The construction of health state utilities

Osch, S.M.C. van

1

Introduction

Weather forecasting is the science of predicting the state of the atmosphere for a future time and location. However, any model cannot include all variables that are relevant to it, e.g. the weather prediction cannot incorporate the wind caused by the wings of a butterfly.

When a decision is to be made, there are, by definition, two or more possible actions. Each action leads in one or more possible outcomes. Some outcomes will be more preferred than others, and some will be more likely to occur. A decision can be based on emotions, it can be taken intuitively or it can be a reasoned decision. A decision maker may determine the likelihood of each outcome to occur and attach a value to that outcome, and make a (possibly rational) decision. In medical decision-making different treatments (actions) can lead to different outcomes certain risks. Utilities can be elicited to measure the health benefits of treatments, in other words, be used to 'attach values to an outcome'. Hence, the measurement of utility is central in medical decision-making. However, the utility measurement methods themselves are fraught with inconsistencies and biases. Payne et al. argue that preferences, and thus utilities are constructed during elicitation rather than elicitation being a form of uncovering existing values (1). This thesis deals with the measurement of health state utilities. The aim of this thesis is to study decision-making from both a normative (prescriptive) and an empirical (descriptive) point of view. Purpose is to improve the measurement of utility in health care through findings from non-expected utility theory, i.e. cumulative prospect theory (PT).

"Economists often criticize psychological research for its propensity to generate lists of errors and biases, and for its failure to offer a coherent alternative to a rational-agent model … this is just another way of saying that rational models are psychologically unrealistic" wrote Kahneman (2). Psychology indeed provides a description of how people make judgments and decisions, and thus proposes a descriptive decision-making model. A normative theory specifies an optimal set of decision rules. Such as, if medical treatment A is preferred over treatment B, and treatment B is preferred over treatment C; it follows that treatment A is to be preferred over treatment C. A descriptive model tries to represent behavior or the anticipation of behavior. The goal is then to obtain an accurate model of the existing decision process. If discrepancies exist between normative theory and observed behavior, a further question is what people should be told to do in order to satisfy certain goals, i.e. what should be prescribed to them. For this we need a prescriptive theory. Prescriptive models are based on normative theories. Thus, descriptive and prescriptive decision-making models differ in how the parameters comprising the models are obtained. Prescriptive and descriptive decision making models bring forward different perspectives. The descriptive decision making model can be evaluated afterwards by assessing the validity of the model through the reproduction of the behavior of the decision maker. A prescriptive model is to be evaluated according to the ability to bring about decisions that are optimal, and that lead to the predicted (desirable) outcomes.

In medical decision-making, the most commonly used decision making theory is expected utility (EU) theory. EU is based on the axioms (i.e. normative decision rules) formulated by von Neumann and Morgenstern (3). To apply EU, it is essential to quantify uncertainties in terms of probabilities, and values of outcomes in terms of utilities. The expected utility of each available action, obtained by combining probabilities and utilities of its associated outcomes, is used to determine the optimal action (4). Empirical evidence of biases in utility measurement is well documented in management science, economics and psychology (5-7). In medicine, however, the awareness of these biases is much more restricted.

Based on normative EU arguments, the standard gamble (SG) method has historically been considered the gold standard for utility measurement. The major violations of EU are explained by Prospect Theory (PT), a descriptive decision making theory (5;11). These violations include loss aversion and probability weighting. However there is ample empirical evidence that EU is not descriptively valid, and that its violations generate upward biases in SG utilities (8-10). Loss aversion refers to the finding that people are more sensitive to losses than to gains. This, for instance, explains why a casino player at the end of the evening may take high risks to work away his losses. Or why a stockholder experiences more difficulties in selling stocks that have decreased in value than stocks that have increased in value. Probability weighting entails that people process probabilities in a nonlinear manner. The pattern that is most often found is that people overweight small probabilities and underweight large probabilities.

The SG generally requires a respondent to compare the certainty of being in the health state to be valued for the remaining life expectancy, with a gamble that offers a chance (probability p) of optimal health for the remaining life expectancy but also entails a risk of immediate death (probability 1-p). In the generally used probability equivalent of the SG, the probability p is varied so as to identify the point at which the respondent is indifferent to the choice between the health state and the gamble. The utility of the health state is calculated by equating the expected utilities of the two alternatives. In health economics, the time trade-off (TTO) has been developed as an alternative to the SG (10). In the TTO, the subject is asked how many years in optimal health she/he considers to be equivalent to a period (e.g. their remaining life expectancy) in a particular impaired health state. A major and basic difference with the SG is that the TTO is not based on expected utility – in the sense of being a product of probability and outcome - and thus provides a riskless measure. TTO utilities have been found to better reflect individual preferences for health than SG utilities do (6), and have a higher face validity. The most well known biases to affect TTO utilities are scale compatibility, utility curvature and loss aversion (6). Unfortunately, these prominent biases in the TTO measurements have rarely been investigated.

With respect to societal decision-making, utility is included in cost utility analysis. In which the benefits of health care programs are expressed in utility terms and are compared to costs. The use of biased utilities will lead to biased resource allocation decisions. Therefore the joint effect of the aforementioned biases should be minimized in the elicitation of utilities, whether using the SG or using the TTO. This requirement creates a need for more knowledge about these biases in health utility measurement, so as to adequately correct for them or at least to be aware of their effect on utilities (4).

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. Therefore, we will evaluate this measure as well. In medical decision making, until now mostly standard gambles have been used to assess risk attitude. Elicitation is a complex task, fraught with biases. We aim to develop the health-risk attitude scale (HRAS) in order to assess health risk attitude.

The present thesis is based on 9 chapters. In the second chapter the SG and TTO are introduced as methods to assess utilities. The potential biases in these methods that are discussed are loss aversion, probability weighting, scale compatibility, and utility curvature for life duration. This chapter describes correction methods for the aforementioned biases that have been advanced in the economic literature, and tests them in the medical domain. No clear conclusions can be drawn yet in this chapter, because information is lacking on some crucial premises regarding the corrections.

Such information is the topic of subsequent chapters. We then explore, with the use of qualitative research, how utilities are constructed, and which themes and biases influence utilities.

Chapter 3 combines qualitative with quantitative data, so as to provide evidence of the reference point in life-year certainty equivalent (CE) standard gambles and to explore the psychological basis of the reference point. Risk behavior has been shown to depend strongly on the perception of the outcome as either a gain or a loss. According to prospect theory, the reference point, i.e. a point of view, determines how an outcome is perceived. However, no theory on the location of the reference point exists for a standard gambles (whether probability equivalent or certainty equivalent). Additionally, for the health domain, there is no direct evidence for the location of the reference point. With knowledge of the reference point, the proper correction method can be applied to counteract the biases probability weighting and loss aversion.

Chapters 4 and 5 also contain a combination of qualitative and quantitative data, but now for the probability-equivalent SG and for the TTO. The effect of the aforementioned biases on SG and TTO utilities is assessed with the use of qualitative data. The first objective was to locate the SG outcome that is the reference point or that seems to lie closest to the reference point. The second objective was to obtain an indication of whether respondents focus more on the bad outcome or on the good outcome, in order to assess whether scale compatibility results in a systematic bias upwards or downwards. To assess this point, we determined the focus of attention. Additionally, we aimed to verify that a main focus on a bad outcome or good outcome will lead to higher or lower utilities, respectively. Relevant themes that were raised by respondents and that could result in a biased utility were also taken into account. Chapter 5 provides similar research for the TTO.

A frequently used method to assess valuations is the visual analog scale (VAS). In chapter 6, we explored approaches to valuing a health state on a VAS. Cognitive processes involved in for instance valuing a health state, can be carried out at two distinct levels, each with qualitatively different mechanisms. Dual-processing theory states that thoughts, behaviors and feelings result from the interplay of automatic (and implicit) and controlled (and explicit) processing (12). We carried out two experiments in which respondents were probed for approaches used in the VAS. Possible approaches were explored in the first experiment, and were systematically examined in the second.

In medicine, as well as in medical decision making, risk and uncertainty are almost always standard ingredients. Consequently, risk attitude is highly relevant to medical decision-making. It is therefore surprising, that risk attitude, is generally not, or only implicitly, taken into account. Differences among individuals, whether patient or doctor, in risk attitude and, consequently, in their response to risky medical situations can provide valuable information with respect to (the understanding of) treatment preferences and clinical decisions. The CE gambles described in chapter 3 provide a formal approach to assessing risk attitude according to EU. No alternative to this formal approach, which is cognitively difficult and time-consuming, was available. We therefore decided to develop a health risk attitude scale (H-RAS), which we introduce in chapter 7. The H-RAS aims to assess how persons value their health and manage health risks. The H-RAS is psychometrically tested.

In Chapter 9, the main findings and their implications for health care will be summarized and discussed.