# The construction of health state utilities

Osch, S.M.C. van

# 8

# Summary and Conclusions

This thesis presents the results of our examination of the construction of health state utilities according to expected utility (EU) and (cumulative) prospect theory (PT). PT extends EU and encapsulates its biases and inconsistencies. The aim of our study is to investigate decision making from both a normative and a descriptive point of view and to improve the measurement of utility in medicine through findings from non-expected utility theory, i.c. PT. This chapter summarizes and draws a number of conclusions from the preceding chapters.

Chapter 1 provided an outline of the thesis. In particular, the distinction between normative, prescriptive, and descriptive models was discussed. Two commonly used methods to measure utilities are the standard gamble (SG) and time trade-off (TTO). The SG is based on normative expected utility (EU) arguments. However, there is much empirical evidence that expected utility is not descriptively valid and that its violations generate upward biases in SG utilities. Among other things, these violations are described in the descriptive model PT. Although lacking the theoretical foundations of the SG, the TTO has emerged as the most frequently used method due to better feasibility, higher discriminative power and better face validity. No empirical research has been carried out yet on the biases inherent in this method.

Chapter 2 provided new insights into existing correction methods for the biases, advanced in the economic literature, and to test them in the medical domain. TTO utilities have been suggested to be pushed downwards by the bias of utility curvature, and upwards by loss aversion and scale compatibility. Thus it has been suggested that TTO biases may neutralize each other. The biases of probability weighting and loss aversion generate an upward effect for SG utilities. The effect of the bias scale compatibility on SG utilities was as yet unknown (see chapter 4 also). In PT, outcomes are described as gains or losses (positive or negative deviations) relative to a (neutral) reference point (32). Therefore, the reference point dictates how to correct SG utilities for probability weighting and loss aversion. However, little is known about the psychology behind the location of the reference point, for which PT does not include a hypothesis. For the health domain, there is no direct evidence concerning the location of the reference point.

In the data presented, the gains-corrected SG showed the best convergence with TTO scores. However, this chapter provided arguments suggesting that the TTO utilities were biased upwards, rather than having balanced biases. Utility curvature was absent at the average level, and, as a result, correcting for utility curvature had little effect at this level.  Moreover, the TTO scores were higher than the theoretically most preferred correction of the SG, the mixed correction. These findings suggest once more that uncorrected SG scores, which were higher than TTO scores, are too high. Further, it is likely that uncorrected TTO scores are also too high. It remains an empirical question, however, whether the theoretically preferred mixed correction of the SG is the appropriate correction. This was evaluated in chapter 4. Qualitative data from both chapters 4 and 5 provided further insights into these findings, and will be summarized and discussed below.

Chapter 3 presented information about the reference point in certainty equivalent (CE) standard gambles. We assessed the outcome in the CE that seemed closest to, or seemed to include, the reference point. Additionally, the psychology behind the reference point was explored. Qualitative data were combined with quantitative data to provide evidence of the reference point in life-year CE gambles and to explore the psychology behind the reference point. On the basis of direct and indirect evidence we concluded that the offered CE most often served as the reference point and, therefore, that a life-year CE gamble is most likely to be perceived as a mixed gamble. This finding can have important consequences for economic evaluations, since for mixed gambles the corrected utilities differ from the uncorrected utilities (11;34). Framing and goals influenced the perception of the reference point. Goals also caused an interaction between the time horizon and the attention paid to an outcome. We argued that motivational constructs should be incorporated into research on the reference point. Because patient preferences are involved in treatment decisions, it is preferable to incorporate a patient's goals into treatment decision making.

Chapter 4 further examined the effect of biases on probability equivalent SG utilities using qualitative data on the reference point and focus of attention. To assess the effect of scale compatibility, correlations were assessed between focus of attention and mean utility. The certain outcome, i.e. the health state to be valued, most often served as the reference point. The SG was then most likely to be perceived as a mixed gamble. Additionally, goals were mostly mentioned with respect to this outcome. These findings can have important consequences for cost-effectiveness analyses, such as correcting preference-based utilities for loss aversion and probability weighting with the appropriate mixed correction formula (11;34). Additionally, we observed that respondents mostly focused on the low outcome of the gamble. Consequently, scale compatibility would still have led to upward biases in the SG utilities.

Chapter 5 provided evidence based on qualitative data that indeed the impaired health state is (most often) used as the reference point in the TTO, as argued by Bleichrodt (4). The use of the impaired health state as reference point was associated with higher utilities as a result of loss aversion. Additionally, it appeared that the main focus for the TTO in our study lay on quantity of life as opposed to quality of life, which would lead to higher utilities through scale compatibility (6). We found little evidence of the bias of utility curvature (34). Moreover, we observed no significant relation between the bias of utility curvature and TTO utilities. It has been argued that patients use less of the utility continuum than the general public when valuing similar health states. We observed that labeling may have a similar effect, but to a lesser extent.

Chapter 6 dealt with cognitive processes underlying the VAS. Strictly speaking, the VAS is not a preference-based measure, and therefore does not provide utilities. It is nevertheless often substituted for SG or TTO, for reasons of feasibility. We therefore evaluated this measure as well. Our experiments showed four not mutually exclusive approaches to valuing a health state by using a VAS. These approaches were, in order of the respondents' preferences, Sort-of, Bisection of line first, Numerical expression, and Division into small segments. The Numerical and Small segments approaches were used less frequently. This coincides with the (visual) concept of the VAS; otherwise a rating scale should be preferred. Respondents appeared to be reluctant to express their valuation numerically. Adjusting the task instruction by indicating the way responses were to be used did not encourage respondents to use a more explicit approach. Next, we examined whether dual processing (an interaction between automatic and controlled information processing) occurred during VAS valuation. We presented evidence of dual processing during health state valuation using the VAS. Controlled processing, for example, being aware of an approach, may enhance reliability of the VAS since the use of a similar strategy on two occasions stands a higher chance of producing the same result. Additionally, awareness is related to explicit approaches (such as bisecting the line first, or dividing it into smaller segments), and also to having confidence in the answer. These findings are an argument in favor of instructing respondents beforehand to determine their approach, to improve reliability.

Chapter 7 dealt with the measurement of health related risk attitude. People differ in their attitude towards health risks. This results in different preventive health risk behavior and treatment preferences. In medical decision making, until now mostly Certainty Equivalent standard gambles have been used to assess risk attitude. Elicitation is a complex task, fraught with biases, as explained above. We therefore developed the health-risk attitude scale (HRAS) in order to assess how persons value their health and manage health risks, and refer to this as health risk attitude. The HRAS aims to predict how a person will resolve risky health decisions in the future. Items for the scale were devised mostly through a literature study and through interviews with patients about their treatment preferences. The psychometric aspects of this instrument were tested in two studies. To assess construct validity, we used general risk taking scales, a domain specific risk scale, standard gambles, the health locus of control scale, and a personality scale. Study 1 described the construction of the first version of the HRAS and documented the validity and reliability. On the basis of a factor analysis and a reliability analysis, the HRAS was reduced to 13 items. Study 2 described the validity and reliability of the final version of the HRAS. Relations with other risk scales were positive. In summary, we have developed a short and simple scale assessing risk attitude in the health context. It showed good reliability (both internal and test-retest) and convergent validity.

Most people working in the medical domain seem to understand intuitively that the SG is biased upward, because most people resort to the TTO to assess utilities. The introduction of PT into medical decision making has provided an explanation for the intuitive gut feelings that the SG utilities generally were much too high. The TTO has removed the standard gamble from its golden throne and has gained power due to its better descriptive abilities. On the basis of qualitative and quantitative data, we have found that scale compatibility generates an upward effect on SG utilities. The qualitative data provided at the end of this thesis on the SG method argued for a mixed-corrected SG. All biases effective in the SG method push SG utilities upwards.

The data concerning the SG reference point further underpins the argument stated in chapter 2 that TTO utilities are still too high, because TTO utilities were higher than the mixed-corrected SG utilities. Additional support for this argument was provided by the qualitative data on TTO utilities, because the counterbalancing bias of utility curvature was infrequently found. This may have been due to our sample, though, because other studies did find this curvature. On the other hand, Stalmeier (105) has pleaded for the use of uncorrected TTOs for utility curvature. The other two biases were observed, as well as the upward generating effect thereof on TTO utilities. We propose to use the mixed-correction formula to correct SG utilities (derived with ping-pong elicitation method) proposed by Bleichrodt, Pinto and Wakker (4). No correction is available as yet for scale compatibility, and more research is clearly needed here. TTO utilities appeared to be too high. However, there is no correction method for scale compatibility today. Awareness of the total effect of the biases affecting TTO utilities is all that we offer now. More research is required for this issue.

The two generic models of cognitive function (dual processing) that were observed during the VAS, appeared also to be present during the TTO and SG. These are the intuitive mode in which judgments and decisions are made and the reasoned (more controlled) mode in which these are made deliberate but in a slower pace. Both appeared to play a role during the elicitation. The distinction between intuition and reasoning has been a topic of interest (106;107). Emotions and beliefs play a big role in decision making for which EU holds no room. EU, like 'traditional theories of economic decision-making' assumes that humans are fundamentally rational creatures. However, humans are reproducibly irrational in a number of characteristic ways.

Utility cannot be divorced from emotion. It has even been found that the framing effect was specifically associated with neural activity in a key emotion center in the human brain. This finding also highlights the importance of incorporating emotional processes within models of decision making (108). In decision analysis, it is commonly assumed that the right normative model for decision under uncertainty is expected utility. Some implicitly assume the EU model to have no deviations and biases by using uncorrected standard gamble utilities. We can conclude that this is incorrect. A further question that occurs is: Can one say that the expected utility theory is correct if so much deviations and biases occur? Bleichrodt, Pinto and Wakker argue that biases and inconsistencies are not to be interpreted as irrationalities on the client's part, but are the result of measurement techniques (4).

However, what is so bad about irrationality? In the case of individual decision making there may be nothing wrong with that. Emotion is often regarded as the antithesis of reason. However emotion (e.g. fear) can often be thought of as a systematic response to observed facts. It is safe to argue that emotion is complex. Emotions and rationality have a variable and complex connection, sometimes they strengthen one and other, sometimes they are each others enemy. I can have emotions that are grounded in judgments with which I do not agree. I can even be angry that I feel such unfounded emotions, such as with respect to my fear of spiders. However, my reason can justifiably be told introspectively that it is being unreasonable. We cannot deny that emotions play a role in the way that people arrive at decisions with respect to their physical well being (109). It is possible to formalize these emotions, and these should be incorporated in descriptive decision models. The difficult question that remains is whether emotions should be incorporated into a prescriptive model. Based on the findings in this thesis we can acknowledge that goals in life and emotions play an important role in decision making, but based on these findings we cannot answer the question posed above. However, a theory of decision making that completely disregards feelings such as anticipated regret and loss aversion is not only descriptively invalid, it also results in prescriptions that do not maximize the utility of outcomes as they are actually experienced (106). If emotions are a standard part of the cognitive process that is decision making, then how can one prescribe how to decide without them?

Considering the available decision models, it is not necessarily true that when utilities are corrected for biases, that these biases should not exist. Rather, these biases are not incorporated into the existing models. On the other hand, all elements that are relevant to decision making cannot be expected to be incorporated in one model, as only a limited number of parameters can be used in order to maintain tractability of a model. As argued before, for the time being, we should then be satisfied with what we have, for decisions must be made now (4). The first step is to have a growing awareness of the effect of biases on utilities in the field of medical decision making. The next step is to actually correct for them. A third step is to further formalize the role of emotions in the decision models. But bear in mind that, it still cannot 'incorporate the wind caused by the wings of a butterfly'.