Whose Preferences – a Matter of Reference Points?

By Michael Happich

Abstract

Background. Whose preferences are to use for Cost-effectiveness-analysis? It has been recommended that community preferences for health states are the most appropriate ones for use in a Reference Case analysis. But critics say that people are not able to properly judge a health state if they have not experienced that condition. This problem has been analysed in the framework of Prospect Theory. It can be argued that different reference points of patients and the general public are responsible for deviating results. In addition, it is argued in this paper that risk attitudes with respect to health-related quality of life are an indicator for reference points. If patients and the general public refer to the same reference point, i.e. they have the same risk attitude, the hypothesis is that deviations are not significantly different any longer. **Methods**. Evaluations of the health condition Tinnitus of 210 patients and 210 unaffected people have been compared. Time Tradeoff and Standard Gamble method have been applied to elicit preferences. Risk attitude has been measured with the question whether participants would undergo a treatment that could either improve or deteriorate their health condition, both with an equal chance (five possible answers between 'in no case' and 'in any case'). Risk attitude and mean evaluations have been compared with the non-parametric Mann-Whitney-U test. **Results**. Affected people indicated significantly higher values for Tinnitus-related quality of life according to the Standard Gamble method comparing 87.93 on average to 80.67 for unaffected people (asymptotic significance < 0,001). The difference between Time Tradeoff values is less dramatic but still significant with, on average, 82.68 for affected people compared to 78.02 for unaffected people (asymptotic significance < 0,05). Affected people are more risk-averse than affected people (asymptotic significance < 0,001). However, significance disappears if mean evaluations in single risk groups (those who answered 'in no case' for example) are compared (asymptotic significance > 0,4 in most cases). **Conclusion**. Prospect Theory is a reasonable framework to tackle the question about whose preferences count. However, community preferences can be used, indeed, if they are corrected for risk attitude with respect to health-related quality of life.

Key Words: health state evaluation, reference points, patients versus general public

Introduction

The question of providing necessary or efficient health care is important in today's context of rising financial pressure on health care systems. The utilitarian philosopher Hutcheson argued "that action is best which procures the greatest happiness for the greatest numbers." In the view of Hutcheson's fellow utilitarian Bentham, any action is right which increases general happiness. Cost-effectiveness analysis can contribute to maximization in the health domain by simply comparing costs with the outcome of a medical intervention. Those interventions with the most favourable cost-effectiveness ratio are served first until a given budget is spent. To properly determine effectiveness in this ratio, the evaluation of health state utilities has recently gained importance in the literature.

However, there is much debate about whose utilities should count [6, 7, 25]. Who decides what medical intervention is really needed? Gold et al. [9] recommend that "community preferences for health states are the most appropriate ones for use in a Reference Case analysis." Kaplan [14] supports this, noting that "preferences should represent the will of the general public."

Intuitively, one should expect that persons affected by a certain health condition are much better prepared to judge the suffering that it entails. Gold et al. [9], however, point to the "veil of ignorance" in claiming just the opposite of this. They state that it is most appropriate to aggregate the utilities of the rational public because it is blind to its own self-interest. Hadorn [11] supports this point of view, believing in a kind of utility-maximizing behaviour on the part of patients, especially when the payer of the medical bills and receiver of medical treatment are not identical. Another argument is that it does not really matter who one asks as long as any illness is sufficiently explained. Stable utilities allow the substitution of information for personal experience [18], i.e., any rational decision maker can base his or her judgement on knowledge of a health condition and come to similar average evaluations as patients. Some empirical evidence does support this [2, 17, 18].

On the other hand, many authors find deviations. This deviation bias generally favours the affected group [6, 16, 20, 21, 22], meaning that the general public assumes a certain health condition to be more serious than do the patients themselves. Personal experience of the illness does matter under such circumstances and can influence any priority setting. They can change the ordinal ranking of prioritised treatments in league tables and, depending on the bias, decide whether one or the other health condition is more strongly supported by society. Looking at this problem in detail, various authors (e.g., [16, 25]) have contributed these deviations to Prospect Theory [12]. This theory is also the theoretical foundation of the present analysis.

Prospect Theory

In opposition to normative theories of decision, the descriptive Prospect Theory assumes individual reference levels to exist that severely influence health state evaluations. If individual health states are taken as the status quo representing the reference level, the reason for deviations are threefold: First, depending on the reference level, a decision maker codes improvements in health as gains and deteriorations as losses. Secondly, value functions are concave in the gain domain but convex in the loss domain. Third, the convex part is steeper than the concave part [12] (see the Fig. 1).

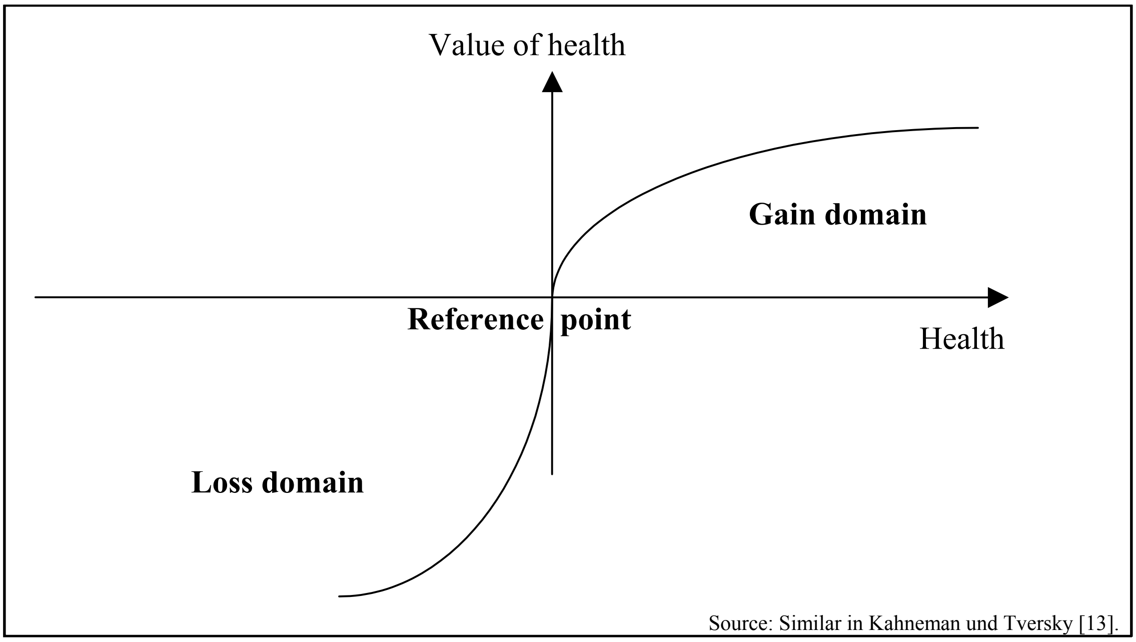

Figure 1. **Value function according to Prospect Theory**.

Hence, the overall utility function of an unaffected person is more convex than those of affected persons, who have much more to gain. This may explain the differing perceptions of the same health condition. Healthy persons should judge a condition much more severely, as shown in Fig. 2.

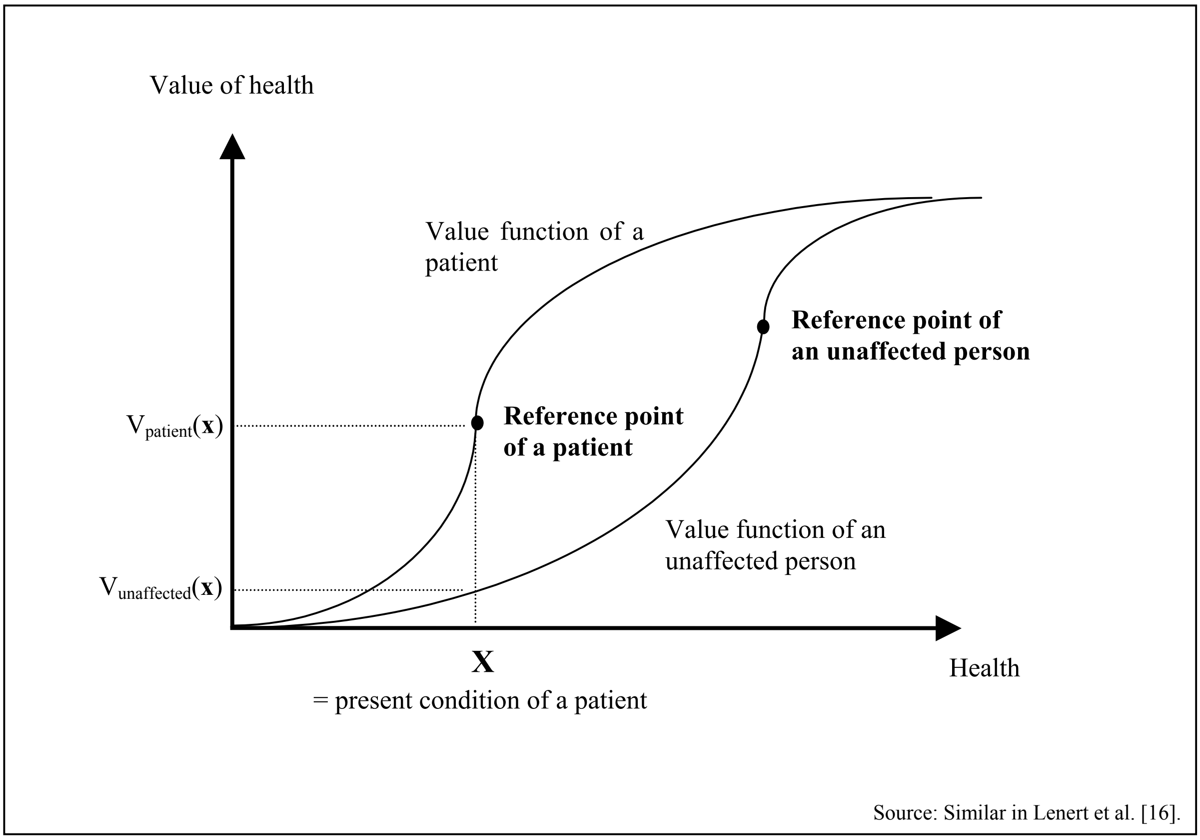

Figure 2. **Impact of Prospect Theory on evaluations.**

Patients and unaffected persons have differing perceptions of the same health state because they have different reference points. The point of reference for an affected person lies lower (and more to the left) than that for an unaffected person. The characteristics of Prospect Theory lead to the typical depicted value curves. The valuation of a health condition X is different.

This leads to the first hypothesis: On average, affected persons rate their own health state less as severe than unaffected persons do. Lenert et al. [16] have discussed this solution and tried to gather evidence in favour of Prospect Theory. They were only partially successful. They considered similar sources of evidence as in the general debate of whose preferences are to be used and found the same mixed results. The only point which they considered to be something of a verification was the lower rating by the general public than by patients. Lenert et al. [16] did not consider relative curvatures of utility functions. These are interpreted as risk attitudes. Convex curves represent risk-seeking behaviour and concave curves risk aversion [15]. What does this mean in the context of Prospect Theory? Unaffected persons evaluate any severe health condition on the convex part of their value function. Assuming a lottery between a treatment that could leave them either better off or worse off- with equal probabilities - or staying in the present condition, they should prefer risk seeking and thus favour treatment. Patients are already at the reference point. Considering the same lottery, their probable improvement is less than their probable loss since the loss function is steeper than the gain function. They should choose against treatment, indicating risk averse behaviour. Therefore, the second hypothesis is: On average, affected persons are more risk-averse than unaffected persons.

The second hypothesis implies that the risk posture can be considered as an indicator of the location of the reference point. Hence. If both hypothesis are influenced by considerations of Prospect Theory, the two referred attributes, health-related quality of life and risk attitude, should be correlated. It is argued above that affected persons evaluate their own condition as less severe while being risk averse in contrast to unaffected persons. Leaving the health-related dichotomy, it can be hypothesized in general that risk-averse persons evaluate their own condition as being less severe while risk-seeking persons consider it more dramatic. This is our third hypothesis: Risk posture correlates with average evaluations. Finally, assuming that risk attitude is the main indicator of deviations between evaluations of affected and unaffected persons, these differences should diminish if persons with the same risk attitude are compared independently of health status, i.e., those with the same reference point evaluate a given health condition similarly. This is the fourth hypothesis.

Value and utility

In theories about decision making, two fundamentally different measurement approaches are generally used to model preferences. The first, founded on difference measurement [23], asks for judgements about strength of preference to derive a value function, v [4]. The second approach to decision making uses preferences among gambles to construct a utility function, u [4]. Hence the term "value" describes preferences derived under certainty while "utility" describes preferences derived under uncertain conditions [2]. Prospect Theory originally refers to value functions whereas the evaluation methods used in this study, namely Standard Gamble (SG) and Time Tradeoff (TTO), refer to utility functions for axiomatic reasons [14]. Reports in the literature have already analysed and tried to combine value and utility (e.g., [4]). Bell and Raiffa [5] say that "it is well known, of course, that a utility function is a bona-fide value function but not the converse." We follow this approach because, according to Bamberg and Coenenberg [3], a utility function can theoretically be split into a value function and an intrinsic risk part. Therefore, anything valid for a value function influences the overall utility function. A changing curvature of the value function bends the overall utility function as well.

Methods

The disease

The health condition which we use in the comparison is tinnitus. The first symptom of this condition is commonly known as a sound in the head. Its characteristic feature is described by Graham [10]: "Tinnitus may be defined [further] as a sensation of sound for which there is no source of vibration outside the individual." The causes of tinnitus are manifold. Organic damages can lead to a permanent stimulus. Menière's disease, hearing loss, blockade of the vertebral column, or metabolic problems may be the cause. Up to 800 causes have been adduced to explain tinnitus [8]. And almost as many treatments as causes are available. But Graham [10] states: "Many cases of recovery are enumerated, but one wonders if these were cures or rather the patients' effort to bring a stop to the treatment."

Unaffected participants were informed of the primary symptoms of tinnitus. To simulate possible sounds and volume levels, participants listened to a recording of sounds produced by a synthesizer and simulating descriptions provided by affected persons. In addition, participants were told of possible secondary symptoms such as sleeplessness, ear trouble, depression, concentration problems, and particularly the inability to cope, since this is a crucial aspect of living with the condition [8]. Participants were asked to imagine such a state and think about their own possible ability to cope. Affected persons were invited to describe their own condition, i.e., individual sounds and secondary symptoms. Interview questions were then put to them.

Health-related quality of life measures

There are numerous ways to measure preference-based, health-related quality of life. One method for evaluating individual health perception asks the maximum number of years that participants would be willing to sacrifice in order to free themselves of the symptoms of tinnitus [24]. A hypothetical medicine is described which may have the effect of freeing the sufferer from symptoms, but which also affects life expectancy. The number of years that are to be sacrificed is progressively increased until the respondent is undecided between taking the medicine and living with the condition. The ratio between remaining and actual life expectancy yields a value between 0 and 1 (or normed between 0 and l00) and defines the individual quality of life of that health condition. This procedure is the TTO method.

A similar utility-based method is the SG [24]: Respondents are asked to state their indecision point of survival probability for a hypothetical operation that would remove any signs of tinnitus. Starting with 100 survival probability, figures are successively lowered until the participant can no longer clearly state whether he or she would take part or refuse such operations. The stated probability determines a point on a scale between 0 and 100 that describes the individually experienced or imagined quality of life of the health condition.

TTO and SG have been widely applied in the literature, and there is considerable concern about whether they actually measure the same phenomenon [19]. For example, much debate has surrounded the incorporation of risk in the method. Considering life years, SG is thought to include such risks in the question while TTO does not. However, our main concern is not about differences between measures but the relevance of the hypotheses on each measure.

On the other hand, TTO and SG deal with two attributes - length of life and quality of life, i.e., both take length of life as the form in which health-related quality of life is measured. We are concerned about risk attitudes of persons considering the second attribute derived from the hypotheses. While the difference between TTO and SG could tell us something about the risk attitude towards the first attribute, length of life, nothing is said about the second. To obtain data about this risk posture we use a corollary provided by Keeney and Raiffa [15]: "A decision maker who prefers the expected consequence of any 50-50 lottery [...] to the lottery itself is risk averse." The opposite holds for risk seeking behaviour. Respondents are asked whether they are willing to accept an operation that can either improve or worsen their health condition, the two having with an equal probability. Answers as to the intensity of risk aversion or seeking are given on a five-point rating-scale with categories: in no case, 1; unlikely, 2; maybe, 3; likely, 4; in any case, 5.

These risk groups can be considered as an indicator of the location of the reference point. In addition, in the questionnaire-based interview tinnitus patients and unaffected persons were asked to evaluate the relationship between life expectancy and willingness to exchange (expected) life years for better health. To define individual life expectancy all participants were asked how old they expected to become. The difference between individual life expectancy and actual age can be defined as remaining life expectancy. This procedure allows reference point biases to be avoided considering life years [26].

The analysis

The first and second hypotheses are investigated by the Mann-Whitney U test, a nonparametric procedure to compare mean values of two groups, in our case affected and unaffected persons. The correlation of risk attitude towards quality of life and mean evaluations in the third hypothesis is investigated with Spear-man's Rho, a nonparametric test of correlation. The fourth hypothesis is also analysed by the Mann-Whitney U test. Mean evaluations of affected and unaffected persons are compared but this time within the five risk groups that are considered to determine different reference points. Returned interview questionnaires were coded. The data were analysed using SPSS software. Numbers not clearly recognizable were coded as missing values.

Participants

A total of 210 patients were interviewed between September and December 2000 (110 women, and 100 men; age 16-85 years, mean 54). Patients were met at four different places in

Berlin (21 at the Tinnitus League, a self-help association; 21 at the Heinrich-Heine Hospital, a hospital with a focus on psychosomatic conditions; 63 at the Ear, Nose, and Throat Department of the Charité University Hospital, and 105 patients of Dr. Berndt, a leading expert in tinnitus treatment). In addition, 210 unaffected persons were interviewed between October 2000 and January 2001 (108 women, 102 men; age 13-81 years, mean 54). Participants were met at four different places in Berlin (46 at Kaiser's Supermarket in Kreuzberg, 57 at the main railway station, 52 at Ring-Center, and 55 at Kaufhof Shopping Center in eastern Berlin). Demographic details of both groups are presented in Table 1.

		Tinnitus patients (N=210)		Tinnitus not affected (N=210)	
		number	column percent	number	column percent
Gender	male	100	47,6	102	48,6
	female	110	52,4	108	51,4
Marital status	married	146	69,5	142	67,6
	single	24	11,4	24	11,4
	widowed	14	6,7	15	7,1
	divorced/ seperated	26	12,4	29	13,8
Years of school attendance	less than 10 years of school	109	54,3	119	56,7
	more than 10 years of school	101	45,7	91	43,3
Occupation	student	1	0,5	5	2,4
	worker	17	8,1	18	8,6
	civil servant	10	4,8	10	4,8
	employee	70	33,3	56	26,7
	self-employed	9	4,3	14	6,7
	housewife	3	1,4	3	1,4
	pensioner	79	37,6	83	39,5
	unemployed	14	6,7	15	7,1
	other	7	3,3	6	2,9

Table 1. **Demographic characteristics of tinnitus patients and unaffected persons in our sample.**

Out of 420 participants, 21 persons responded that they could not answer the SG question or refused to do so (10 affected, n unaffected), and 29 did not answer to the life expectancy and TTO question (16 affected, 13 unaffected). However, only 8 persons did not answer to the question about operation risks (4 affected, 4 unaffected). One of the participants broke off the interview.

Results

The hypotheses

Affected persons indicated substantially higher values for tinnitus-related quality of life according to the SG method (87.93 vs. 80.67). The difference between TTO values was less dramatic but still considerable (82.68 vs. 78.02). The Mann-Whitney U test confirms that these differences are significant. The asymptotic significance was close to zero, i.e., the probability of incorrectly assuming a difference although there is none is very low. The exact figures are 2.7 probability for TTO scores and less than 0.005 for SG scores. These results are in accordance with the first hypothesis.

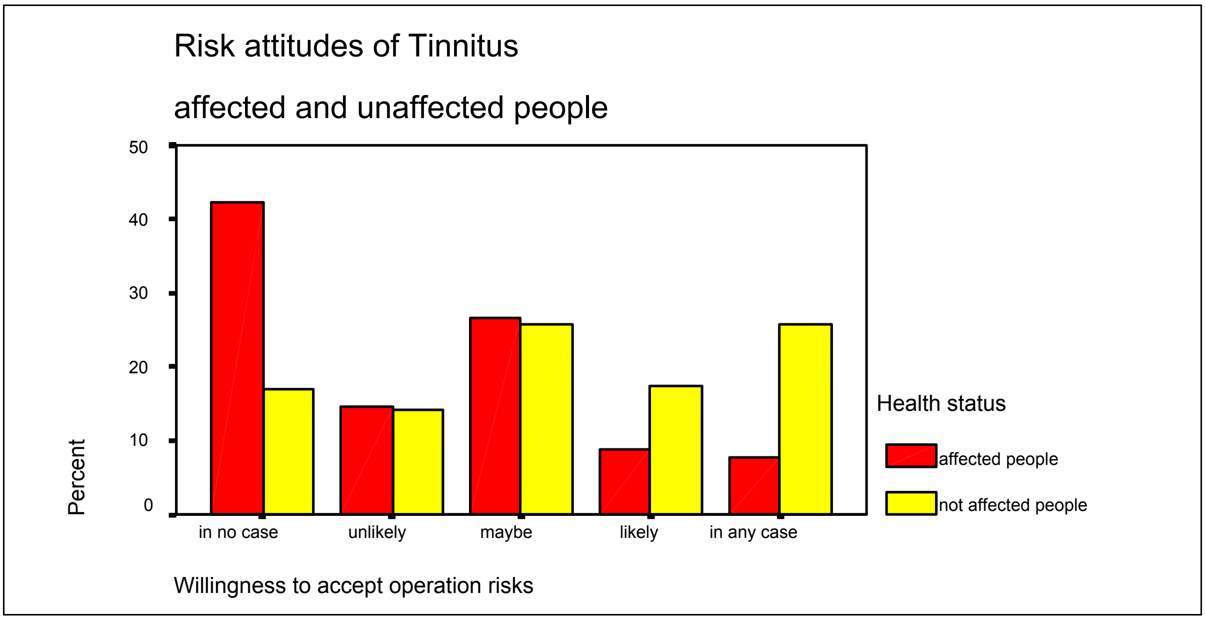

Figure 3. **Risk attitudes with respect to health-related quality of life of tinnitus patients and unaffected persons.**

The affected and unaffected groups answered significantly differently on the question of a possible operation that could either improve or deteriorate the tinnitus condition. The majority of those with tinnitus would not (42) or probably not (15) be willing to accept an operation. This is in contrast to unaffected persons, only about 30 of whom would avoid such an operation if they had to face tinnitus (Fig. 3).

The Mann-Whitney U test confirmed the statistical significance of this finding. The asymptotic significance was close to zero, i.e., the probability of incorrectly assuming a difference although there is none, was less than 0.005. These results are also in accordance with the second hypothesis.

The next question is whether risk attitudes and mean evaluations are correlated. Using SG as an example, Fig. 4 shows that the more risk-averse persons, regardless of whether affected or unaffected, have higher mean scores ($P < 0.005$, Spearman's Rho). This was the case with both SG and TTO. This confirms the third hypothesis: The more willing persons are to accept risks, the lower is their evaluation of tinnitus.

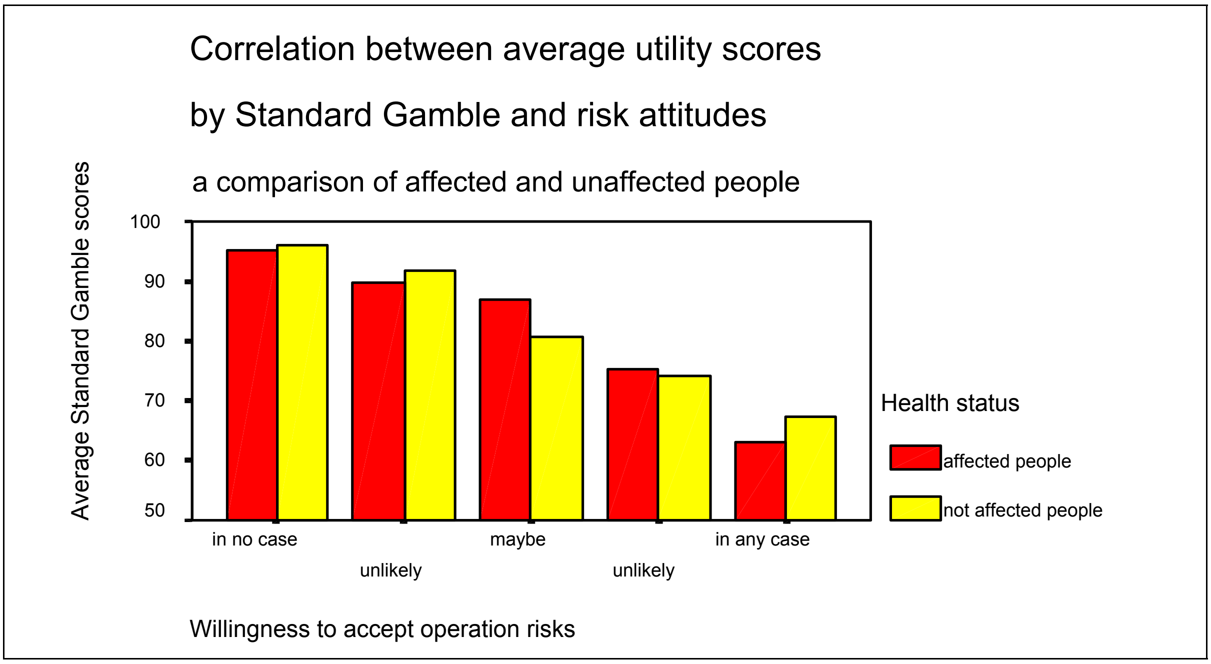

Figure 4. **Correlation between mean values for tinnitus and willingness to accept operation risks.**

However, within risk groups differences were seldom significant. Scores of tinnitus patients were not consistently higher than those of the unaffected group. For example, patients who said that they would "in no case" take part in an operation, evaluated tinnitus on average almost identically as the unaffected group (Fig. 4), as in other risk groups as well. Differences were not significant ($P>0.05$) on either measure (Table 2). Even at the significance level of 10, differences between patients and unaffected persons were statistically significant only in one group (SG-maybe). Although no proof is possible that evaluations are equal in general, the result is seen to support the fourth hypothesis: Considering the same reference level (the same risk attitude), the two groups come to similar evaluations.

Risk group	Standard Gamble	Time Tradeoff
In no case	0,630	0,425
Unlikely	0,133	0,856
Maybe	0,060	0,419
Likely	0,882	0,704
In any case	0,656	0,630

Table 2. **Significance of differences according to the Mann-Whitney-U Test for sample means of affected and unaffected persons.**

Conclusion

This analysis demonstrates that decision processes in the evaluation of a health state are consistent with predictions of Prospect Theory. Given the confirmation of the first three hypotheses, we are left in an uneasy situation. A recommendation for the use of evaluation scores of unaffected persons does not appear justified. It does matter who is asked. This has been confirmed many times in the past. However, our analysis does not support the implication that unaffected persons cannot be involved in health care decision making. The difference is that unaffected persons are indeed able to judge properly but not all are able to anticipate the shift in reference point, obviously caused by the disease. Persons must anticipate the "right" reference point to evaluate a certain health state correctly. If this finding can be confirmed for other illnesses as well, it is possible generally to correct for this bias. A decision maker who wants to combine evaluations of the general public and the experience of affected persons can use scores of the first group and weight them with the risk attitude of the second. Mathematically speaking:

$$\sum_{i=1}^{5} SG_{iN} \times R_{iB} = \varnothing SG_{iN}$$

as **Eq.1** in the case of the Standard Gamble method

$$\sum_{i=1}^{5} TTO_{iN} \times R_{iB} = \varnothing TTO_{iN}$$

and as **Eq.2** in case of the Time Tradeoff method

where
i = number of risk class
SG = Standard Gamble evaluation values
TTO = Time Tradeoff evaluation values
N = unaffected
B = affected
R = percentage of affected people in risk class i (see figure 3).

How these formulas can be used is demonstrated in the following example: Table 3 shows mean evaluations for single risk classes:

		Health status	
Method	**Risk group**	affected	unaffected
Standard Gamble	In no case	95.27	96.06
	Unlikely	89.80	91.72
	Maybe	86.82	80.55
	Likely	75.33	74.12
	In any case	63.13	67.45
	Mean	87.93	80.67
Time Tradeoff	In no case	86.30	85.08
	Unlikely	86.13	84.40
	Maybe	81.35	77.43
	Likely	75.57	71.59
	In any case	70.05	73.50
	Mean	82.68	78.02

Table 3. **Mean evaluations of single risk groups.**

Applying Eqs. 1 and 2 leads to the following average scores:

Standard Gamble:

$$\begin{aligned} & 42.2\% * 96.06 \\ &+ 14.6\% * 91.72 \\ &+ 26.7\% * 80.55 \\ &+ 8.7\% * 74.12 \\ &+ 7.8\% * 67.45 = \underline{\mathbf{87.14}} \end{aligned}$$

Time Tradeoff:

$$\begin{aligned} & 42.2\% * 85.08 \\ &+ 14.6\% * 84.40 \\ &+ 26.7\% * 77.43 \\ &+ 8.7\% * 71.59 \\ &+ 7.8\% * 73.50 = \underline{\mathbf{80.86,}} \end{aligned}$$

which are much closer to the values of affected persons than the unweighted (mean) values in Table 3. Resulting values lie within the 95% confidence interval for SG scores (between 85.87 and 90.48) and the 95 confidence interval for TTO scores (between 80.01 and 85.87).

Discussion

This analysis is based on utility theory with its compelling advantage of simplicity. Preferences are simply aggregated by adding individual scores. Arrow argued in 1951 that inconsistencies can occur, and the measurement of cardinal preferences might not even be possible [1]. Furthermore, before this procedure can be applied in a broader framework, the relationship between individual reference points, risk attitudes, and evaluation of health state utilities must first be shown for other diseases. If this relation-ship can be confirmed, present league tables in the health domain are questionable at best. These league tables are meant to rank medical interventions according to costs and effectiveness. However, ordinal rankings are severely biased when effectiveness measures refer to "wrong" reference points and hence over- or underestimate the true underlying impact of a medical intervention.

In addition, our analysis depends on several "ifs," and although the results seem structurally valid since all theoretically derived hypotheses were confirmed, it is open to debate whether these results indeed prove what they seem to. Two questions appear to be related which deal with almost identical subjects, namely risky operations. It seems to be a straightforward assumption that persons with lower scores for certain health states risk more to improve their condition. However, correlations were confirmed on items in which such connections were not expected, for example, in the case of the TTO measure. In addition, it is astonishing that the relationship between evaluations of health states and risk attitudes towards health-related quality of life have been neglected in the past. We therefore hope to contribute to establishing better instruments to cover more aspects of quality of life.

The last question is what causes shifts in reference levels. Kahneman und Tversky [14] proposed such a theory in case there is a shift. They later suggested that rapid adaptation might induce such a shift. For future analysis this may also be a crucial point for better understanding evaluations in a cost-effectiveness framework.

Acknowledgements. We are indebted to the Tinnitus League, to Professor Seefeldt from Heinrich-Heine Hospital, and to Hartmut Berndt for their support of our tinnitus questionnaire. We also thank Roland Habich for helpful comments.

References

1. Arrow KJ (1951) Social Choice and individual values. Wiley, New York
2. Balaban DJ, Sagi PC, Goldfarb Nl, Nettler S (1986) Weights for scoring the quality of wellbeing instrument among rheumatoid arthritics. Med Care 24:973-980
3. Bamberg G, Coenenberg AG (1996) Betriebswirtschaftliche Entscheidungslehre. Vahlen, Munich
4. Barron FH, von Winterfeldt D, Fischer GW (1984) Empirical and theoretical relationships between value and utility functions. Acta Psychol 56:233-244
5. Bell DE, Raiffa HJ, Tversky A (1988) Decision making-descriptive, normative, and prescriptive interactions. Cambridge University Press, Cambridge
6. Boyd NF, Sutherland HJ, Heasman KZ, Tritchler DL, Cummings BJ (1990) Whose utilities for decision analysis? Med Decis Making 10:58-67

7. Dolan P (1999) Whose preferences count. Med Decis Making 19:482-486
8. Feldmann H (1998) Tinnitus: Grundlagen einer rationalen Diagnostik und Therapie. Thieme, Stuttgart
9. Gold MR, Siegel JE, Russell LB, Weinstein MC (1996) Cost-effectiveness in health and medicine. Oxford University Press, New York
10. Graham JT (1965) Tinnitus aurium. Almquist & Wiksells, Uppsala
11. Hadorn DC (1991) The role of public values in setting health care priorities. Soc Sci Med 32:773-781
12. Kahnemann D, Tversky A (1979) Prospect Theory: an analysis of decision under risk. Econometrica 47:263-291
13. Kahnemann D, Tversky A (1983) Choices, values, and frames. Am Psychologist 39:341
14. Kaplan RM (1995) Utility assessment for estimating quality-adjusted-life-years. In: Sloan FA (ed) Valuing health care. Cambridge University Press, Cambridge
15. Keeney RL, Raiffa H (1976) Decisions with multiple objectives: preferences and value tradeoffs. Wiley, New York
16. Lenert LA, Treadwell JR, Schwartz CE (1999) Associations between health status and utilities implications for policy. Med Care 37:479-489
17. LIewellyn-Thomas HA, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF (1984) Describing health states: methodologic issues in obtaining values for health states. Med Care 22:543-552
18. LIewellyn-Thomas HA; Sutherland HJ, Thiel EC (1993) Do patients' evaluations of a future health state change when they actually enter that state? Med Care 31:1002-1012
19. Nord E (1992) Methods for quality adjustment of life years. Soc Sci Med 34:559-569
20. Rosser R, Kind P (1978) A scale of valuations of states of illness: is there a social consensus? Int J Epidemiol 7:347-358
21. Sackett DL, Torrance GW (1978) The utility of different health states as perceived by the general public. J Chronic Dis 31:697-704
22. Sloan FA, Viscusi WK, Chesson HW, Conover CJ, Whetten-Goldstein K (1998) Alternative approaches to valuing intangible health losses: the evidence for multiple sclerosis. J Health Econ 17:475-497
23. Suppes P, Winet M (1955) An axiomatization of utility based on the notion of utility differences. Manage Sci 1:259-270

24. Torrance GW (1986) Measurement of health state utilities for economic appraisal. J Health Econ 5:1-30
25. Treadwell JR, Lenert LA (1999) Health values and Prospect Theory. Med Decis Making 19:344-352
26. Verhoef LCG, de Haan AFJ, van Daal WAJ (1994) Risk attitude in gambles with years of life: empirical support for Prospect Theory. Med Decis Making 14:194-200