ELSEVIER

Value in Health

ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Methodology

The Corrective Approach: Policy Implications of Recent Developments in QALY Measurement Based on Prospect Theory

Stefan A. Lipman, MSc,* Werner B.F. Brouwer, PhD, Arthur E. Attema, PhD

Erasmus School of Health Policy & Management, Erasmus University Rotterdam, DR Rotterdam, The Netherlands

ABSTRACT

Background and Objectives: Common health state valuation methodology, such as time tradeoff (TTO) and standard gamble (SG), is typically applied under several descriptively invalid assumptions, for example, related to linear quality-adjusted life years (QALYs) or expected utility (EU) theory. Hence, the current use of results from health state valuation exercises may lead to biased QALY weights, which may in turn affect decisions based on economic evaluations using such weights. Methods have been proposed to correct responses for the biases associated with different health state valuation techniques. In this article we outline the relevance of prospect theory (PT), which has become the dominant descriptive alternative to EU, for health state valuations and economic evaluations.

Methods and Results: We provide an overview of work in this field, which aims to remove biases from QALY weights. We label this "the corrective approach." By quantifying PT parameters, such as loss aversion, probability weighting, and nonlinear utility, it may be possible to correct TTO and SG responses for biases in an attempt to produce more valid estimates of preferences for health states. Through straightforward examples, this article illustrates the effects of this corrective approach and discusses several unresolved issues that currently limit the relevance of corrected weights for policy.

Conclusions: Suggestions for research addressing these issues are provided. Nonetheless, if validly corrected health state valuations become available, we argue in favor of using these in economic evaluations.

Keywords: corrective approach, loss aversion, prospect theory, QALYs

VALUE HEALTH. 2019; 22(7):816–821

Introduction

Health economic evaluations provide important information to policy makers,[1] such as determining incremental cost-effectiveness ratios (ICERs) of interventions, that is, the incremental costs per unit of health gained. In cost-utility analyses health gains are commonly expressed in quality-adjusted life-years (QALYs), which are obtained by multiplying life duration with the utility weight(s) of the health state(s) experienced. These QALY weights are normalized such that 0 and 1 represent the utility of health states judged equivalent to being dead or perfect health, respectively. It is well known that QALY weights differ between health state valuation (HSV) methods used to obtain them: standard gamble (SG) weights are typically higher than weights obtained with time trade-off (TTO) methodologies.[2–4] Bleichrodt[5] proposed that these differences occur as result of bias because of the "classical elicitation assumption," that is, applying expected utility (EU) theory to analyze individual choices.[6] Although research in behavioral economics and psychology has established many systematic violations of EU (for a review of these violations in the monetary domain, see Starmer[7]), its axioms still underlie QALY weight calculations applied in HSV exercises.[8–10] To better inform healthcare decisions, it has been suggested that these biased QALY weights could be corrected by applying calculations based on alternative utility models such as prospect theory.[6,11]

Prospect theory (PT)[12,13] by now is a well established behavioral theory that assumes people judge states relative to some reference point (such as the current position). Changes relative to that point are perceived as either losses or gains. Furthermore, utility increases for gains are lower than utility decreases for equally sized losses; that is, people are loss averse. People, moreover, are not "perfect calculators." They tend to overweight small probabilities and underweight large ones. This is labeled as "probability weighting."[12,13] It has been suggested that reference points, loss aversion, and probability weighting affect decisions about health (eg,[8,11,14–16]), perhaps more pronouncedly than in financial decision making.[17] Importantly, these insights may

* Address correspondence to: Stefan A. Lipman, MSc, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Email: lipman@eshpm.eur.nl

https://doi.org/10.1016/j.jval.2019.01.013

provide an explanation for the systematic differences between HSV methods[5] and can be used in pursuit of obtaining QALY weights that may more accurately reflect tradeoffs relevant to specific methodologies. For instance, SG responses can be corrected for bias owing to probability weighting, and TTO responses can be corrected for loss aversion. It has been argued that such a *corrective approach* may lead to better (ie, less biased) QALY weights and hence may have relevance for HSV.[5,11] TTO and SG (see Box 1 for examples) have been the focus of the corrective approach because these methods are especially relevant to HSV for generic utility classifications, with EQ-5D tariffs frequently being determined via TTO,[18] whereas SF-6D tariffs have been obtained via SG.[19]

Hence, in this article, we focus on the corrective approach in the context of TTO and SG by providing an overview of developments in the corrective approach. These developments opened up at least two challenges to research and policy. First, applying the corrective approach (with current estimates) may affect ICER calculations and allocation decisions—especially when perfect health is involved. Second, even though loss aversion may lead to bias in HSV, it could reflect a real preference many individuals may hold. Thus, distinguishing between gains and losses may still be seen as relevant in healthcare decision making. Preventing health losses may, for example, have higher societal value than achieving health gains of a similar size (relative to a relevant reference point). Hence, we explore how a loss aversion premium for prevented health losses could be applied if and when deemed relevant by responsible policy makers. Finally, we outline policy implications and important steps for future research.

BOX 1. The Time Trade-Off (TTO) and Standard Gamble (SG) methods

TTO exercises involve choices between living longer (say 10 years) in a imperfect health state or shorter (<10 years) in perfect health. The number of years in imperfect health is varied until the respondent is indifferent between the two options. Assuming the linear QALY model,[57] the utility of the imperfect health state is given by dividing the duration in full health by the duration in imperfect health (eg. as in work by Walters).[19] Hence, if a person considers 6 years in perfect health to be equal to 10 years with severe pain, the utility of this health state is 6/10 = 0.6. The worse the health state, the greater is the reduction in years in perfect health that people would be willing to accept. Similarly, SG methods entail asking subjects to choose between living some period of time (eg, 10 years) in some imperfect health state for sure and a gamble with two outcomes: full health (FH) for the same period of time or immediate death (D). By varying the probability of immediate death, one may derive the utility of the imperfect health state. Under EU (and with the utility of perfect health normalized to 1 and that of death to 0) this utility equals probability 1-p. For instance, if people accept a maximum risk of 10% of immediate death to live the rest of their lifespan in perfect health rather than with moderate back pain, this implies the utility of the health state "moderate back pain" is 0.9. If the health state is worse, people would accept a higher risk of immediate death to regain health, leading to lower QALY weights.

The Corrective Approach: Rationale and Overview of Earlier Work

Acknowledging that decisions about health may be reference dependent, as is done in prospect theory, changes the implications of responses in TTO and SG exercises.[5] These implications crucially depend on the location of the reference point in HSV exercises, which was the topic of some empirical studies (eg, van Osch[20]). Given that this work suggested that the time spent in the imperfect health state was the most frequently applied reference point (coinciding with how TTO and SG are typically framed), we will assume this reference point in HSV throughout the present article. Under this assumption, TTO involves trading off losses in life duration for gains in quality of life, whereas picking the risky option in SG indicates a preference for a mixed gamble generating either a gain in quality of life or a catastrophic loss of life (ie, immediate death). As such, loss aversion may exert upward bias in both methods because of the negative utility of (possible) losses subjects are willing to incur in TTO and SG is amplified by loss aversion and thus signifies a larger utility decrement than assumed under the classical elicitation assumption.[5]

Probability weighting only affects SG and generally has an upward influence on SG weights. This upward bias results from subjects' overweighting of generally small chances of death, and underweighting of the typically large chance of obtaining full health in this method.[5] In other words, if subjects weight probabilities in this manner, accepting a 10% chance of death in SG may signify a larger utility decrement than traditionally assumed. In addition, linear utility of life duration is often assumed in health state valuation (ie, the linear QALY model). Nevertheless, many authors have found utility of life years to deviate from linearity in the ranges typically considered in TTO,[16,21–23] where the severity of this deviation may even depend on how duration is described.[24] Such utility curvature will only affect TTO weights because this method depends critically on tradeoffs in duration. As shown in Bleichrodt,[5] if utility of life years is concave (ie, each extra year of life is worth less) instead of linear, TTO weights are biased downward. Inversely, when utility of life years is convex (ie, each subsequent year is worth more than the previous) instead of linear, then the TTO weights are biased upward.

Although EU is often considered the "right" normative theory,[12,25–27] retaining the classical elicitation assumption mistakes the empirical nature of HSV, in which deviations from EU are likely if not inevitable,[6] with the normative relevance that QALYs may have in economic evaluations. Consequently, several studies exist that applied a corrective approach,[28–34] each using the same two steps: i) quantify the deviations from EU and the linear QALY model, such as loss aversion and nonlinear utility, and ii) use corrective formulas[6,11,29] to account for their confounding effect on HSV. Considerable differences exist between empirical studies regarding both steps, with researchers using different techniques to quantify PT or applying corrective formulas based on different assumptions about decision making. A frequently applied approach is to pre-emptively assume a certain degree of loss aversion, utility curvature, and probability weighting in all respondents.[35] In this type of work, average parameters elicited in earlier work (eg, loss aversion coefficients of 2.25) are applied to each individual. Nevertheless, typically large differences in loss aversion, utility curvature, and probability weighting are observed between individuals; that is, not everyone is equally loss averse or weighs probabilities the same way.

Therefore, other attempts at correcting TTO or SG weights apply an individual approach, in which PT parameters are elicited separately for each respondent, applying corrections for loss aversion[31] or nonlinear utility of life duration,[29,33] for example. In this work, utility of life duration or probability weighting is typically estimated by assuming specific functional forms.[30,33,34,36]

Although such parametric analyses may be practical and efficient, the mathematical properties of the chosen parametrical form may not fit well for some extreme cases, for example.[37] Indeed, literature exists documenting that parametric analysis may result in biases in individual estimates for PT.[38,39] Recently, the nonparametric method by Abdellaoui et al[38] was adapted to correct TTO and SG weights without parametric assumptions.[11] In that study, as was expected under PT, concave utility for life year gains and convex utility for losses were observed, with considerable loss aversion and probability weighting for both gains and losses. After applying the corrective approach, TTO and SG weights converged, as predicted by Bleichrodt.[5] Nevertheless, the resulting corrected QALY weights seemed quite low and compressed, raising questions about their validity (see Appendix I for numerical examples of corrections based on this study).

Collectively, these developments in PT measurement and the corrective approach could be important for health policy because they suggest that it may be possible to move beyond the classical elicitation assumption for TTO and SG weights, which still dominates applications of HSV.

The Impact of the Corrective Approach on Health Policy

Regardless of these developments, the corrective approach currently does not affect the policy domain: only a single study[28] exists that estimated corrected tariff lists (ie, without assuming EU), and no country has adopted the corrective approach in guidelines for economic evaluations. Of course, this gap between the current state-of-the-art in the literature and policy may in part be caused by unresolved questions about validity or feasibility of the corrective approach. We return to these important questions in the final section of this article, for now disregarding them to address two currently understudied corollaries of applying the corrective approach. First, we illustrate with currently available weights that moving from the classical elicitation assumption to a corrective approach may substantially affect ICERs and allocation decisions, especially when treatments involve perfect health. Second, we explore how loss aversion, which produces bias that we argued needs correction in HSV, could still have relevance in the context of health policy.

To Correct or Not to Correct: It Makes a Difference

Currently, TTO and SG weights (or weights derived from classification systems using these methods) are commonly elicited assuming EU or the linear QALY model. Hence, at least implicitly, the classical elicitation assumption is still applied. Our focus is to compare this status quo to the situation in which the corrective approach would be applied. We will refer to TTO and SG weights calculated under the classical elicitation assumption as *classical weights* and refer to *corrected weights* when the corrective approach is applied. Without correction, TTO and SG typically yield different QALY weights,[2–4] and hence, it is obvious that ICERs for the same treatment could vary substantially (and systematically) depending on which method is used to value health benefits—especially for treatments dealing with full health. If we choose to apply corrections, we could observe converging TTO and SG weights, and hence converging ICERs for both methods (see Box 2 for an example using currently available estimates). Similarly, applying the corrective approach may affect allocation decisions in different situations compared with using classical weights (see Box 3 for an example). In both cases, applying the corrective approach will likely lead to a lower valuations of impaired health states.[5,11]

BOX 2. The impact of the corrective approach on ICERs

Imagine a group of patients who experience moderate problems with walking about, slight problems with usual activities and slight pain or discomfort (31221 in EQ-5D nomenclature, β2 in Appendix A). In a study by Lipman et al[11], the classical TTO and SG weights for β2 were elicited at 0.605, and 0.706, respectively. We let $U(.)$ represent the utility assigned to health states. Assume that a treatment is evaluated that returns these patients to full health for 30 years, and the costs for treatment are € 20,000 per year. Without discounting, we then obtain the following ICERs:

$$ICER_{TTO} = \frac{€20{,}000 * 30\ years = €600{,}000}{30 * (U(FH) - U(\beta2)) = 30 * 0.395 = 11.85} = 50{,}632€/QALY,$$

$$ICER_{SG} = \frac{€20{,}000 * 30\ years = €600{,}000}{30 * (U(FH) - U(\beta2)) = 30 * 0.294 = 8.82} = 68{,}027€/QALY.$$

If we repeat our calculations using corrected SG and TTO weights, which were 0.442 and 0.456 respectively (see Appendix A), we obtain the following ICERs:

$$ICER_{TTO-C} = \frac{€20{,}000 * 30\ years = €600{,}000}{30 * (U(FH) - U(\beta2)) = 30 * 0.558 = 16.74} = 35{,}842€/QALY,$$

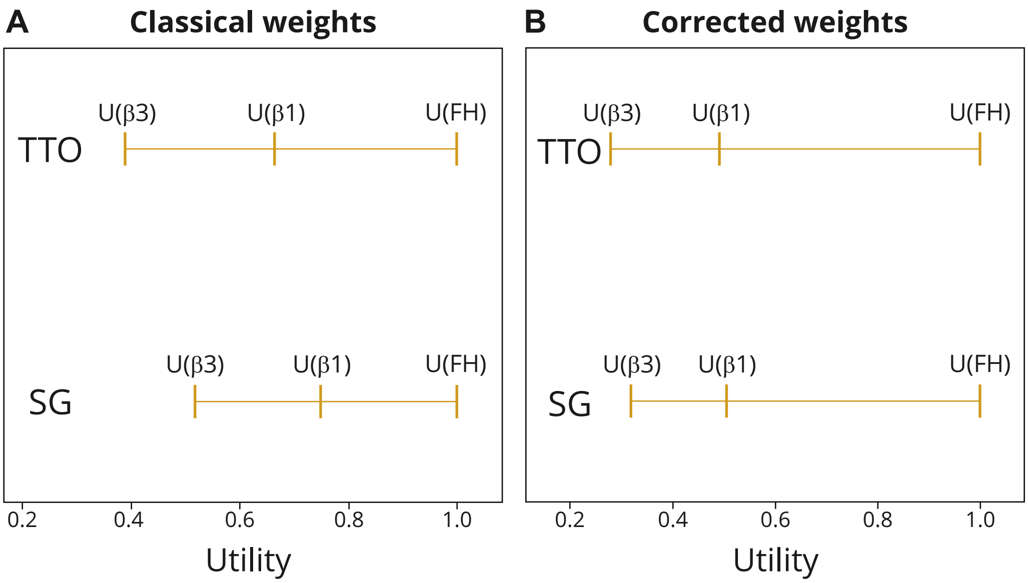
$$ICER_{SG-C} = \frac{€20{,}000 * 30\ years = €600{,}000}{30 * (U(FH) - U(\beta2)) = 30 * 0.551 = 16.32} = 36{,}765€/QALY.$$

Although applying a corrective approach in calculations of QALY weights from HSV exercises will likely improve understanding of choices in TTO and SG, it is not yet clear to what extent this corrective approach ultimately yields QALY weights that better reflect preferences for health states. Obviously, the exact impact of utilizing corrected weights instead of classical weights on subsequent economic evaluations will depend on the respective valuations of health states associated with the treatment and control groups, which raises two crucial issues. First, whereas Box 2 and Box 3 illustrate that the corrective approach may have considerable effects on ICERs and allocation decisions for treatments moving patients between impaired health states and full health, the effects of correction on movements between health

BOX 3. The impact of the corrective approach on allocation decisions

Imagine two patient populations that have the same initial quality of life: a mild health state characterized by slight problems in mobility and self-care (21211, β1 in Appendix A). Treatment A will return population P_a to perfect health, that is, treatment A is curative. Treatment B, on the other hand, will prevent population P_b from a sure loss in quality of life, from state β1 to β3 (32341), a health state characterized by moderate problems with mobility, slight problems with self-care, moderate problems with usual activity, and severe pain. In other words, we have to choose between funding A, which involves P_a gaining $U(FH)-U(\beta1)$, while funding B prevents a loss in quality of life of $U(\beta3)-U(\beta1)$ for P_b. Under the classical elicitation assumption, we observe that the utility differences between $U(\beta3)-U(\beta1)$ and $U(FH)-U(\beta1)$ are of similar magnitude, independent of which method is used to elicit these weights (see Fig. 1A). After correction, however, utility for β1 (see Fig. 1B) has dropped substantially, which could change the allocation decision problem between A and B in favor of Treatment A (ceteris paribus).

Figure 1. A: TTO and SG weight differences with classical estimation. B: TTO and SG weight differences under the corrective approach.



SG indicates standard gamble; TTO, time trade-off.

states that differ only slightly are currently unknown. Given that treatments yielding full recovery are likely to be rare, more insight into how such small improvements or deteriorations in health status are affected by deciding to correct or not to correct (for both SG and TTO) is an important avenue for future research. Second, as can be seen from Box 2 and Box 3, another corollary of applying the corrective approach is that a "perfect-health gap" may be exacerbated. Note that whether or not such a gap emerges depends on final corrected weights. Nevertheless, currently available corrected weights suggest that the distance in utility between the mildest impaired health states, which in current estimates receive lower QALY weights after correcting for bias owing to loss aversion, utility curvature, and probability weighting, and the utility of perfect health, which remains stable at 1.00, increases. As a result, applying the corrective approach may especially affect ICERs of and allocation decisions for treatments involving patients losing or returning to perfect health. Incremental cost-effectiveness thus increases (as shown in Box 2) for treatments that return patients to full health, with potential policy and allocation implications (as in Box 3).

It is yet unclear whether this "perfect health gap" is simply the result of poor correction of bias in TTO and SG, and as such an unintended and undesirable by-product of applying the corrective approach, or reflects actual individual or societal preferences. Implicitly, a perfect health gap already exists in many applications of tariff estimations for utility classification systems.[18,40,41] Whether the larger gap aligns with preferences needs to be established further, especially because correction may enlarge the gap, emphasizing the special status perfect health may have. Nevertheless, earlier work applying a similar corrective approach outside the health domain found that correcting for PT may lead to compression of utility weights.[6] It was suggested that this compression was unrelated to individuals' preferences, but rather resulted from the specific parameterized correction process applied. Such compression of corrected weights could explain the enlarged perfect health gap. Indeed, if the utility estimates of impaired health states are compressed and come closer to the midpoint of the 0 to 1 scale (as in Appendix I), while perfect health remains fixed at 1, this inevitably leads to a (larger) gap.

Moreover, if this compression effect is strong enough, it could also explain the convergence of TTO and SG valuations (as all values cluster in the middle of the scale). The convergence of valuations using both methods has been interpreted in earlier work as evidence of successful correction.[11,34] Hence, it is crucial to determine whether the corrective approach leads to such unwarranted compression of QALY weights, and indeed whether corrected weights better reflect preferences for health states than classical weights. The move toward individual corrective approaches,[11] combined with, for example, ex-post validation of corrected weights in personal interviews in future work, could shed light on this issue. Such insight in the validity of classical and corrected weights is pivotal in interpreting the observed convergence of health state valuations obtained through different methods as well as the increased perfect health gap, and we believe that it is required before the corrective approach is applied in economic evaluation.

To Prevent Is Better Than to Cure: Exploring the Loss Aversion Premium

Applying the corrective approach implies correcting for bias in TTO and SG weights that results from loss aversion, probability weighting, and utility curvature. This may be desirable because TTO and SG are not designed to reflect these time and risk preferences; they were designed to reflect preferences for health states. As such, in our view, if time and risk preferences are deemed relevant for health policy, HSV is not the context in which they should be considered. Rather, this should occur within economic evaluations if deemed appropriate. For time preferences this is already common practice: often a discount rate is applied to future life years in cost-effectiveness analyses,[1] which may reflect societal time preferences for health outcomes.[42] Nevertheless, to avoid "double discounting," TTO weights should be adjusted for individual utility curvature (or time preferences) before applying such societal discount rates in economic evaluations.[43] Thus, individuals' discounting in TTO should be corrected for initially in HSV, and policy makers can decide if and which societal time preference is to be incorporated in economic evaluations. Application of the corrective approach would extend the first step of this sequence to also correct for loss aversion and probability weighting. Nevertheless, no work exists on providing a rationale or methods for also applying this second step for loss aversion or probability weighting, even though it is well known that loss aversion and probability weighting apply to health outcomes as well.[11,14,44–46]

Several authors provided arguments that loss aversion and probability weighting, although yielding bias in TTO and SG, need not be irrelevant or irrational. For example, Huber et al[47] wrote, "In many settings, one cannot tell whether loss aversion is a bias or merely a reflection of the fact that losses have more emotional impact than gains of equal magnitude." Similarly, Diecidue and Wakker[48] argued that probability weighting could reflect individuals' decision that some outcomes are especially important and should receive more attention than equally likely outcomes. As such, just as societal time preference, both probability weighting and loss aversion *could* provide information relevant for economic evaluations and healthcare decision makers. They may signal that (possible) health losses are perceived to have large emotional impact by many members of society. Hence, this preference information could be viewed as a relevant input in decisions based on economic evaluations of health technologies dealing with (risks of) health losses. Below, we will explore how policy makers may include such behavioral insights in such economic evaluations, with a focus on loss aversion (developing a similar approach for probability weighting is beyond the scope of this article and less intuitive in the context of economic evaluation).

To interpret or apply insights based on loss aversion in economic evaluations, it is important to consider which reference point is taken—otherwise losses and gains are undefined. For example, one could take individuals' current health as reference point, which implies that preventive treatments reduce health losses, while curative treatments generate health gains (typically after some loss is incurred). Loss aversion could then refer to a social preference for preventive treatments over curative treatments (ceteris paribus). Nevertheless, extensive literature on equity weighting in health exists suggesting that people on average prefer to treat those worse off (eg, Van de Wetering[49]). Furthermore, research has also documented that age-dependent expectations about length and quality of life could also serve as reference point.[50–52] Collectively, these findings indicate that if a similar approach is to be developed as for time preference, more research on reference points in decisions about health is required. Nonetheless, in Appendix II, we provide a first suggestion as to how insights from loss aversion may be included in economic evaluations, by incorporating *a loss aversion premium*. When and why policy makers should include a loss aversion premium in economic evaluations may be explored in future work taking a broad view of relevant factors in the decision-making process. For simplicity, this approach, which involves deliberately adjusting the value assigned to changes between health states that involve losses, was applied with current health as reference point. Such a loss aversion premium could be used when this is deemed relevant and normatively acceptable.

Conclusions: Research Agenda and Policy Implications

Besides more discussion on corollaries of the corrective approach, such as the perfect health gap and a loss aversion premium, several steps can be outlined for future research. We suggest that these are necessary for successful potential application of the corrective approach in the policy context. First, the robustness and validity of PT parameters obtained through the corrective approach should be determined, both individually and combined, because differences were observed between studies using different methods.[11,14,44–46,53] A head-to-head comparison of these methods could provide a more in-depth analysis of these differences and their impact on correction. Second, research could focus on replicating and extending earlier work on the corrective approach (eg, Lipman[11]), preferably with a sample representative of the relevant population and test the validity of individually corrected QALY weights. Third, future research should aim to clarify the effect of PT on QALY weights elicited with discrete choice experiments (DCE) because these are employed more frequently in large-scale valuation studies (eg, Versteegh[18]). Given that orthogonal comparisons of TTO, SG, and DCE are nonexistent, only suggestive evidence exists showing that DCE weights are similar to classical TTO and SG weights.[54,55] Nevertheless, given that DCEs are typically applied assuming random utility models and because they use an aggregate approach to HSV, it may be difficult to reconcile with corrections at the individual level. Fourth, as mentioned, the corrective approach crucially depends on assumptions about the reference point. Future work should explore the role and nature of the reference point(s) further, especially for TTO, for example, with an approach as in van Osch,[20] and develop corrections for PT that are applicable when outcomes other than the time spent in reduced health are taken as reference point. Finally, if the results of future research on correcting biases are encouraging, national tariffs using the corrective approach for the relevant health-utility classification, for example, EQ-5D-5L or SF-6D, could be obtained to facilitate the incorporation of the corrective approach within health policy as in Perpiñán.[28]

Summarizing, if future research indeed demonstrates the merit of the corrective approach, our suggestion would be to apply the corrective approach in QALY measurement also in the context of actual decision making, which entails several steps.

First, in HSV exercises, for example, large-scale valuation studies, measure each subject's degree of deviation from EU with the most accurate methods available and adjust individuals' responses accordingly. Although work exists that challenges some of its core presuppositions (eg, suggesting no stable preferences exist at all[56]), PT appears to best capture these deviations.

Second, if these corrected weights are found to be valid (and a better representation of health state preferences than classical weights), national tariffs could be calculated based on corrected weights. These could be used in economic evaluations informing policy makers.

Third, some of the correction factors used to "clean" health state valuations may still be informative for policy makers outside the context of HSV. We have explored how this may be true for loss aversion, in relation to the distinction between interventions producing health gains and those preventing health losses.

To conclude, despite developments and increased research efforts into the corrective approach, many unresolved issues still exist that caution against its widespread use. This suggests that the quest for improving methods for HSV, economic evaluations, and decision making has clearly not ended yet. With this article we hope to have encouraged both researchers and policy makers alike to explore these new opportunities.

Acknowledgments

An earlier version of this article was presented at the 4th EUHEA PhD/Supervisor Conference in September 2017 in Lausanne, Switzerland, discussed by Giacomo Pasini. The authors gratefully acknowledge helpful comments during this session. Additional helpful comments were provided by Matthijs Versteegh.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2019.01.013.

REFERENCES

1. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press; 2015.
2. Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *J Health Econ*. 1997;16(2):155–175.
3. Read JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein MC. Preferences for health outcomes: comparison of assessment methods. *Med Decis Making*. 1984;4(3):315–329.
4. Torrance GW. Toward a utility theory foundation for health status index models. *Health Serv Res*. 1976;11(4):349.
5. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ*. 2002;11(5):447–456.
6. Bleichrodt H, Pinto JL, Wakker PP. Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Manage Sci*. 2001;47(11):1498–1514.
7. Starmer C. Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *J Econ Lit*. 2000;38(2):332–382.
8. Bleichrodt H, Abellan-Perpiñan JM, Pinto-Prades JL, Mendez-Martinez I. Resolving inconsistencies in utility measurement under risk: tests of generalizations of expected utility. *Manage Sci*. 2007;53(3):469–482.
9. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till J, Boyd N. The measurement of patients' values in medicine. *Med Decis Making*. 1982;2(4):449–462.
10. Treadwell JR, Lenert LA. Health values and prospect theory. *Med Decis Making*. 1999;19(3):344–352.
11. Lipman SA, Brouwer W, Attema AE. QALYs without bias? non-parametric correction of time trade-off and standard gamble weights based on prospect theory. SSRN. 2017. Available at: https://ssrn.com/abstract=3051140. Accessed November 1, 2017.
12. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica*. 1979;47:263–291.
13. Tversky A, Kahneman D. Advances in prospect theory: cumulative representation of uncertainty. *J Risk Uncertainty*. 1992;5(4):297–323.
14. Attema AE, Brouwer WB, l'Haridon O, Pinto JL. An elicitation of utility for quality of life under prospect theory. *J Health Econ*. 2016;48:121–134.
15. Jonker MF, Attema AE, Donkers B, Stolk EA, Versteegh MM. 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*. 2017;26(12):1534–1547.
16. Bleichrodt H, Pinto JL. The validity of QALYs under non-expected utility. *Econ J*. 2005;115(503):533–550.
17. Suter RS, Pachur T, Hertwig R. How affect shapes risky choice: distorted probability weighting versus probability neglect. *J Behav Decis Mak*. 2016;29(4):437–449.
18. Versteegh MM, Vermeulen KM, Evers SM, et al. Dutch tariff for the five-level version of EQ-5D. *Value Health*. 2016;19(4):343–352.
19. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res*. 2005;14(6):1523–1532.
20. van Osch SM, van den Hout WB, Stiggelbout AM. Exploring the reference point in prospect theory: gambles for length of life. *Med Decis Making*. 2006;26(4):338–346.
21. Attema AE, Bleichrodt H, Wakker PP. A direct method for measuring discounting and QALYs more easily and reliably. *Med Decis Making*. 2012;32(4):583–593.
22. Wakker P, Deneffe D. Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Manage Sci*. 1996;42(8):1131–1150.
23. Abellan-Perpinan JM, Pinto-Prades JL, Mendez-Martinez I, Badia-Llach X. Towards a better QALY model. *Health Econ*. 2006;15(7):665–676.
24. Craig BM, Rand K, Bailey H, Stalmeier PF. Quality-adjusted life-years without constant proportionality. *Value Health*. 2018;21:1124–1131.
25. Savage LJ. *The Foundations of Statistics*. New York: Wiley; 1954.
26. Harsanyi JC. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *J Polit Econ*. 1955;63(4):309–321.
27. Wakker PP. *Prospect Theory: For Risk and Ambiguity*. Cambridge: Cambridge University Press; 2010.
28. Perpiñán JMA, Martínez FIS, Pérez JEM, Martínez IM. *Debiasing eq-5d tariffs. New Estimations of the Spanish EQ-5D Value Set Under Nonexpected Utility*. Sevilla, Spain: Centro de Estudios Andaluces; 2009.
29. Attema AE, Brouwer WB. The correction of TTO-scores for utility curvature using a risk-free utility elicitation method. *J Health Econ*. 2009;28(1):234–243.
30. Martin AJ, Glasziou P, Simes R, Lumley T. A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *Int J Technol Assess Health Care*. 2000;16(1):137–147.
31. Oliver A. The internal consistency of the standard gamble: tests after adjusting for prospect theory. *J Health Econ*. 2003;22(4):659–674.
32. Wakker P, Stiggelbout A. Explaining distortions in utility elicitation through the rank-dependent model for risky choices. *Med Decis Making*. 1995;15(2):180–186.
33. Stiggelbout AM, Kiebert GM, Kievit J, et al. Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. *Med Decis Making*. 1994;14(1):82–90.
34. van Osch SM, Wakker PP, van den Hout WB, Stiggelbout AM. Correcting biases in standard gamble and time tradeoff utilities. *Med Decis Making*. 2004;24(5):511–517.
35. Pinto-Prades J-L, Abellan-Perpiñan J-M. When normative and descriptive diverge: how to bridge the difference. *Soc Choice Welfare*. 2012;38(4):569–584.
36. van der Pol M, Roux L. Time preference bias in time trade-off. *Eur J Health Econ*. 2005;6(2):107–111.
37. Wakker PP. Explaining the characteristics of the power (CRRA) utility family. *Health Econ*. 2008;17(12):1329–1344.
38. Abdellaoui M, Bleichrodt H, L'Haridon O, van Dolder D. Measuring loss aversion under ambiguity: a method to make prospect theory completely observable. *J Risk Uncertainty*. 2016;52(1):1–20.
39. Abdellaoui M, Bleichrodt H, Paraschiv C. Loss aversion under prospect theory: a parameter-free measurement. *Manage Sci*. 2007;53(10):1659–1674.
40. Xie F, Pullenayegum E, Gaebel K, et al. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Med Care*. 2016;54(1):98.
41. Kim S-H, Ahn J, Ock M, et al. The EQ-5D-5L valuation study in Korea. *Qual Life Res*. 2016;25(7):1845–1852.
42. Attema AE, Brouwer WBF, Claxton K. Discounting in economic evaluations. *Pharmacoeconomics*. 2018;36(7):745. https://doi.org/10.1007/s40273-018-0672-z.
43. MacKeigan LD, Gafni A, O'Brien BJ. Double discounting of QALYs. *Health Econ*. 2003;12(2):165–169.
44. Attema AE, Brouwer WB, l'Haridon O. Prospect theory in the health domain: a quantitative assessment. *J Health Econ*. 2013;32(6):1057–1065.
45. Attema AE, Bleichrodt H, L'Haridon O. Measuring ambiguity preferences for health. *Health Econ*. 2018;27:1699–1716.
46. Kemel E, Paraschiv C. Deciding about human lives: an experimental measure of risk attitudes under prospect theory. *Soc Choice Welfare*. 2018;51(1):163–192.
47. Huber J, Ariely D, Fischer G. Expressing preferences in a principal-agent task: a comparison of choice, rating, and matching. *Organ Behav Hum Decis Process*. 2002;87(1):66–90.
48. Diecidue E, Wakker PP. On the intuition of rank-dependent utility. *J Risk Uncertainty*. 2001;23(3):281–298.
49. Van de Wetering E, Stolk E, Van Exel N, Brouwer WB. Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall. *Eur J Health Econ*. 2013;14(1):107–115.
50. van Nooten F, Brouwer W. The influence of subjective expectations about length and quality of life on time trade-off answers. *Health Econ*. 2004;13(8):819–823.
51. Brouwer WB, van Exel NJ, Stolk EA. Acceptability of less than perfect health states. *Soc Sci Med*. 2005;60(2):237–246.
52. Wouters S, van Exel NJ, Rohde KI, Brouwer WB. Are all health gains equally important? An exploration of acceptable health as a reference point in health care priority setting. *Health Qual Life Outcomes*. 2015;13:79.
53. Lipman SA, Brouwer WB, Attema AE. A QALY loss is a QALY loss is a QALY loss. *Eur J Health Econ*. 2018. https://doi.org/10.1007/s10198-018-1008-9 [Epub ahead of print].
54. Stolk EA, Oppe M, Scalone L, Krabbe PF. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. *Value Health*. 2010;13(8):1005–1013.
55. McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. *J Health Econ*. 2006;25(3):418–431.
56. Slovic P. The construction of preference. *Am Psychol*. 1995;50(5):364.
57. Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Oper Res*. 1980;28(1):206–224.