly larger than 0, and larger for TTO

weights compared to SG weights across all three health states (Wilcoxon tests, p 's < 0.002). We validated these SLE effects using a mixed effects regression, which also showed that our conclusions appear to be unaffected by acceptability of the health states Q1, Q2 and Q3 or gender (see Online Supplements). Finally, we tested whether the typical difference between TTO and SG weights is affected by moving the gauge duration below SLE. To this end, for each subject, we calculated a difference score between TTO and SG per health state, with difference scores being obtained within versions (e.g. TTO-gains vs. SG-gains). This TTO-SG difference was smaller for losses compared to gains (Wilcoxon tests, all p 's < 0.02), but differences remained significantly larger than 0 (Wilcoxon tests with $\eta = 0$, all p 's < 0.04). Collectively, these findings suggest that moving the gauge duration below SLE increases QALY weights, with this SLE effect being larger for TTO than for SG.

4.3. Discussion

This section briefly discusses the results of Study 1 and the main limitations of this experiment that are remedied in Study 2. A more elaborate discussion of the results and limitations can be found in section 6 ('General Discussion').

In accordance with our theoretical predictions, we observed a reduced willingness to give up life years in the TTO-loss version compared to the TTO-gain version (i.e. SLE-effect for TTO). For SG, similar to the TTO results, subjects were reluctant to risk losing life years when deciding about life years that fell short of their expectations (i.e. SLE-effect for SG). When comparing normalised TTO and SG weights (calculated in the common way, based on EU and the linear QALY model), we observed that QALY weights increased when gauge durations were moved below SLE, albeit to a larger extent for TTO. Hence, the difference between TTO and SG was smaller for loss versions. However, the QALY weights elicited for Q1, Q2, and Q3 were low, especially compared to earlier work on health state valuation with general population samples (Versteegh et al., 2016; Devlin et al., 2018). The results from Versteegh et al. (2016) allowed calculating a QALY weight for health states representative of the Dutch general public's valuation (i.e. tariffs). For example, Q1, Q2 and Q3 were assigned valuations of 0.88, 0.79 and 0.68, respectively, which is considerably larger than the valuations in Study 1 (especially those elicited with gain versions).

The low QALY weights elicited in Study 1 suggest that students were willing to give up large proportions of their remaining life or accept high risks of immediate death, just to avoid living in health states with relatively minor problems. At least two reasons can be provided to doubt the validity of such responses to TTO and SG. First, students were paid course credits for participation in this study. Generally, in behavioural experiments in health it is preferred to use financial incentives to motivate respondents to carefully consider their responses (Galizzi and Wiesen, 2018). As such, without an incentive to provide effort in our modified versions of TTO and SG, it could be hypothesized that students invested too little effort in considering the

⁵ Derivations of TTO and SG weights under PT are available (see Lipman et al., 2018), but require assumptions about or measurements of $L^i(T)$, $w^i(p)$ and λ . This is beyond the scope of this paper.

Table 4
Sample characteristics for sample older respondents (Study 2).

Demographic	Categories	Frequency	%
Age	60–65	202	61.6
	66–70	81	24.7
	70+	41	12.5
Sex	Male	127	38.7
	Female	201	61.3
Has chronic disease	Yes	120	36.6
	No	208	63.4
Nationality	United Kingdom	220	67.1
	United States	84	24.4
	Other	24	7.3
Relationship status	Married/in a relationship	187	57.0
	Divorced/Separated/Single	58	17.7
	Widowed	24	7.3
Highest education completed	Doctorate degree (e.g. Ph.D.)	6	1.8
	Graduate degree (e.g. M.Sc.)	39	11.9
	Undergraduate degree (e.g. B.Sc.)	93	28.4
	Technical/Community college	45	13.7
	High school diploma	77	23.5
	No formal diploma	6	1.8
Household income (GBP)	£10,000 - £29,999	127	38.7
	£30,000 - £49,999	68	20.7
	£50,000 - £69,999	34	10.4
	£70,000 - £99,999	13	4.0
	£100,000 or more	12	3.7

consequences of their choices. Hence, to resolve this issue, in Study 2 respondents were provided with a monetary reward for participation. Second, this first experimental exploration of the process by which SLE affects valuations relied on a convenience sample of students. Obviously, this sample is small and not representative for the Dutch population in terms of age and education level. All students were required to imagine being much older than their current age and living in health states they were unlikely to have experienced. This could be problematic, as earlier work has shown that individuals may experience difficulty accurately predicting their future choices (i.e. projection bias, see: [Loewenstein et al., 2003](#)). Furthermore, earlier work has found that SLE is associated with both age and education level ([Brouwer and van Exel, 2005](#); [Péntek et al., 2014](#); [Rappange et al., 2016](#)), and TTO depends on attitudes regarding ageing and end-of-life ([Van Nooten et al., 2016](#)), which may also be different for students compared to older populations. Hence, in Study 2 we applied our empirical tests in a sample of older persons to investigate the external validity of our findings.

5. Study 2

In the second experiment, to test the external validity of our findings, we aimed to replicate our predicted SLE effects for TTO and SG in an online experiment with individuals aged 60 years and older. The methods were almost identical to that of Study 1, and as such we will only highlight modifications to the method below. Furthermore, seeing as we applied a similar analysis strategy, we will present the results of Study 2 without repeating the (more) detailed descriptions found in section 4.2.

5.1. Methods

Study 2 used the same measures of expectations about length and quality of life, health states and operationalisations of TTO and SG as were used in Study 1. However, the experimental task (programmed in Qualtrics) was now distributed online to a sample of 328 people aged 60 years and older. This was done through Prolific, a platform for online research with a large sample of individuals, who mostly live in the UK and US. It allows screening for a wide array of demographics, including age. When this experiment was run, Prolific had around 2600 users that were eligible (i.e. 60 years and older) and active in the last 90 days. Respondents were rewarded 3€ for taking part in this experiment. On average it took respondents 24 min to complete the experiment. Only a single question was added to the original set up in Study 1, i.e. a question to investigate experience with chronic illness.⁶ Demographic characteristics for this sample of older people can be found in [Table 4](#).

5.2. Results

[Table 5](#) reports descriptive statistics on expectations about length and quality of life. The findings for SLE were similar to Study 1 with median SLE being 85 years old. Compared to the students in Study 1, respondents wished to become significantly less old (i.e. SLE-max was smaller) and considered impaired health states acceptable from a higher age onwards (Wilcox tests, p 's < 0.001). Consequently, only

⁶ 'Do you have a health condition or disease with long lasting effects (i.e. a chronic condition)? Examples of chronic conditions are: arthritis, COPD, asthma, diabetes, hepatitis, AIDS, and cancer.'

Table 5

Medians, inter-quartile range (first quartile, third quartile), means and standard deviations for measured health outcomes for older sample ($n = 328$).

Outcomes	Median	IQR	Mean	SD
SLE	85.00	(80.00, 90.00)	84.68	9.56
SLE-min	84.00	(80.00, 86.00)	82.43	7.30
SLE-max	95.00	(90.00, 100.00)	93.93	9.38
Acceptable age Q1	70.00	(65.00, 80.00)	70.21	11.23
Acceptable age Q2	75.00	(70.50, 82.00)	75.50	10.52
Acceptable age Q3	80.00	(75.00, 86.25)	80.79	9.79

53% (Q1), 36% (Q2) and 16% (Q3) of the respondents considered these health states acceptable at all ages considered in the experiment. These were significantly smaller proportions than observed in Study 1 (all χ^2 's ($2, N = 328$) > 47.95, all p 's < 0.001). A risk of having older respondents completing the experiment is that t_a^l (i.e. the age they are asked to imagine to be in the loss versions of TTO and SG) is lower than their current age. This was the case for 32 respondents (10% of the sample). However, excluding these respondents did not affect our results (see Online Supplements). Furthermore, compared to Study 1 for all conditions and health states we found larger amounts of non-trading with rates of non-trading ranging from 12.5% to 34% of the sample (all χ^2 's ($2, N = 328$) > 3.93, all p 's < 0.05). As for Study 1, several analyses were performed to check if acceptability of health states affected QALY weights or the main conclusions of our study (see Online Supplements). We also included having a chronic disease in these analyses. Acceptability did not affect QALY weights, but experience with chronic disease was associated with higher QALY weights. However, our main results were similar for those with and without experience with disease (see Online Supplements). Hence, we report on the full sample below.

5.2.1. Testing predicted SLE effects for TTO and SG

Table 6 shows aggregate results for TTO and SG responses in both versions. As in Study 1, fewer life years were given up in the loss versions of TTO compared to gain versions for all health states (Wilcoxon tests, all p 's < 0.001). We observed that for Q1, Q2 and Q3 respectively 50%, 49%, and 41% of the sample gave up fewer life years in loss versions (rather than more or the same), which was a significant majority (all χ^2 's ($2, N = 328$) > 36.67, all p 's < 0.001). As can be seen by comparing Tables 3 and 6, the SLE effect

Table 6

Median years given up in TTO and probability of death risk in SG, including within-subject differences between gain and loss versions.

	Gains		Losses		Diff.	
	Median	IQR	Median	IQR	Median	IQR
TTO						
Years given up						
Best (Q1)	2.92	(0.25, 5.31)	1.17	(0.00, 3.50)	0.00***	(0.00, 2.63)
Middle (Q2)	3.29	(1.00, 5.71)	2.00	(0.08, 4.40)	0.08***	(0.00, 2.00)
Worst (Q3)	4.92	(2.00, 6.65)	3.08	(0.92, 5.50)	0.08***	(0.00, 2.31a)
SG						
Probability of D						
Best (Q1)	10.00	(5.75, 35.00)	10.00	(2.25, 30.00)	0.00	(0.00, 5.00)
Middle (Q2)	13.00	(20.00, 40.00)	12.00	(1.00, 30.00)	0.00	(0.00, 2.00)
Worst (Q3)	22.00	(24.00, 50.00)	21.00	(3.00, 40.75)	0.00	(-1.00, 3.00)

Note: *, **, *** indicate that differences between gain and loss version were significant at $p < 0.05$, $p < 0.01$, and $p < 0.001$, for the Wilcoxon signed-rank test.

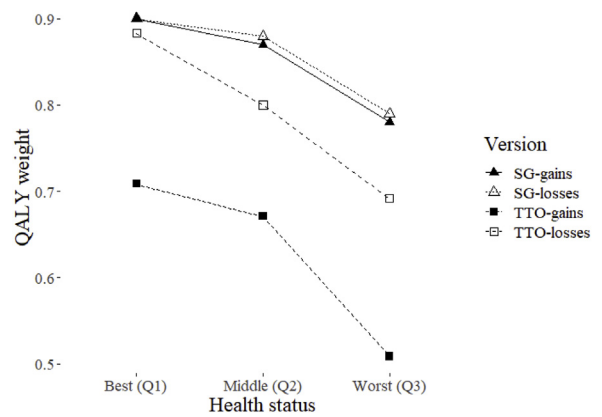


Fig. 4. Median QALY weights for different versions and health states for older sample ($n = 328$).

for TTO appears smaller for this older sample, but this difference was never significant (Wilcoxon test, all p 's > 0.06). In contrast to Study 1, we found no SLE-effect for SG, i.e. no evidence for lower probabilities of immediate death risked for the loss version compared to the gain version. We observed that for Q1, Q2 and Q3 respectively, 35, 30, 31% of our sample was willing to take a smaller risk of immediate death in loss versions (with similar proportions of the sample taking higher risks for loss versions). As can be seen by comparing Tables 3 and 6, the SLE effect for SG was smaller in Study 2, and this difference was indeed significant for all three health states (Wilcoxon test, all p 's < 0.002). Finally, we explored whether excluding non-trading responses affected our findings for SLE effects for TTO and SG. Although this indeed increased effect sizes for TTO, the conclusions remained qualitatively similar (see Online Supplements).

5.2.2. Comparing SLE effect between TTO and SG weights

Fig. 4 illustrates the aggregate normalised QALY weights for each version. For each condition, QALY weights were significantly higher compared to Study 1 (Wilcoxon tests, p 's < 0.04), except for SG losses for Q1 (Wilcoxon test, $p = 0.11$). We also compared our results against the QALY weights calculated using the results by Devlin et al. (2018), which represent QALY weights for a sample representative of the UK (i.e. the country of residence for most of

our sample). We found that the QALY weights elicited in Study 2 were significantly closer to the estimates by [Devlin et al. \(2018\)](#) than those elicited in Study 1 (Wilcoxon test, p 's < 0.04), except for SG-losses for Q1 (Wilcoxon test, $p = 0.11$). Still, our older persons sample reported QALY weights that were significantly different from the [Devlin et al. \(2018\)](#) estimates for all conditions (Wilcoxon tests, p 's < 0.01), except for SG-gains and losses for Q2 and SG-gains for Q3 (Wilcoxon tests, p 's > 0.16). Within versions (i.e. gains or losses), SG weights were significantly higher than TTO weights (Wilcoxon tests, all p 's < 0.007). QALY weights for health state valuation exercises involving losses (compared to gains) produced significantly higher QALY weights for TTO (Wilcoxon tests, all p 's < 0.002), but not for SG (Wilcoxon tests, all p 's > 0.09). Shifting gauge duration below SLE resulted in an average increase in TTO weights of between 0.10 and 0.11. For SG, no such pattern was observed, where moving life years below SLE increased SG weights by 0.001 to 0.02. As in Study 1, the Online Supplements report a qualification of these findings by means of linear mixed-effects regression. The TTO-SG difference was smaller for losses compared to gains (Wilcoxon tests, all p 's < 0.001), but differences remained significantly larger than 0 for all health states (Wilcoxon tests with $\eta = 0$, all p 's < 0.007). Collectively, these findings suggest that moving gauge duration below SLE increases QALY weights for TTO, but not for SG (which leads to smaller TTO-SG differences for loss versions).

6. General discussion

The goal of this paper was to (further) explore SLE effects for TTO and SG by means of a within-subjects approach. We constructed a theoretical model based on prospect theory, which allowed us to test its predictions using different versions for TTO and SG, with a gauge duration occurring either completely below (i.e. losses) or above SLE (i.e. gains). Although EU and the QALY model give no reason to expect differences between these TTO and SG versions, prospect theory, on the other hand, implies that if SLE functions as RP, loss aversion and sign-dependent evaluation of life years and probabilities can give rise to discrepancies between different versions. It was predicted that fewer years would be given up in TTO when elicited below SLE, i.e. TTO weights would be higher. Furthermore, for SG our predictions based on prospect theory suggest that, depending on their loss aversion and probability weighting functions, individuals would be willing to either increase or decrease their risk of immediate death, i.e. the effect on SG weights is ambiguous.

We tested these predictions in two studies with a student (Study 1) and sample of individuals aged 60 years and older (Study 2). We find SLE to be similar to estimates from earlier work ([Brouwer and van Exel, 2005](#); [Van Nooten et al., 2009](#); [Péntek et al., 2014](#); [Rappange et al., 2016](#)). Furthermore, SLE falls in between maximum and minimum aspired ages, suggesting that it could indeed be taken as RP within prospect theory, as this is typically seen as a neutral position ([Wakker, 2010](#)). In accordance with our theoretical predictions, if life years in TTO occurred below SLE, we observed less willingness to give up life years in

both Study 1 and Study 2. Hence, our results for TTO confirm the SLE effect observed in earlier work ([van Nooten and Brouwer, 2004](#); [Van Nooten et al., 2009](#); [Heintz et al., 2013](#); [van Nooten et al., 2014](#)) where similar comparisons were made between individuals expecting to live longer than TTO gauge duration or shorter. Furthermore, seeing as it occurs both in student and samples with older respondents, it appears to be robust to individuals' current age, which provides some support for the external validity of the effect of SLE on TTO. These findings (according to our model) suggest that: a) subjects refrain from giving up life years compared to SLE as a result of loss aversion (as suggested by [Van Nooten et al., 2009](#)), and/or b) subjects show less diminishing marginal utility for life duration for losses compared to gains with respect to SLE.

For SG, the results for Study 1 were similar to those for TTO, i.e. students were more reluctant to risk losing life years when deciding about life years that fall short of their expectations. That is, when the gauge duration occurs below SLE, lower chances of immediate death were taken, which is, to our knowledge, a novel finding. However, these results were not replicated for people aged 60 years and older in Study 2. Given that our model based on prospect theory yields ambiguous predictions for SG, it can provide an explanation for this null result in Study 2. For Study 1, our findings suggest that loss aversion decreased willingness to risk immediate death for gauge durations below SLE. Our model predicts that probability weighting for gains and losses may have offset part of the effect due to loss aversion, which may explain the weaker effect of SLE for SG in Study 1, and perhaps the null result in Study 2. To explain the non-significant SLE effect for SG in Study 2 individuals aged 60 years and older should be less loss averse and/or had larger differences in probability weighting between gains and losses than the student sample in Study 1 had.

Although we derived predictions based on assumptions about loss aversion, utility curvature, and probability weighting, we did not include an empirical measurement of these prospect theory parameters. Instead, our predictions were based on earlier work on prospect theory for both health outcomes (e.g. [Attema et al., 2013, 2016](#); [Attema et al., 2018](#); [Kemel and Paraschiv, 2018](#); [Lipman et al., 2019b, 2018](#)) and monetary outcomes ([Abdellaoui et al., 2008](#); [Bruhin et al., 2010](#); [Kemel and Paraschiv, 2018](#)), where substantial loss aversion and differences in curvature of probability weighting and/or utility functions between gains and losses were observed. As such, our study does not allow to directly test our theoretical explanations for the SLE effect for TTO, nor to determine why SLE effects could be observed in Study 1 but not in Study 2. Hence, combining our experimental approach with measurement of prospect theory parameters is a promising avenue for future research.

However, existing work suggests that older individuals are typically more loss averse and loss aversion decreases with education level ([Gächter et al., 2007](#); [Arora and Kumari, 2015](#)), which would lead us to expect stronger loss aversion in the sample of Study 2. For probability weighting the evidence is inconclusive. [Donkers et al. \(2001\)](#) find some evidence that suggests more pronounced weighting of probabilities for higher ages, but they do not differentiate

between gains and losses. As such, at least two alternative explanations for the null result for SG in Study 2 appear relevant. First, QALY weights for health states Q1, Q2 and Q3 were considerably higher in Study 2 compared to Study 1, with especially SG weights for these mild health states nearing 1.00. It may be possible that no SLE effect for SG is observed due to a ceiling effect, which could be tested by incorporating more severe states in future work. Second, it is possible that this null result is explained by differences between the student and older samples in how they perceive life (and death) at the ages considered in this experiment. Our results provide some indication for this, with students indicating to find health problems acceptable from younger ages than individuals aged 60 years and older. Future work could explore, for example using qualitative interview techniques, the influence of these perceptions on the effects of SLE on QALY weights.

Before arriving at the conclusion that SLE serves as RP in TTO and SG, several alternative explanations, not related to reference-dependence with respect to SLE, and methodological limitations should be considered. First, subjects in both studies were asked to imagine being older than their current age. If subjects did not adopt the instructions in our experiment, the gauge durations in this experiment would be strongly discounted (van der Pol and Roux, 2005; Attema and Brouwer, 2010). Given that loss versions of TTO involved years below SLE, these would necessarily occur earlier in time than years given up in gain versions if current age is adopted instead of SLE. Thus, compared to their current age, life years in gain versions are likely to be discounted more strongly, and given up more willingly compared to life years for the loss version of TTO (i.e. this would predict higher QALY weights for loss versions). Similarly, if subjects used their current age instead of hypothetical ages in our experiment, the time dimension may explain higher utility for SG-losses compared to gains, as for monetary outcomes it is well-known (Abdellaoui et al., 2011; Baucells and Heukamp, 2012; Noussair and Wu, 2006) that risk-seeking increases when lotteries are resolved in the future. As such, SG-gains are resolved further away in the future than loss versions, and thus higher risk-seeking could explain the higher risks of death accepted for gain versions of SG. Hence, although it is not possible to make sure subjects indeed adopted our instructions, in the Online Supplements we show that if subjects did not adopt the ages in our experiment the effects of discounting would be negligible. Hence, given that we do find significant SLE effects, this is not likely to result from failure to adopt t_a^g and t_a^l .

Second, scale compatibility has been suggested to bias both SG and TTO (Bleichrodt, 2002; van Osch and Stiggelbout, 2008). Our manipulation, i.e. shifting life years around SLE, may have caused subjects to focus on life duration in TTO and SG. Given that life duration is fixed and equal in both options in SG, while in TTO life duration is varied along the choice list, this may explain the stronger effect of our manipulation on TTO, especially as the RP was also operationalised on the scale of life duration. Third, even though we provided respondents in Study 2 a monetary incentive to diligently complete our experiment, their rewards were not contingent on their choices (such

incentive compatibility is typically preferred in economic experiments, Galizzi and Wiesen, 2018). Although earlier work in economics suggests that the use of hypothetical choices as opposed to incentive-compatible choices has little to no effect on preferences (Hertwig and Ortmann, 2001; Camerer and Hogarth, 1999), we encourage the exploration of incentive-compatible choices in the context of health. Fourth, all theoretical predictions in this study were based on prospect theory, and hence, we explicitly assumed prospect theory to hold for decisions about health. Several authors have found violations of prospect theory, mostly for monetary outcomes (Birnbau, 2006; Payne, 2005; Bateman et al., 2007), but also for health (Feeny and Eng, 2005). As such, future work could explore if TTO and SG can be modelled in other reference-dependent models (Kőszegi and Rabin, 2006). Finally, to accommodate our subjects and avoid confusion or unnecessary errors, we maintained a consistent ordering throughout the experiment. Future work could explore whether this lack of counterbalancing between-subjects could have affected our conclusions, although other authors find no effects of order on gain-loss framing (e.g. De Dreu et al., 1994).

7. Conclusion

Whereas it is well-known that TTO and SG weights are typically different (e.g. Read et al., 1984; Torrance, 1976), earlier work on the role of SLE has exclusively focused on TTO. Our work suggests that decision-making in both health state valuation methods may be affected by subjective expectations about length of life, with QALY weights being higher for TTO and (to a lesser extent) SG when gauge durations are below SLE; i.e., SLE may serve as RP in health state valuation. This SLE effect could be relevant for the current practice in health state valuation, as this typically involves short gauge durations, which imply losses compared to their SLE for a large part of the sample. For example, when obtaining nationally representative TTO tariffs for EQ5D, EuroQoL typically uses a 10 year duration for health states preferred to death (Oppe et al., 2014), which must fall short of SLE for many subjects. Applying derivations based on EU or linear QALYs will then yield TTO or SG weights that are too high.

Although finding a solution for this biasing effect attributable to SLE seems warranted, as discussed by Heintz et al. (2013), it can be complex to choose an appropriate duration for health state valuation. Durations below SLE may induce reluctance to lose any life years at all, while durations above SLE may yield lower QALY weights as individuals are more willing to lose some of these 'bonus years'. To our knowledge, no compelling normative argument exists to prefer either of these scenarios, suggesting that it may be necessary to acknowledge these possible biases and derive health state utility in a reference-dependent model (as discussed by: Abellan-Perpiñan et al., 2009; Lipman et al., 2019a, b). Therefore, we hope that our attempt to unify earlier work on reference points in health state valuation into a formal model based on prospect theory provides some insight into the consequences of not being able to live up to expectations about length of life.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRedit authorship contribution statement

Stefan A. Lipman: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Werner B.F. Brouwer:** Conceptualization, Writing - original draft, Writing - review & editing, Supervision. **Arthur E. Attema:** Conceptualization, Methodology, Validation, Investigation, Formal analysis, Writing - original draft, Writing - review & editing, Supervision.

Online Supplements

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jhealeco.2020.102318>.

Declaration of Competing Interest

None.

References

- Abdellaoui, M., 2000. Parameter-free elicitation of utility and probability weighting functions. *Manage. Sci.* 46, 1497–1512.
- Abdellaoui, M., Bleichrodt, H., L'haridon, O., 2008. A tractable method to measure utility and loss aversion under prospect theory. *J. Risk Uncertain.* 36, 245.
- Abdellaoui, M., Diecidue, E., Öncüler, A., 2011. Risk preferences at different time periods: an experimental investigation. *Manage. Sci.* 57, 975–987.
- Abdellaoui, M., Bleichrodt, H., L'haridon, O., Van Dolder, D., 2016. Measuring loss aversion under ambiguity: a method to make prospect theory completely observable. *J. Risk Uncertain.* 52, 1–20.
- Abellan-Perpınan, J.M., Pinto-Prades, J.L., Mendez-Martinez, I., Badia-Llach, X., 2006. Towards a better QALY model. *Health Econ.* 15, 665–676.
- Abellan-Perpınan, J.M., Bleichrodt, H., Pinto-Prades, J.L., 2009. The predictive validity of prospect theory versus expected utility in health utility measurement. *J. Health Econ.* 28, 1039–1047.
- Arora, M., Kumari, S., 2015. Risk taking in financial decisions as a function of age, gender: mediating role of loss aversion and regret. *Int. J. Appl. Psychol.* 5, 83–89.
- Attema, A.E., Brouwer, W.B., 2010. The value of correcting values: influence and importance of correcting TTO scores for time preference. *Value Health* 13, 879–884.
- Attema, A.E., Bleichrodt, H., Wakker, P.P., 2012. A direct method for measuring discounting and QALYs more easily and reliably. *Med. Decis. Making* 32, 583–593.
- Attema, A.E., Brouwer, W.B., L'haridon, O., 2013. Prospect theory in the health domain: a quantitative assessment. *J. Health Econ.* 32, 1057–1065.
- Attema, A.E., Brouwer, W.B., L'haridon, O., Pinto, J.L., 2016. An elicitation of utility for quality of life under prospect theory. *J. Health Econ.* 48, 121–134.
- Attema, A.E., Bleichrodt, H., L'haridon, O., 2018. Measuring ambiguity preferences for health. *Health Econ.*, 1–18.
- Bateman, I., Dent, S., Peters, E., Slovic, P., Starmer, C., 2007. The affect heuristic and the attractiveness of simple gambles. *J. Behav. Decis. Mak.* 20, 365–380.
- Baucells, M., Heukamp, F.H., 2012. Probability and time trade-off. *Manage. Sci.* 58, 831–842.
- Birnbaum, M.H., 2006. Evidence against prospect theories in gambles with positive, negative, and mixed consequences. *J. Econ. Psychol.* 27, 737–761.
- Bleichrodt, H., 2002. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ.* 11, 447–456.
- Bleichrodt, H., Johannesson, M., 1997. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *J. Health Econ.* 16, 155–175.
- Bleichrodt, H., Miyamoto, J., 2003. A characterization of quality-adjusted life-years under cumulative prospect theory. *Math. Oper. Res.* 28, 181–193.
- Bleichrodt, H., Pinto, J.L., 2005. The validity of qalys under non-expected utility. *Econ. J.* 115, 533–550.
- Bleichrodt, H., Pinto, J.L., Wakker, P.P., 2001. Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Manage. Sci.* 47, 1498–1514.
- Bleichrodt, H., Schmidt, U., Zank, H., 2009. Additive utility in prospect theory. *Manage. Sci.* 55, 863–873.
- Brazier, J., Roberts, J., Deverill, M., 2002. The estimation of a preference-based measure of health from the SF-36. *J. Health Econ.* 21, 271–292.
- Brouwer, W.B., Van Exel, N.J.A., 2005. Expectations regarding length and health related quality of life: some empirical findings. *Soc. Sci. Med.* 61, 1083–1094.
- Brouwer, W.B., Van Exel, N.J., Stolk, E.A., 2005. Acceptability of less than perfect health states. *Soc Sci Med* 60, 237–246.
- Bruhın, A., Fehr-Duda, H., Epper, T., 2010. Risk and rationality: uncovering heterogeneity in probability distortion. *Econometrica* 78, 1375–1412.
- Camerer, C.F., Hogarth, R.M., 1999. The effects of financial incentives in experiments: a review and capital-labor-production framework. *J. Risk Uncertain.* 19, 7–42.
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Erlbaum Associates, Hillsdale.
- De Dreu, C.K.W., Carnevale, P.J.D., Emans, B.J.M., Van De Vliert, E., 1994. Effects of gain-loss frames in negotiation: loss aversion, mismatching, and frame adoption. *Organ. Behav. Hum. Decis. Process.* 60, 90–107.
- Devlin, N.J., Shah, K.K., Feng, Y., Mulhern, B., Van Hout, B., 2018. Valuing health-related quality of life: an EQ-5 D-5 L value set for England. *Health Econ.* 27, 7–22.
- Donkers, B., Melenberg, B., Van Soest, A., 2001. Estimating risk attitudes using lotteries: a large sample approach. *J. Risk Uncertain.* 22, 165–195.
- Feeny, D., Eng, K., 2005. A test of prospect theory. *Int. J. Technol. Assess. Health Care* 21, 511–516.
- Gächter, S., Johnson, E.J., Herrmann, A., 2007. Individual-level Loss Aversion in Riskless and Risky Choices.
- Galizzi, M.M., Wiesen, D., 2018. Behavioral experiments in health economics. In: *Oxford Research Encyclopedia of Economics and Finance*.
- Heintz, E., Krol, M., Levin, L.-Å., 2013. The impact of patients' subjective life expectancy on time tradeoff valuations. *Med. Decis. Mak.* 33, 261–270.
- Hertwig, R., Ortmann, A., 2001. Experimental practices in economics: A methodological challenge for psychologists? *Behav. Brain Sci.* 24, 383–403.
- Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica*, 263–291.
- Kemel, E., Paraschiv, C., 2018. Deciding about human lives: an experimental measure of risk attitudes under prospect theory. *Soc. Choice Welfare* 51, 163–192.
- Köszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. *Q. J. Econ.* 121, 1133–1165.
- Lipman, S.A., Brouwer, W.B., Attema, A.E., 2018. A QALY loss is a QALY loss is a QALY loss. *Eur. J. Health Econ.*
- Lipman, S.A., Brouwer, W.B.F., Attema, A.E., 2019a. The corrective approach: policy implications of recent developments in QALY measurement based on prospect theory. *Value Health* 22, 816–821.
- Lipman, S.A., Brouwer, W.B.F., Attema, A.E., 2019b. QALYs without bias? Non-parametric correction of time trade-off and standard gamble weights based on prospect theory. *Health Econ.* 28, 843–854.
- Llewellyn-Thomas, H., Sutherland, H.J., Tibshirani, R., Ciampi, A., Till, J., Boyd, N., 1982. The measurement of patients' values in medicine. *Med. Decis. Mak.* 2, 449–462.
- Loewenstein, G., O'donoghue, T., Rabin, M., 2003. Projection bias in predicting future utility. *Q. J. Econ.* 118, 1209–1248.

Please cite this article in press as: S.A. Lipman, W.B.F. Brouwer and A.E. Attema, Living up to expectations: Experimental tests of subjective life expectancy as reference point in time trade-off and standard gamble. *J. Health Econ.* (xxxx), <https://doi.org/10.1016/j.jhealeco.2020.102318>

- Miyamoto, J.M., Eraker, S.A., 1989. Parametric models of the utility of survival duration: tests of axioms in a generic utility framework. *Organ. Behav. Hum. Decis. Process.* 44, 166–202.
- Noussair, C., Wu, P., 2006. Risk tolerance in the present and the future: an experimental study. *Manage. Decis. Econ.* 27, 401–412.
- Oppe, M., Devlin, N.J., Van Hout, B., Krabbe, P.F., De Charro, F., 2014. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health* 17, 445–453.
- Payne, J.W., 2005. It is whether you win or lose: the importance of the overall probabilities of winning or losing in Risky Choice. *J. Risk Uncertain.* 30, 5–19.
- Péntek, M., Brodzsky, V., Gulácsi, Á.L., Hajdú, O., Exel, J., Brouwer, W., Gulácsi, L., 2014. Subjective expectations regarding length and health-related quality of life in Hungary: results from an empirical investigation. *Health Expect.* 17, 696–709.
- Rappange, D.R., Brouwer, W.B., Van Exel, J., 2016. A long life in good health: subjective expectations regarding length and future health-related quality of life. *Eur. J. Health Econ.* 17, 577–589.
- Read, J.L., Quinn, R.J., Berwick, D.M., Fineberg, H.V., Weinstein, M.C., 1984. Preferences for health outcomes: comparison of assessment methods. *Med. Decis. Mak.* 4, 315–329.
- Starmer, C., 2000. Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *J. Econ. Lit.* 38, 332–382.
- Torrance, G.W., 1976. Toward a utility theory foundation for health status index models. *Health Serv. Res.* 11, 349.
- Treadwell, J.R., Lenert, L.A., 1999. Health values and prospect theory. *Med. Decis. Mak.* 19, 344–352.
- Tversky, A., Kahneman, D., 1991. Loss aversion in riskless choice: a reference-dependent model. *Q. J. Econ.* 106, 1039–1061.
- Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5, 297–323.
- Van Der Pol, M., Roux, L., 2005. Time preference bias in time trade-off. *Eur. J. Health Econ.* 6, 107–111.
- Van Nooten, F., Brouwer, W., 2004. The influence of subjective expectations about length and quality of life on time trade-off answers. *Health Econ.* 13, 819–823.
- Van Nooten, F., Koolman, X., Brouwer, W., 2009. The influence of subjective life expectancy on health state valuations using a 10 year TTO. *Health Econ.* 18, 549–558.
- Van Nooten, F., Koolman, X., Busschbach, J., Brouwer, W., 2014. Thirty down, only ten to go?! Awareness and influence of a 10-year time frame in TTO. *Qual. Life Res.* 23, 377–384.
- Van Nooten, F., Van Exel, J., Eriksson, D., Brouwer, W., 2016. Back to the future: Influence of beliefs regarding the future on TTO answers. *Health Qual. Life Outcomes* 14 (1), 4.
- Van Osch, S.M., Stiggelbout, A.M., 2008. The construction of standard gamble Utilities. *Health Econ.* 17, 31–40.
- Van Osch, S.M., Wakker, P.P., Van Den Hout, W.B., Stiggelbout, A.M., 2004. Correcting biases in standard gamble and time tradeoff utilities. *Med. Decis. Making* 24, 511–517.
- Van Osch, S.M., Van Den Hout, W.B., Stiggelbout, A.M., 2006. Exploring the reference point in prospect theory: gambles for length of life. *Med. Decis. Making* 26, 338–346.
- Verhoef, L.C., De Haan, A.F., Van Daal, W.A., 1994. Risk attitude in gambles with years of life: empirical support for prospect theory. *Med. Decis. Mak.* 14, 194–200.
- Versteegh, M.M., Vermeulen, K.M., Evers, S.M., De Wit, G.A., Prenger, R., Stolk, E.A., 2016. Dutch tariff for the five-level version of EQ-5D. *Value Health* 19, 343–352.
- Wakker, P.P., 2010. *Prospect Theory: for Risk and Ambiguity*. Cambridge University Press.
- Wouters, S., 2016. *Absolutely Relative: on the Value of Health Outcomes*.
- Wouters, S., Van Exel, N.J., Rohde, K.L., Brouwer, W.B., 2015. Are all health gains equally important? An exploration of acceptable health as a reference point in health care priority setting. *Health Qual. Life Outcomes* 13, 79.