Health State Valuation Methods and Reference Points

Happich, Michael; Moock, Jörn; von Lengerke, Thomas

Health State Valuation Methods and Reference Points: The Case of Tinnitus

Michael Happich, PhD,[1] Jörn Moock, MSc,[2] Thomas von Lengerke, PhD[1,3]

[1]Helmholtz Zentrum München—German Research Center for Environmental Health, Institute of Health Economics and Health Care Management (IGM), Munich, Germany; [2]University of Greifswald, Institute of Community Medicine, Greifswald, Germany; [3]Hannover Medical School (MHH), Medical Psychology Unit (OE 5430), Hannover, Germany

ABSTRACT

Objectives: Many studies support the finding that patients, compared to the general public, valuate a given health condition differently. Based on Prospect Theory, this difference can be explained by adaptation processes resulting in differences in individual reference points. Using tinnitus as a case in point, our objective is to analyze empirically to what extent differences in risk attitudes (as a proxy to reference points) mediate differences in health valuations.

Methods: Two hundred ten tinnitus patients and a similar number of unaffected persons indicated their willingness to undergo, hypothetically, an intervention (surgery or treatment) that would either improve or worsen the condition, thus revealing their risk attitudes. Utilities were elicited using three different methods: visual analogue scale (VAS), time trade-off (TTO), and standard gamble (SG). Repeated measure analysis of variance was used to test for mediation of utility differences by reference points.

Results: Health status (affected–unaffected) has a significant effect on tinnitus utilities and risk attitude; at the same time, the latter is significantly associated with utilities. Adjusting for risk attitude, differences by health status disappear for SG and TTO, and are alleviated for VAS.

Conclusion: Reference points in terms of risk attitudes are a potential confounder in the valuation of health states. Taking into account theoretical predictions and issues in measuring SG, TTO, and risk attitudes, these results cast doubt on the construct validity of SG and TTO, and point to the need to recognize and further clarify the role of reference points in health valuation research.

Keywords: health valuation methods, Prospect Theory, reference point, risk attitude.

Introduction

Rising financial pressure on health-care systems requires prioritization of resources in health-care financing [1]. Usually, different medical interventions are compared with respect to both their costs and effectiveness to allow a ranking of health interventions [2]. Nevertheless, most authors agree that such rankings of interventions depend on who is asked, and whose utilities are used [3–7].

Recently, pharmacoeconomic guidelines have been published by independent institutions or administrations of several countries (e.g., the United States, Canada, Australia, New Zealand, and the United Kingdom) for the purpose of health economic evaluations [8]. One issue in these guidelines is recommendations on which method should be used to assess benefits of health care and whose utilities should be implemented into evaluation of benefits. For example, the guideline provided by the National Institute for Health and Clinical Excellence recommends that values should be based on public preferences. This public or societal perspective can be supported by the argument that the general public as tax payers provides the monetary frame for the health system [9]. With some exceptions [6,10], studies usually find that patients provide higher utilities to their own health condition than does the general public for whom the condition is hypothetical [3,7,11–14]. Expressed in a modeling framework; this implies that health status as an independent variable influences health-state valuations ("basic impact model").

Address correspondence to: Jörn Moock, Institute of Community Medicine, University of Greifswald, Walther-Rathenau-Straße 48, 17475 Greifswald, Germany. E-mail: joern.moock@uni-greifswald.de

Attempts to explain these findings may draw on different lines of argument [15] relating to the different stages of interpretation, judgment, and response which follow exposition to the stimulus of one's own health state or respective scenario [16]. First, discrepancies might occur because of different interpretations of health state descriptions if, for example, comorbidities are neglected (or, alternatively, are missing in the scenario because of a lack of scope). Second, in the judgment phase a variety of effects can occur. A "focussing illusion" [17] can affect valuations if people forget obvious aspects of the health state under consideration. A "contrast effect" takes place if people are influenced by extremely negative life events that level the understanding of the valued health condition [18]. Furthermore, adaptation processes may explain differences in perception [19]: patients become accustomed to a chronic health condition that appears, from the outside, highly undesirable. Finally, in the response phase, recalibration response shift may happen in patients in that they change their internal standards.

Potentially, these aspects place obstacles to consistent decision-making and assessments given people with different health states are involved. Until now, rational and thus normatively consistent decision-making (under risk) has been predominantly based on a normative theory by von Neumann and Morgenstern [20] (i.e., Expected Utility Theory). For some decades, however, this has been increasingly challenged by less rigorous approaches that allow and incorporate "inconsistent" aspects in a more descriptive manner. Among these, the most influential descriptive theory of decision-making under risk is the so-called Prospect Theory by Kahneman and Tversky [21].

Prospect Theory distinguishes an editing and an evaluation phase. In the editing phase, outcomes are coded as gains or losses rather than as final asset positions. This coding is certainly
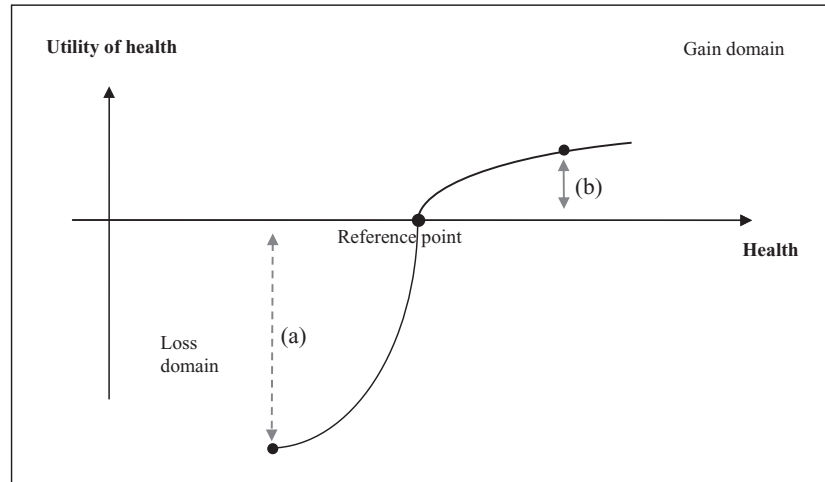
Figure 1 Utility differences depending on reference point. Situation (a) (dotted arrow): Utility differences (a) between health with and without disorder are substantial because of loss aversion if the reference point reflects health without disorder. Situation (b) (solid arrow): Utility differences (b) between health with and without disorder are small if the individual reference point reflects health with disorder.

influenced if, for instance, possible "focusing illusions" or adaptation processes have taken place beforehand. Gains and losses, however, are coded relative to some neutral reference point, which splits the evaluation space into a gain and a loss domain. Kahneman and Tversky assume that the shape and position of individual utility functions significantly depend on the position of the reference point. Specifically, the "gain function" is concave while the "loss function" is convex (see Fig. 1). Taking into account the observation that "losses loom larger than gains," labeled "loss aversion," the utility function is asymmetrical: the function for losses is steeper than the corresponding function for gains. In terms of utilities, equal deviations from the reference point are perceived more intensely in the loss than the gain domain, as shown in Figure 1.

If the status quo, that is, the present individual health condition, serves as the reference point, observed differences in utility of the same health condition between patients and the general public are explicable. Assume a participant has to rate a health condition worse than his present unaffected one. If he codes his present condition as the reference point, the worse condition will be coded as a loss, thus being on the steep loss function to the left of the reference point. The associated utility difference between both health conditions is quite substantial (situation (a) in Fig. 1). Assume on the other hand, the rated worse health condition is perceived as the reference point. In that case, the compared unaffected condition lies in the gain domain, thus on the less steep concave gain part of the utility function. Utility differences in the gain domain are less pronounced (situation (b)) than in the loss domain. It is hypothesized that the situation (a) in Figure 1 corresponds to the situation of unaffected people while situation (b) corresponds to the situation of patients.

Nevertheless, following Kahneman and Tversky, the location of the reference point can be affected by expectations, experience, or adaptation; thus, gains or losses can be coded relative to a point that differs from the status quo. Though plausible, a strict classification of patients and the general public with respect to different reference points does not seem justified. Possibly, patients have not yet adapted to their condition, or unaffected people can comprehend more or less how it is to suffer a particular impairment. In principle, the entire range of reference points may then be observed across and within both groups.

At any rate, the reference point is a potential explanatory variable for the impact of health status on health state valuations. In other words, we propose that reference points mediate the association of health status and health state valuation, and hence account for differences between people with a disorder and those without. As visualized in Figure 2, this implies the following four hypotheses, which relate to the four standard steps of mediational analysis first delineated by Baron and Kenny [22]:

Hypothesis 1: Health status (here: tinnitus vs. no tinnitus) has an effect on the respective health state valuations if the reference point is not considered (see Arrow 1 in Fig. 2); specifically, patients are predicted to assign higher utilities to their condition than unaffected people asked about the same condition. This represents the "basic impact model" described in the introductory paragraph of this article. Technically, this implies to show that the independent variable is correlated with the dependent variable (i.e., synonymously, the outcome) via modeling the latter by prediction through the independent variable. If significant, this establishes the very effect which may be mediated.

Hypothesis 2: Health status influences the reference point, that is, on average patients are predicted to refer to other points than unaffected people (Arrow 2). This implies to show that the independent variable is correlated with the potentially mediating variable via modeling the latter by prediction through the independent variable.

Hypothesis 3: Reference points influence health state valuations, that is, people with different reference points are predicted to value the same health condition differently (Arrow 3). This implies to show that the potential mediator is associated with the dependent variable via modeling the latter by prediction through the independent variable. Nevertheless, it is not sufficient just to link the mediator to the outcome since
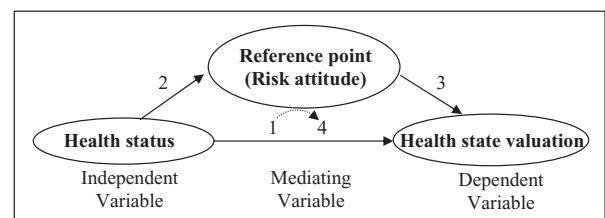
Figure 2 Mediated model of health status and health state valuation.

they may be correlated because they are both caused by the independent variable. Thus, the independent variable must be adjusted for in establishing the effect of the mediator on the outcome.

Hypothesis 4: The association of health status and health state valuations is lessened or even offset if the reference point (here: risk attitude) is adjusted for (Arrow 4). That is, risk attitude should explain a substantial part of the mean difference in utilities between patients and unaffected people. Complete mediation implies to show that the effect of the independent variable on the outcome adjusting for the mediator is zero. Technically, Hypothesis 3 and 4 are estimated in the same model.

Unfortunately, testing these hypotheses is not entirely straightforward because reference points are not consistently defined. In some cases, they have been used to ex plain results ex post but it is hard to define them ex ante. Nevertheless, Kahneman and Tversky [21] observed that "most people find symmetric bets [. . .] distinctly unattractive" (p. 280) if deciding something from the reference point while "a person who has not made peace with his losses is likely to accept gambles that would be unacceptable to him otherwise" (p. 288). People will be risk-seeking if the health state under consideration is perceived as a loss; people will be risk-averse if the health state coincides with their reference point. Thus, the relationship between risk attitude and reference point that is addressed in Prospect Theory suggests that the risk attitude with respect to health implies the position of the reference point in our context.

Against this background, we analyze the hypotheses stated previously for the case of tinnitus (i.e., "health status" in Fig. 2). The main symptom of this condition is described by Graham [23] "as a sensation of sound for which there is no source of vibration outside the individual" (p. 5). This impairment is quite common in industrial societies with about 40% of all adults having experienced temporary or permanent ear trouble, and 10% have to cope with it daily [24]. No personal characteristics can be linked to the appearance of tinnitus. Secondary symptoms are the main problem of tinnitus. Fifty-eight percent of affected Germans suffer from sleeping disorders, 38% have difficulty understanding conversations properly, and 36% are depressive or desperate [25]. The resulting stress can aggravate the situation leading to a "vicious circle."

Methodically, we apply three well-established direct valuation techniques to assign utilities to tinnitus-related quality of life (health state valuation in Fig. 2). Standard gamble (SG) is the theoretically most profound technique [26] and based on the axioms of von Neumann and Morgenstern [20]. Time trade-off (TTO) has been developed as an alternative measure to value health-related quality of life because many people have difficulties to understand the probability concept of SG [27]. Visual analogue scale (VAS) is a psychometric measure that is easily understood, and extensively used [28].

Methods

Sample and Procedure

Two equally sized samples of tinnitus patients and respondents from the general public (n = 210 each) were recruited and matched for sex and age. Patients were contacted at different places in Berlin, Germany: the German Tinnitus League (a self-help association), the Heinrich-Heine-Hospital (an institution focusing on psychosomatic conditions), an otolaryngology department (at Charité University Hospital), and a private clinic

for the treatment of tinnitus; they were interviewed between September and December 2000. The control group from the general population was interviewed at different public places in Berlin as well (predominantly randomly chosen pedestrian and shopping malls, and public transport localities) between October 2000 and January 2001 until matching to the patients sample was ensured. To be eligible for the study, respondents had to be Berlin residents (determined by their reported zip code). They were considered noneligible if they reported current tinnitus. After providing informed consent and agreeing to participate the interview was accomplished. All interviews were conducted by the same interviewer and took between 10 and 20 minutes.

Of those 210 who did participate from the general public, one person terminated the interview. Women and men are approximately equally represented in the patients (110 women and 100 men) and general public group (108 women and 102 men); also, age distributions are equivalent (for more information, see [29]). Regarding nonresponse to the different elicitation methods, overall 21 respondents were unable to answer SG or refused to do so (10 patients and 11 public respondents), while 30 did not reply to TTO (16 and 14, respectively), and five to VAS (2 and 3, respectively).

For the valuation tasks, participants unfamiliar with tinnitus listened to an example of tinnitus sounds which had been assembled by the German Tinnitus League in accordance with reports of patients. Also, a scenario description for tinnitus, developed jointly with experienced physicians and patients on the basis of the literature, was provided. Information about secondary symptoms and their prevalence among patients was explained as follows: "Permanent ear trouble is an increasingly bothering problem for many people. About one in ten in Germany already experienced permanent ear trouble in their life. Please imagine listening to permanent whistling or whooshing, hissing or pounding all day long. Interviews of patients revealed that more than half the participants have had sleeping disorders, and about a third communication, and concentration problems, depressions, and times of desperation, or despondence. Take your time to imagine such a situation" (note: English translation of original German version). SG and TTO were placed at the end of the interview to allow participants to "warm up"; VAS and assessments of age and sex among general public participants preceded them.

Elicitation Methods

Visual analogue scale. Following the procedure described by Gold et al. [1], participants were asked to place a mark for the condition "tinnitus" on a horizontal 100-mm line on or between the two anchor states "worst imaginable health state" and "best imaginable health state." They were not divided in "millimeter" or verbal clues to avoid memory effects and clustering [25]. The distance in millimeter (expressed in percent) between the lower anchor state and the mark was assumed to reflect the perceived severity of tinnitus on a scale between 0 and 1.

Standard gamble. With SG, respondents repeatedly face changing decision pairs until indifference is reached [26]. Specifically, participants were asked to envisage two sets of health-related circumstances that involve risky choices considering length of life: subjects chose either life with tinnitus or, as the second choice, a hypothetical treatment that either cured tinnitus with probability (p) or otherwise resulted in immediate death with probability (1-p). The individual utility score of tinnitus was determined by varying the level of P in a ping-pong mode until

Table 1 Valuations of tinnitus by between-subject factor HEALTH STATUS (tinnitus patients vs. controls without tinnitus) and within-subject factor METHOD (elicitation methods)*

Source of variation	SS	DF	MS	F	P
Tests of between-subject effects					
HEALTH STATUS	33,065.1	1	33,065.1	48.9	<0.001
Error	257,054.5	380	676.5		
Tests involving within-subject effect					
METHOD	379,605.5	2	189,802.8	773.1	<0.001
METHOD * HEALTH STATUS	10,763.6	2	5,381.8	21.9	<0.001
Error	186,597.6	760	245.5		
Simple effects of HEALTH STATUS within values of METHOD					
HEALTH STATUS within VAS	35,517.5	1	35,517.5	110.8	<0.001
Error	121,793.6	380	320.5		
HEALTH STATUS within TTO	2,577.9	1	2,577.9	5.2	0.023
Error	187,558.0	380	493.6		
HEALTH STATUS within SG	5,733.4	1	5,733.4	16.2	<0.001
Error	134,300.6	380	353.4		

*Repeated measure analysis of variance.
DF, degrees of freedom; MS, mean squares; SG, standard gamble; SS, sum of squares; TTO, time trade-off; VAS, visual analogue scale.

the participant remained indifferent. Participants were offered colored probability wheels, that is, color-coded pie chart segments, to facilitate the understanding of probabilities [30].

Time trade-off. In TTO, respondents balance life-years for better health. Two alternatives were offered, either an entire life (y) with tinnitus or a shorter period without.

The individual utility score of tinnitus was determined by varying the number of life-years spent disease-free (x) in a ping-pong mode until the participant remained indifferent. The maximum number of years the interview partner was willing to give up in relation to the original life expectancy determined the utility ratio (x/y). As Verhoef et al. remark, the aspiration level of survival seems to change with age [31]. To correct for individual aspiration levels considering life-years, individual life expectancy was operationally defined with the question of how old each participant guessed to become. The difference between individual life expectancy and actual age determined original life time.

Reference Points

As a proxy for the context-specific reference point, we used risk attitude with respect to tinnitus-related quality of life (variable "RISK" hereinafter). The operationalization of the risk attitude in the interview-based questionnaire was based on a corollary by Keeney and Raiffa [32]: "A decision maker who prefers the expected consequence of any 50–50 lottery . . . to the lottery itself is risk averse" (p. 150). The opposite holds in case of risk-seeking behavior. In our questionnaire, respondents were asked whether they were willing to accept a surgery or treatment for tinnitus that could either improve or worsen the condition, with an equal likelihood and to an equal extent. Literally, the item read as follows: "Would you undergo surgery or treatment that could improve or worsen your present health condition with Tinnitus, both with an equal probability and to an equal extent?" (English translation; German original available from the authors). Five possible answers: 1) "in no case," 2) "unlikely," 3) "maybe," 4) "likely," and 5) "in any case," were supposed to reflect five different risk attitude levels. Conceptually, the answers "in no case" and "unlikely" reflect risk aversion, while "likely" and "in any case" reflect risk seeking, and "maybe" is risk-neutral. Nevertheless, in order not to miss any potentially important variability in risk attitudes, the original five-point scale rated

by the study participants was used throughout the analyses reported hereinafter.

Statistical Analysis

To test Hypothesis 1, that is, for differences between the groups "tinnitus patients" and "controls without tinnitus" (variable "HEALTH STATUS") in tinnitus valuations based on the three different elicitation techniques (VAS, TTO, SG), a repeated measure analysis of variance was performed using the general linear model (GLM) function (SPSS for Windows, version 15.0.1 [SPSS Inc., Chicago, IL, USA] for this and all following analyses). In addition to the between-subject factor HEALTH STATUS, elicitation methods were entered as a within-subject factor ("METHOD") to account for this source of variance and to test the interaction, that is, whether HEALTH STATUS has differential effects on valuations depending on METHOD. Finally, contrast analysis in terms of simple effects [33] of HEALTH STATUS within each elicitation method was conducted using MANOVA design command options, thus testing Hypothesis 1 for each method. Table 1 and Figure 3 relate these analyses (see Results section).

Hypothesis 2, that is, differences in risk attitudes between participants affected versus unaffected by tinnitus, was explored by cross-tabulating HEALTH STATUS and RISK. As measures of association, a chi-square statistic was calculated and a test of mean differences in RISK reported in the text. Table 2 relates to the results of this analysis (see Results section).

Finally, Hypotheses 3 and 4 were tested in a repeated measure analysis of covariance, again using a GLM. In addition to the between-subject factor "HEALTH STATUS" and the within-subject factor "METHOD," risk attitude ("RISK") was entered in the equation as a covariate, thus being adjusted for. Again, contrast analyses in terms of simple effects within each elicitation method were conducted as described previously (see Hypothesis 1). Table 3 and Figures 4 and 5 relate to these analyses (see Results section).

Results

Hypothesis 1 predicted that patients assign higher utilities to their condition than unaffected people asked about the same condition ("basic impact model"; see Fig. 2, Arrow 1). Table 1
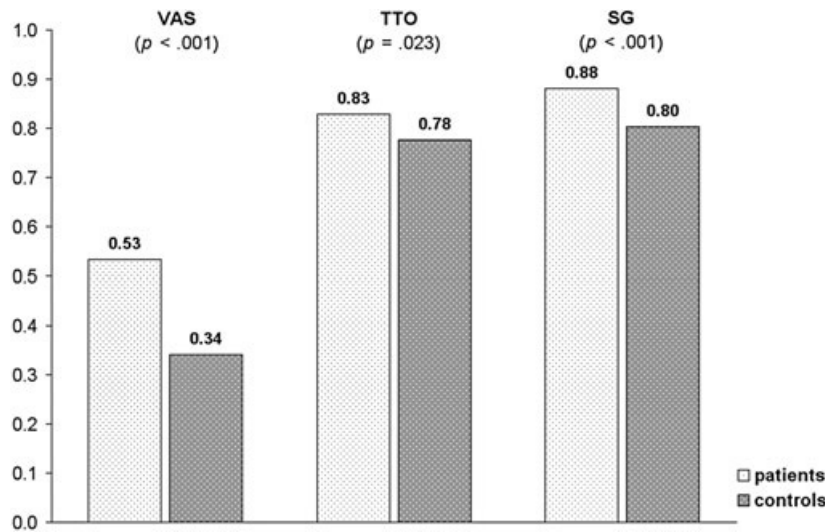
Figure 3 Utility differences depending on health status (tinnitus patients vs. controls without tinnitus), unadjusted for risk attitude. SG, standard gamble; TTO, time trade-off; VAS, visual analogue scale.

and Figure 3 show the results of the respective repeated measure analysis of variance and contrast analyses. As predicted, tinnitus patients differ significantly from controls both overall (HEALTH STATUS) and within each elicitation method (HEALTH STATUS within VAS, TTO and SG, respectively; see Table 1). As Figure 3 shows, within every method tinnitus patients assign higher utili-

ties to tinnitus as a health state than controls. This difference is most pronounced for VAS (0.53 vs. 0.34), followed by SG (0.88 vs. 0.80) and TTO (0.83 vs. 0.78). Furthermore, Table 1 indicates that both level of valuation (within-subject effect METHOD) and the differences between patients and controls (METHOD * HEALTH STATUS) vary significantly across

Table 2 Association of health status (tinnitus patients vs. controls without tinnitus) and risk attitude*

| HEALTH STATUS		In no case	Unlikely	Maybe	Likely	In any case	Total
				RISK risk attitude			
Patients	n	87	30	55	18	16	206
	row %	42.2	14.6	26.7	8.7	7.8	100
Controls	n	35	29	53	36	53	206
	row %	17.0	14.1	25.7	17.5	25.7	100
Total	n	122	59	108	54	69	412
	row %	29.6	14.3	26.2	13.1	16.7	100

*$\chi^2_{4,412} = 48.1$, $P < 0.001$.

Table 3 Valuations of tinnitus by between-subject factor HEALTH STATUS (tinnitus patients vs. controls without tinnitus), within-subject factor METHOD (elicitation methods), and RISK (risk attitude as a proxy for reference point) as covariate*

Source of variation	SS	DF	MS	F	P
Tests of between-subject effects					
RISK (covariate)	41,211.1	1	41,211.1	72.1	<0.001
HEALTH STATUS	11,314.8	1	11,314.8	19.8	<0.001
Error	214,281.2	375	571.4		
Tests involving within-subject effect					
METHOD	372,776.2	2	186,388.1	761.6	<0.001
METHOD * HEALTH STATUS	10,458.6	2	5,229.3	21.4	<0.001
Error	184,037.8	752	244.7		
Simple effects of HEALTH STATUS within values of METHOD					
RISK (covariate) within VAS	4,030.6	1	4,030.6	12.9	<0.001
HEALTH STATUS within VAS	24,686.0	1	24,686.0	79.2	<0.001
Error	116,833.0	375	311.6		
RISK (covariate) within TTO	8,823.7	1	8,823.7	18.7	<0.001
HEALTH STATUS within TTO	396.1	1	396.1	0.8	0.360
Error	176,614.3	375	471.0		
RISK (covariate) within SG	37,711.4	1	37,711.4	148.1	<0.001
HEALTH STATUS within SG	52.1	52.11	1.0	0.2	0.631
Error	95,517.2	375	254.7		

*Repeated measure analysis of covariance.
DF, degrees of freedom; MS, mean squares; SG, standard gamble; SS, sum of squares; TTO, time trade-off; VAS, visual analogue scale.
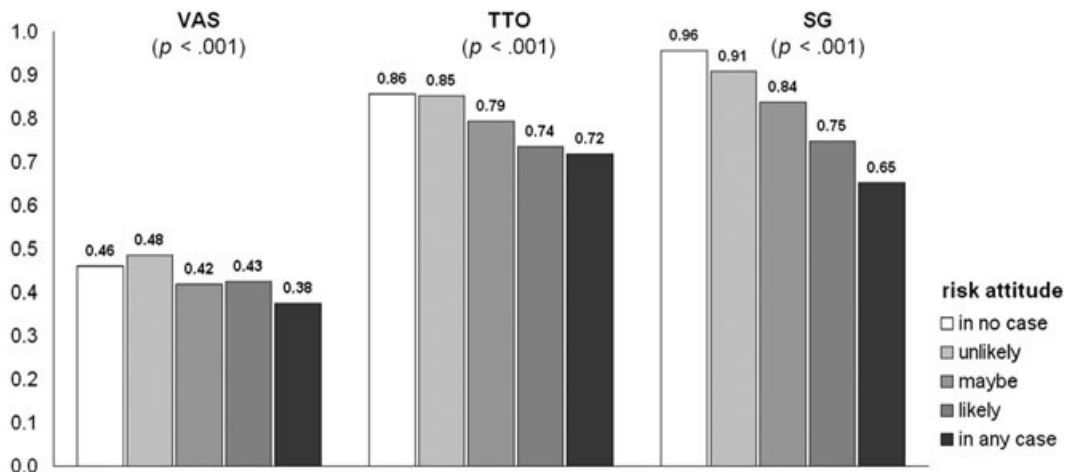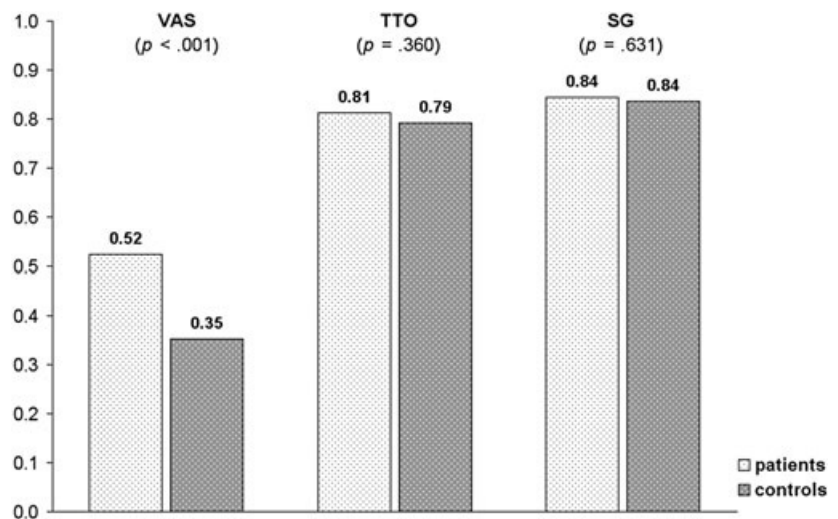
Figure 4 Utility differences depending on risk attitude, adjusted for health status (tinnitus patients vs. controls without tinnitus). SG, standard gamble; TTO, time trade-off; VAS, visual analogue scale.

methods. This latter finding indicates that the degree of discrepancy between health state valuations of patients and the general public depends on which specific elicitation method is utilized.

Table 2 shows the test of Hypothesis 2, stating that on average patients refer to other points than unaffected people (see also Fig. 2, Arrow 2). The majority of patients would "in no case" (42.2%) or "unlikely" (14.6%) accept an intervention that may improve or—with the same probability and to the same extent—worsen their situation, while the majority of tinnitus-unaffected controls from the general public would more often do so "likely" (17.5%) or "in any case" (25.7%). Both the results of the cross-tabulation (chi-square) and an additional analysis of variance not shown in the table ($F_{1,375} = 51.3$, $P < 0.001$) underline that this difference is most probably not due to chance variations.

Finally, Table 3 and Figures 4 and 5 present the results of the repeated measure analysis of covariance for the mediated model, that is, for Hypotheses 3 and 4 (see also Arrows 3 and 4 in Fig. 2). As regards Hypothesis 3, the risk attitude introduced as a covariate (RISK) significantly predicts valuations both overall and within methods (RISK within VAS, TTO and SG, respectively; see Table 3). As shown in Figure 4, the largest range can be observed for SG, with a utility score of 0.96 if the risky intervention would be accepted "in no case," and 0.65 if the answer was "in any case." For TTO, utilities range from 0.72 to 0.86 and follow a linear pattern similar to SG. Regarding VAS, results are not so pronounced and linearly patterned, but it can be seen that risk-averse groups generally elicited higher values (0.46 and 0.48) than the risk-seeking respondents (0.43 and 0.38). Finally, coming to Hypothesis 4 (see Arrow 4 in Fig. 2), that is, that the impact of health status and health state valuations is lessened or even offset if risk attitude is adjusted for, results show that indeed this is the case. Most prominently, compared to Table 1 the simple effect analyses show that significant differences between patients and the general public sample disappear for TTO and SG, and are lessened for VAS. Numerically, it can be seen in Figure 5 that this is most striking for SG, with regard to which the estimated means for patients and controls virtually converge at 0.84. Likewise, the mean difference is now diminutive for TTO as well (0.81 vs. 0.79), while for VAS, this difference is scaled

Figure 5 Utility differences depending on health status (tinnitus patients vs. controls without tinnitus), adjusted for risk attitude. SG, standard gamble; TTO, time trade-off; VAS, visual analogue scale.

down only marginally, even thought the F-value has been reduced from 110.8 (see Table 1) to 79.2 in the present analysis (see Table 3).

Taken together, in line with the mediated model in Figure 2 the effect of being affected versus unaffected by tinnitus (i.e., HEALTH STATUS) on tinnitus valuations is fully mediated by risk attitude (RISK) as a proxy for reference points for the elicitation methods TTO and SG, and to a minor extent for VAS.

Discussion

The present results suggest that health state-specific reference points defined as risk attitudes play an important role for health state valuations. Utility differences between tinnitus patients and control respondents without tinnitus are considerable for all three elicitation methods VAS, TTO, and SG in that patients assign higher values to tinnitus, indicating better health [29]. Nevertheless, significance is statistically lessened for VAS, and disappears for TTO and SG, if risk attitude is adjusted for. Since simultaneously, both health status significantly predicted risk attitudes (patients more risk-averse than controls) and risk attitude predicted valuations (higher utilities in the risk-averse groups), utility differences between patients and the general public can be attributed to different reference points—given, of course, that it is correct to assume risk attitudes as indicative of, or surrogate to, reference points. As such, risk attitude could be explored as an adjusting variable for raw utility scores.

We concede, however, that the risk attitude approach has some difficulties. First, responses may in part reflect understandings of the intervention's effectiveness, disutility associated with the condition if untreated, and any disutility attached to undergoing the procedure itself. Second, and conceptually probably even more important, risk attitude is usually defined when outcomes are numerical data. The approach applied here attempts to measure risk intensity without being able to specify the deviation from the status quo (only "equal extent"). Nevertheless, we think this is the closest one that can get to Keeney and Raiffa's [32] definition of risk in the health field if quasi-intervals (e.g., days of illness, as in Stalmeier et al. [34]) are to be avoided. Third, we assessed reference points with a risk scale that explicitly mentioned the investigated health state, tinnitus. On one hand, it should be stated that persons may be risk-seeking in one context and risk-averse in another, and that risk attitudes elicited by our item were comparable across subjects because the context was fixed (i.e., tinnitus). On the other hand, as this measurement of risk attitudes is fairly similar to SG, it may be of little surprise to find a strong correlation, even though SG considers risk with respect to length of life, while the risk question used as a proxy to the reference point considers quality of life. Nevertheless, this kind of measurement does in fact comply with the framework and assumptions of Prospect Theory. Plus, and more importantly, results for TTO resemble those for SG—and strike us as much less straightforward. In all, this points to the original interpretation of risk attitude as reference point rather than undue conceptual overlap with health valuations. Finally, TTO and SG are quite similar considering the trade-off aspect, as reflected in their correlation among each other and with risk attitude.

Some limitations of our study warrant discussion. First, the results of the general population were based on data gathered from a street sample and one has to question the extent to which this might have affected the data. Careful checks were employed to ensure that the patients and the street sample were matched for sex and age to minimize bias. Second, widely used measures of health utilities as SG and TTO were performed in our study.

These are direct instruments, involving gambling on a hypothetical medication that may cause perfect health or death (SG), or trading off part of future life reduced time in perfect health (TTO). Indirect methods of obtaining health utilities involve reports of current health on a standardized questionnaire such as the parsimonious EQ-5D, the SF-6 dimensions, or the Health Utility Index Mark 2 or 3. Nevertheless, such indirect measures of health utilities are not included in our study. Further research may address comparisons on both direct and indirect measures of health utilities.

Also, the limitations of the present study advise that further research is needed. Starting with the investigated health condition, results need to be replicated for other diseases to be able to generalize the findings. Also, tinnitus severity as an essentially subjective experience depends crucially on descriptions of patients, which makes it difficult to find a valid description of tinnitus that can be adequately understood and appreciated by the general public. In our study, the general public valued the tinnitus simulation via a recording, and there was no representative sample of sounds and loudness of tinnitus beyond this recording available. In other words, an "average" description was applied which does not necessarily reflect the entire range of potential tinnitus impairments. Hence, differences in these simulations to the tinnitus suffered by participating patients may also partly explain our findings.

Results underpin the importance of Prospect Theory as a descriptive framework for decision analysis (though not all tests of Prospect Theory have shown results that it works as predicted [35]). Furthermore, the study is a first attempt to measure the reference point that is crucial to eliminate utility differences. Of course, several aspects can influence valuations. To name only one in the context of Prospect Theory, Kahneman and Tversky find that low probabilities are overstated while higher probabilities are understated. Especially in the range close to 1, utilities elicited with SG are biased upward because people need to overstate the associated probability to express their "true" preferences [36]. Although loss aversion is said to be mainly responsible for deviating results, this probability weighting could be equally applied as an explanation. It will be necessary for future research to distil each impact variable accurately.

Nevertheless, if further investigations confirm that the explanatory model proposed here bears further conceptual and empirical validity, the question remains whether these results have any consequences for medical decision-making. Setting priorities for recommendations which medical interventions should be supported is a normative exercise, vaguely related to utility maximization [37]. It is of less value to know how people do act but how they should act to allow a maximization of health in priority setting. Otherwise, priorities are based on cost-effectiveness ratios that apply descriptively determined preferences. An understanding of descriptive decision-making is therefore essential—an endeavor which Prospect Theory allows for. Utilities should be based on descriptive techniques which state how people do act.

Against this background, present prioritization in the health domain should be reconsidered. Ordinal rankings change if measures of effectiveness depend on different reference points. At least some interventions near defined decision thresholds could—depending on the "misperception" of the general public—very much be subject to the adjustment procedure. As a result, the necessity of different medical interventions may be misinterpreted. The nature of the reference point is a crucial issue for future analysis if evaluations aresupposed to properly reflect effectiveness. Also, this should be elucidated using different elicitation methods.

Acknowledgments

References

1 Gold MR, Siegel JE, Russell LR, et al. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.

2 Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? Soc Sci Med 1993;37:33–40.

3 Boyd N, Sutherland H, Heasman K, et al. Whose utilities for decision analysis? Med Decis Making 1990;10:58–67.

4 Dolan P. The effect of experience of illness on health state valuations. J Clin Epidemiol 1996;49:551–64.

5 Dolan P. Whose preferences count? Med Decis Making 1999;19: 482–6.

6 Llewellyn-Thomas H, Sutherland H, Thiel EC. Do patients' evaluations of a future health state change when they actually enter that state? Med Care 1993;31:1002–12.

7 Sackett D, Torrance G. The utility of different health states as perceived by the general public. J Chronic Dis 1978;31:697–704.

8 Tarn T, Dix Smith M. Pharmacoeconomic guidelines around the world. ISPOR Connections 2004;10:5–15.

9 Torrance G. Measurement of health state utilities for economic appraisal: a review. J Health Econ 1986;5:1–30.

10 Llewellyn-Thomas H, Sutherland H, Tibshirani R, et al. Describing health states: methodologic issues in obtaining values for health states. Med Care 1984;22:543–52.

11 Jansen S, Stiggelbout A, Wakker P, et al. Unstable preferences: a shift in valuation or an effect of the elicitation procedure? Med Decis Making 2000;20:62–71.

12 Kind P, Dolan P. The effect of past and present illness experience on the valuations of health states. Med Care 1995;33:AS255–63.

13 Llewellyn-Thomas HA, Thiel EC, McGreal MJ. Cancer patients' evaluations of their current health states: the influences of expectations, comparisons, actual health status, and mood. Med Decis Making 1992;12:115–22.

14 Lenert L, Treadwell J, Schwartz C. Associations between health status and utilities: implications for policy. Med Care 1999;37: 479–89.

15 Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. Qual Life Res 2003; 12:599–607.

16 Stiggelbout AM, de Vogel-Voogt E. Health state utilities: a framework for studying the gap between the imagined and the real. Value Health 2008;11:76–87.

17 Legrenzi P, Girotto V, Johnson-Laird PN. Focussing in reasoning and decision making. Cognition 1993;49:37–66.

18 Tversky A, Griffin D. Endowment and contrast in judgments of well-being. In: Argyle M, Schwarz N, Strack F, eds. Subjective Well-Being: An Interdisciplinary Perspective. Oxford: Pergamon, 1991.

19 Adang E. A comment on "response shift and adaptation in chronically ill patients" by Postulart and Adang. Med Decis Making 2001;21:423–4.

20 von Neuman J, Morgenstern O. Theory of Games and Economic Behaviour (3rd ed.). Princeton, NJ: Princeton University Press, 1953.

21 Kahneman D, Tversky A. Prospect Theory: an analysis of decision under risk. Econometrica 1979;47:263–91.

22 Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations. J Pers Soc Psychol 1986;51:1173–82.

23 Graham J. Tinnitus Aurium. Uppsala: Almquist & Wiksells, 1965.

24 Feldmann H. Tinnitus: Grundlagen Einer rationalen Diagnostik und Therapie. Stuttgart: Georg Thieme Verlag, 1998.

25 Goebel G. Fortschritte bei der verhaltensmedizinischen Diagnostik und Behandlung quälender chronischer Ohrgeräusche. Otorhinolaryngol Nova 1995;5:178–89.

26 Torrance G. Measurement of health state utilities for economic appraisal. J Health Econ 1986;5:1–30.

27 Torrance G, Thomas W, Sackett D. A utility maximization model for evaluation of health care programs. Health Serv Res 1972;7: 118–33.

28 Torrance GW, Feeny D, Furlong W. Visual analog scales: do they have a role in the measurement of preferences for health states? Med Decis Making 2001;21:329–34.

29 Happich M, von Lengerke T. Valuing the health state "tinnitus": differences between patients and the general public. Hearing Res 2005;207:50–8.

30 Drummond M, O'Brien B, Stoddart G, et al. Methods for the Evaluation of Health Care Programmes. New York: Oxford University Press, 1997.

31 Verhoef LC, de Haan AF, van Daal WA. Risk attitude in gambles with years of life: empirical support for prospect theory. Med Decis Making 1994;14:194–200.

32 Keeney R, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. New York: Wiley & Sons, 1976.

33 Levine G. A Guide to SPSS for Analysis of Variance. Erlbaum: Hillsdale, 1991.

34 Stalmeier PF, Bezembinder TG. The discrepancy between risky and riskless utilities: a matter of framing? Med Decis Making 1999;19:435–47.

35 Feeny D, Eng K. A test of prospect theory. Int J Technol Assess Health Care 2005;21:511–18.

36 Bleichrodt H. Probability weighting in choice under risk: an empirical test. J Risk Uncertainty 2001;23:185–98.

37 Garber A, Phelps C. Economic foundations of cost-effectiveness analysis. J Health Econ 1997;16:1–31.