

# Severity as an independent determinant of the social value of a health service

Jeff R. J. Richardson · John McKie ·  
Stuart J. Peacock · Angelo Iezzi

Received: 29 July 2009 / Accepted: 15 April 2010 / Published online: 9 May 2010  
© Springer-Verlag 2010

**Abstract** This paper has two objectives, first to review the relevant literature concerning the social importance of severity of pre-treatment condition, and second to present the results of a new analysis of the relationship between social value, individual assessment of health improvement and the severity of illness. The present study differs methodologically from others reported in the literature. The underlying hypothesis is that members of the public have an aversion to patients being in a severe health state irrespective of the reason for their being there, and that this aversion will affect the social valuation of a health program after taking account of the magnitude of the health improvement. This effect will be observable in a program which (compared to another) takes a person out of a severe health state—the usual case discussed in the literature—or in a program which (compared to another) leaves a person in a severe health state. The present study tests this second implication of the hypothesis. We present data consistent with the view that after taking account of health improvement, health programs are preferred which do not leave

people in severe health states. Alternative explanations are considered and particularly the possibility that data reflect a social preference for individuals achieving their health potential. Both explanations imply the need to reconsider the rules for prioritising programs. In this analysis, Person Trade-Off (PTO) scores are used to measure social preferences ('value' or 'social utility') and Time Trade-Off (TTO) scores are used to measure individual assessments of health improvement and initial severity. Econometric results suggest that severity is highly significant and may more than double the index of social value of a health service.

**Keywords** Severity · Cost utility analysis (CUA) · Social preferences · Assessment of quality of life (AQoL) · Person trade-off (PTO)

**JEL Classification** I18

## Introduction

Within the framework of conventional Cost-Utility Analysis (CUA) the initial health state of a person is only of importance to the extent that health improvement depends on health-related quality of life (HRQoL) before and after treatment. The initial health state of a person per se is irrelevant. However, when informed of the fact that individual patients find two health improvements to be of identical benefit, members of the public generally express a strong preference for allocating resources to those patients with the worst initial health state. This result has been independently derived in Norway, Australia, the USA, Spain and the UK (see Table 1). It is true that moderately ill patients can only benefit moderately from treatment, whereas severely ill patients can benefit more. But when

---

J. R. J. Richardson · J. McKie (✉) · A. Iezzi  
Centre for Health Economics, Monash University,  
Clayton, VIC 3800, Australia  
e-mail: john.mckie@buseco.monash.edu.au

J. R. J. Richardson  
e-mail: jeff.richardson@buseco.monash.edu.au

S. J. Peacock  
Centre for Health Economics in Cancer,  
British Columbia Cancer Agency,  
Vancouver, BC, Canada

S. J. Peacock  
School of Population and Public Health,  
University of British Columbia, Vancouver,  
BC, Canada

**Table 1** Studies of social preferences for severity of illness

Study	Country	Number of Subjects	Type of Subjects	Elicitation Method	Quantified	Unquantified
Nord 1991 [16]	Norway	1,141	Doctors/Bio /GenPop	PTO	✓	
Nord 1993 [10]	Norway	150	Politicians	DCM		✓
Nord, Richardson et al. 1993 [18]	Norway/ Australia	486	GenPop/Students/Nurses	PTO	✓	
Nord 1993 [17]	Norway	10	NIPH	PTO	✓	
Ubel, Loewenstein et al. 1996 [21]	USA	42	Economics Students	PTO	✓	
Prades 1997 [20]	Spain	30	Economics Students	PTO	✓	
Richardson 1997 [22]	Australia	78	Economics Students	PTO	✓	
Ubel, Spranca et al. 1998 [19]	USA	289	Prospective Jurors	DCM	✓	
Dolan and Green 1998 [11]	England	28	Sec/Admin/Academic	PTO	✓	
Ubel 1999[12]	USA	479	Prospective Jurors	DCM		✓
Dolan and Tsuchiya 2005 [15]	England	100	GenPop	RE		✓
Present Study	Australia	430	GenPop	PTO	✓	

The shaded rows indicate studies raising evidence challenging the social preference for the more severely ill

<sup>a</sup> Bio = Bioengineers

<sup>b</sup> GenPop = General Population

<sup>c</sup> NIPH = Staff at the National Institute of Public Health in Oslo

<sup>d</sup> Sec/Admin/Academic = Secretarial, administrative and academic university staff

<sup>e</sup> PTO = Person Trade-Off

<sup>f</sup> DCM = Discrete Choice Method

<sup>g</sup> RE = Ranking Exercise

<sup>h</sup> Quantified = Cardinal (numerical) preferences

<sup>i</sup> Unquantified = Ordinal preferences only

patients are expected to derive the same benefit, and all else is equal, conventional CUA provides no basis for distinguishing between them. There is no value associated with the severity of the initial health state itself. Conventional CUA disregards entirely the following sort of sentiment, as expressed by Callahan: ‘Our bias, I contend, should be to give priority to persons whose suffering and inability to function in ordinary life is most pronounced, even if the available treatment for them is comparatively less efficacious than for other conditions’ [1, p. 463]; a point also noted by Cohen: ‘society may want to direct resources preferentially to those who are farthest from good health, even if larger aggregate benefits could be obtained under a different distribution’ [2, p. 287].

The idea that the worse off—e.g. the more severely ill—have a moral claim for special consideration has considerable intuitive appeal. It can be found in official government guidelines in several countries, and in reports of government-appointed commissions [3–5]. It is encapsulated in Rawls’s ‘Difference Principle’, which states that social and economic inequalities are justifiable only in so far as they are to the greatest benefit of the least advantaged

members of society [6]. Daniels has explicitly extended Rawls’s Difference Principle to health care, arguing that fair equality of opportunity requires the provision of health care to those in greatest need—in the present context, the more severely ill [7].

The importance of severity is not a purely theoretical matter. In the US severity has been the dominating factor in the allocation of heart and liver transplants (when need exceeds supply). Those with the best prognosis after receipt of an organ are those with the least severe illness, and maximum health gain would be achieved by giving this group priority. By contrast, the actual policy gives a very high weighting to those with the most severe problem. This results in the counterintuitive situation where the relatively healthy must wait until their health has deteriorated sufficiently for them to satisfy the severity criterion [8, 9]. However, this policy is only counterintuitive against a back-drop that assumes that maximising health gains is the overriding social objective. In the present case, health production is explicitly of secondary importance to severity (Kidney transplantation is not included in this policy since dialysis is available as an alternative—that is, the condition

without transplantation is not sufficiently severe for inclusion in the policy).

### Discordant evidence concerning social preferences

To test the importance of severity, Nord surveyed 150 Norwegian politicians involved in health-care decision making and found that 38% would give priority to the treatment of an illness that gives ‘severe’ problems in preference to an illness that gives ‘moderate’ problems, even though treatment would help those with the severe illness only ‘a little’ whereas it would help those with the moderate illness ‘considerably’ [10]. Another 45% would divide any increase in funding evenly between the two, leaving only 11% who would follow the health maximisation strategy of conventional CUA. Only general descriptors were used in this study and no attempt was made to measure strength of preference, but the results give some indication of the direction of the preferences, albeit with a large egalitarian component.

However, a study by Dolan and Green produced results apparently at odds with this [11]. Using EuroQoL health states, they asked subjects to select a health state, D, such that they would consider a move from C to D to be of equivalent value to a move from A to B, where A is a more severe health state than C. They then asked respondents to do a PTO involving treatment T1, which would take patients from A to B, and treatment T2, which would take them from C to D. In contrast to a concern for the more severely ill, only 7 out of 28 respondents preferred T1 whereas 17 preferred T2.

It is possible respondents misunderstood aspects of this task. A hint of this comes from the qualitative results, which are curiously non-explanatory. Of the 17 respondents who preferred the T2 treatment, 12 commented: ‘*I would say that treatment 2 was a definite improvement... in treatment 1 the difference isn’t so great*’. If subjects were saying they considered T2 of greater value from a personal perspective, then this would seem to be inconsistent with their choice of D. If they meant that T2 was of greater value from a social perspective it is not clear why. Lack of reflection is a possibility explanation for these results. That notwithstanding, 25% of respondents still chose T1 which would benefit more severely ill patients.

In a subsequent study using prospective jurors as subjects, Ubel replicated in the US the earlier study by Nord [10]. Like Nord, he found that many people are prepared to sacrifice overall health gains to benefit those with the worst initial health state. Of 479 subjects, 9% gave priority to patients with ‘moderate’ health problems, 26% gave priority to those with ‘severe’ health problems, and 64% chose to divide resources equally between the two groups.

However, Ubel noted that when subjects are not given the option of dividing resources, only a small majority favoured giving priority to patients with severe health problems. This raises the possibility that subjects in Nord’s original study may not have been expressing a preference for severity per se, but ‘may have simply been unwilling to make a difficult treatment choice’ [12, p. 897]. Also, Ubel found that responses were sensitive to the wording of the options. For example, when subjects were reminded about how much improvement each group was expected to undergo (‘a little’ for the severely ill, and ‘considerably’ for the moderately ill), fewer participants gave priority to the severely ill (6% compared with 26% without the reminder), and more gave priority to the moderately ill (21% compared with 9% without the reminder). This does not negate the significance of severity. However, it does indicate that considerable caution must be exercised in calculating precise severity weights.

A recent study by Dolan and Tsuchiya produced an even more startling result [13]. They asked subjects to rate a number of patient groups that differed, inter alia, in terms of the quality and length of their lives without treatment. They found that respondents gave consistently higher priority to patients with better prospects without treatment—be it in terms of life expectancy or quality of life—than patients with poorer prospects. If these results are to be believed, being more severely ill gives patients *lower* priority for health care from a social point of view.

Nord offered one possible explanation for this anomalous finding: respondents may have thought that the group with the better prospects had better prospects *as a result of treatment* [14]. Nord calls this the ‘fatal misunderstanding’ hypothesis. Dolan and Tsuchiya did not explicitly discuss this possibility in their original article, but pointed out, and reiterated in a rejoinder, that ‘respondents may not have processed the information as intended’ [15]. However, they do not dismiss the results entirely. They point out that the support for the more severely ill detected in other studies could likewise have resulted from respondents mistakenly thinking that the benefit to the more severely ill will be greater.

In light of these inconclusive findings, it is clear that a final assessment of the importance of severity can only be reached by weighing up the evidence for and against. The number and quality of studies in support of severity must be balanced against the number and quality of those against. Importantly, the conditions under which severity should be taken into account, and the weight it should be given, must also be derived from the evidence. The present paper aims to make a contribution to this debate by deriving inferred severity weights for QALYs based on data collected for the Assessment of Quality of Life (AQoL-II) project. This is undertaken in Sects. 4–6. First,

however, we briefly review the remaining empirical evidence in Sect. 3, which supports the importance of severity. One objective in the following section is to observe the different contexts in which the severity hypothesis has been supported.

### Supporting evidence

In one of the earliest studies on social concerns for severity—certainly one of the earliest that attempted to quantify preferences for severity—Nord found that returning one person to full health from the following state—‘unable to work, unable to pursue family and leisure activities, strong pain, depressed’—was considered as valuable as returning 50 people to full health from the following state—‘unable to work, moderate pain’. However, the utility values for these states (assigned by the participants using a rating scale) implied that curing one person in the more severe state should be equivalent to curing two people in the less severe state [16, 17]. This indicates that the social value assigned to treating the more severely ill was much higher than would be expected from the patient utility scores.

In a joint Norwegian-Australian study, also using the PTO, Nord, Richardson et al. surveyed members of the general public (in Norway) and students and nurses (in Australia). Subjects were asked to adopt the perspective of members of Parliament and to choose between two equally expensive special health units. Unit A would save ten people per year from dying and restore them to full health. Unit B would restore to full health a larger number of patients suffering from a chronic illness. The PTO results were again higher than would be expected from the individual utility scores for these states. The utility scores seriously underestimate the social value placed upon the health states when the alternative is death. The authors comment, however, that subjects’ responses to the PTO exercise were highly dispersed, ‘indicating the likelihood of a high sampling error for the median values’ [18, p. 467].

In another study, Nord asked a convenience sample of ten individuals from the National Institute of Public Health in Oslo to compare improvements in health on a seven-level disability scale with approximately equal distances between the levels [17]. For example, subjects were asked to indicate how many patients moving from level 5 to level 1 on the scale they considered equivalent to a smaller number of patients moving from level 6 to level 4 (where lower numbers represent better functioning). Again, the results showed a marked preference for treating the more severely ill (see Table 2). Although taking a patient from level 5 to level 1 should be twice as valuable as taking a patient from level 6 to level 4, taking into account only

**Table 2** Numbers of different outcomes that may be considered equivalent in social value

Severity without intervention	Severity with intervention						
	7	6	5	4	3	2	1
1							
2							128
3						83	64
4					54	42	32
5				35	27	21	16
6			22	17	13	10	8
7		15	12	9	7	5	4
Dead	?	3	2	2	1	1	1

Source: [11]

Seven step severity scale: 1 No problems with walking, 2 Can move about without difficulty anywhere, but has difficulties with walking more than a kilometre, 3 Can move about without difficulty at home, but has difficulties in stairs and outdoors, 4 Moves about without difficulty at home, 5 Can sit. Needs assistance to move about—both at home and outdoors, 6 To some degree bedridden. Can sit in a chair part of the day if helped up by others, 7 Completely bedridden

gains in HRQoL, Nord’s subjects judged them to be approximately equal: taking 16 patients from level 5 to level 1 and taking 17 patients from level 6 to level 4 were both found to be equivalent to taking one person from dying to a state of healthy (the latter was used as a reference state). One of the important aspects of this study was that the levels on the scale were judged approximately equal by the subjects themselves. This makes it difficult to explain the discrepancy between the utility-based predictions and the direct measurements by arguing that the health states were not placed on an interval scale.

Ubel, Spranca and colleagues, using the same seven-step scale as Nord, found that the observed preference for more severely ill patients extends to preventative interventions [19]. Using prospective jurors as subjects, Ubel, Spranca, et al. found only a slight preference for preventative over curative interventions when they brought similar benefits at similar costs. However, there was a significant preference for helping the more severely ill patients in both contexts. Moreover, this preference was observed, in both the curative and preventative context, even when the more severely ill would benefit less.

In another study, Pinto-Prades asked subjects in Spain to assume the role of health planners in an exercise designed to compare the Visual Analogue Scale (VAS), the Standard Gamble (SG) and (three forms of) the Person Trade-Off (PTO) [20]. The study used four EuroQoL health states for comparison. The values assigned to health improvements from these health states are shown in Table 3. Again, it can be seen that the PTO places much higher value on life saving, and ameliorating severe conditions, than the VAS

**Table 3** Sizes of intervals as measured by VAS, SG and PTO

	Mean				
	VAS	SG	PTO-1	PTO-2	PTO-3
Death-32,331	16	44	59	79	59
32,331-23,232	13	27	25	11	22
23,232-21,312	23	10	11	7	8
21,312-12,121	23	14	3.5	2	6
12,121-perfect health	25	5	1.5	0.8	0.8

Source: [13]

Results for each instrument are measured on a 0–100 scale. Thus, moving from death to EuroQoL health state 32,331 (a severe health condition) increased instrument scores by 16 (VAS), 44 (SG) and 59, 79, 59 (PTO 1, 2, 3, respectively). In contrast, improving health from EuroQoL state 12,121 (relatively good health) to the EuroQoL best health state increased scores by 25, 5, 1.5, 0.8, 0.8, respectively

or SG, and less weight on curing milder conditions. For example, to gain the equivalent of returning 4 people from health state 12,121 (a mild health state) to perfect health according to the VAS ( $0.75 \times 4 = 3$ ) it would be necessary to return only 3 people to full health according to the TTO ( $0.99 \times 3 = 2.9$ ), whereas to gain the equivalent of saving 5 people's lives who would be left in health state 32,331 (a severe health state) according to the VAS ( $0.84 \times 5 = 4.2$ ) it would be necessary to save 10 people's lives according to the TTO ( $0.41 \times 10 = 4.1$ ).

In a study in the US, Ubel, Loewenstein and colleagues conducted a study with economics students also using the VAS, the SG and TTO. They measured the utility associated with three health states—ganglion cyst of the hand, ligament damage to the knee, and severe headache. They then tested the same students one to 3 weeks later using the PTO to elicit their rationing choices for groups of patients with the same three conditions [21]. They also added a fourth, fatal condition—appendicitis. Participants were asked the following question concerning appendicitis and meningioma (and similar questions concerning the other conditions):

A. Which do you think would bring the most benefit?

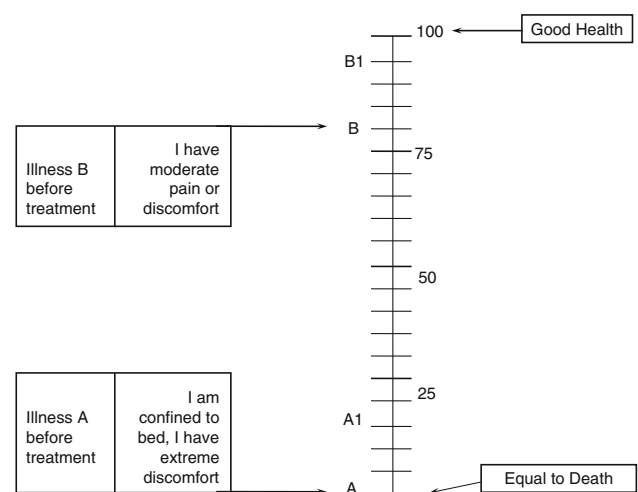
- ten people cured of appendicitis
- \*\* people cured of meningioma
- indifferent

The questions were tailored to individual participants—that is, the double asterisks were replaced by the number at which each participant was predicted to be indifferent, based on their answers to the utility elicitation questions. Subjects did not agree with the rationing implications of their answers to the utility elicitation questions. Higher values were assigned to health states that were initially more severe (despite this being taken into account in the evaluation of the utility scores) and the preference to treat

more severely ill people was 'consistent across all six rationing choices and all three methods of elicitation' [21, p. 113].

Finally, in an Australian study involving 78 economics students, Richardson [22] ensured that the value of health improvement to patients at different levels of severity was perceived as being the same by informing subjects that the patients would be prepared to pay \$30,000 for either the treatment of illness A or illness B; alternatively, that the patients considered health improvement from these treatments to be sufficiently valuable that they would sacrifice 1 year of their life to receive them in both cases. Subjects were then asked to adopt a social perspective by imagining that they were on a health committee of Parliament and had to prioritise the two treatments (see Fig. 1). Higher priority was given to illness A and illness B by 57% and 16% of respondents respectively, and equal priority (indicating that social value equals individual value) by 28%. When asked to nominate the number of people that would need treatment for illness B to generate the same social value as 100 people receiving treatment A the mean and median values were 318 and 200 respectively; that is, the value of the treatment for the more severe illness was valued between 2 and 3 times more highly than the value of the less serious illness despite patients valuing the treatments equally.

Taking into consideration the information on public preferences revealed in the previous studies, Nord [23 p. 37–38] divides health states into three classes—'severe', 'considerable' and 'moderate'—and assigns them values consistent with the emerging empirical data outlined above. See Table 4. These values give rise to some 'rules of thumb' concerning severity: saving someone from death is something like 3–6 times better—has greater social value—than curing someone of a severe health problem (and returning them to full health),



**Fig. 1** Quality of Life Scale

**Table 4** Rules of thumb concerning severity (after Nord 1999)

Health state	Description	Value range
Death		0.0
Severe	For instance, a person who has to sit in a wheelchair, has pain most of the time, and is unable to work	0.65–0.85
Considerable	For instance, a person who must use crutches to walk, has light pain intermittently, and is unable to work	0.90–0.94
Moderate	For instance, a person who has difficulty moving about outdoors and slight discomfort, but is able to do some work and has only minor difficulties at home	0.98–0.995
Full health		1.0

something like 10–15 times better than curing someone of a moderate problem, and 50–200 times better than curing someone of a moderate problem. According to Nord: ‘Quantitative models that purport to be useful for estimating the social value of health care activities in these countries (Australia, England, Norway, Spain and the US), as well as in other countries with similar values, must reflect this structure of concern’ [23 p. 38]. By contrast with the health-state (QALY) values used in CUA—derived, for example, by means of the SG, TTO or RS—these social values are higher, particularly at the upper end of the scale. The value structure encapsulated in Table 4 therefore ‘compresses health states to the upper end of the scale’ [23 p. 38].

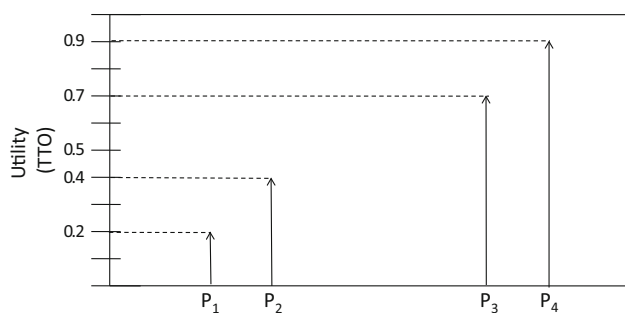
It would appear that, from a social perspective, conventional CUA underestimates the relative value of curing severe health problems, including life-saving treatments. Treatments for the more severely ill are favoured over those for the less severely ill, both when respondents themselves consider the two treatments of equivalent value from a personal point of view, and when they are told that (hypothetical) patients consider them equal. It is supported even when giving priority to the more severely ill means reducing overall health gains (severity has more than just tie-breaking significance). This result has been observed in the case of preventative and curative interventions, and when non-fatal conditions of different severity are compared as well as when life-threatening and non-fatal conditions are compared.

Nevertheless, a small number of studies have produced contrary results. At this stage, the significance of these noncompliant studies is unclear: they may indicate the importance of other factors (e.g. expected benefit, final health state, age) that can override severity, misunderstanding on the part of respondents, or national differences. Clearly, more evidence on the social importance of severity is desirable. In the remainder of this paper we present evidence relevant to this question, based on data derived from the construction of the Assessment of Quality of Life (AQoL-II) project. These are used to test a general relationship between severity and utility gain as distinct from a single test of the main hypothesis.

### A new study: methodology and data

In the present study, PTO values are used to indicate the social value of a movement from death to a health state less than full health. TTO scores are used to measure the individual utility of each of these final health states. The disutility of a health state (1—utility) is the measure of severity of the state in which the person will be left by the treatment. Differences in PTO scores are regressed upon the difference in the utilities (health improvement) and also the severity of the more severe health states in which patients would remain without further treatment. The analysis is based upon three key assumptions. The first two, that the PTO and TTO measure the social value of a health program and the individual utility of a health state respectively are widely, but not universally, accepted. The relevant arguments are outside the scope of the present article. The third is that the importance of severity declines as health improves: that the severity affect is subject to diminishing returns with respect to health. This implies that differences in health improvement as judged by the TTO will be more highly valued by society when health states are poor. The relevant health states are at the severe end of the health state spectrum.

The hypothesis is illustrated with a simple numerical example in Fig. 2. In this, four programs  $P_1 \dots P_4$  will each save a person’s life and leave them in a health state with utility scores 0.2, 0.4, 0.7 and 0.9 respectively.  $P_2$  and  $P_4$  result in TTO scores that are 0.2 higher than  $P_1$  and  $P_3$  respectively. However  $P_1$  leaves a patient in a more severe health state (TTO = 0.2) than  $P_3$  (TTO = 0.7) and creates greater social disutility than the severity of other health states. The study hypothesis is therefore that the difference in the PTO scores between  $P_2$  and  $P_1$  will be greater (say 0.3) than the difference between  $P_4$  and  $P_3$  (say 0.15). This is because  $P_2$  does not leave a patient in a health state as severe as  $P_1$ . While  $P_4$  does not leave patients in  $P_3$  this is of less concern as  $P_3$  is a less severe health state than  $P_1$  and the importance of severity is subject to diminishing returns. The differences ( $P_2-P_1$ ) and ( $P_4-P_3$ ) are therefore hypothesised to be positively related to the magnitude of severity after standardising for differences in TTO scores.



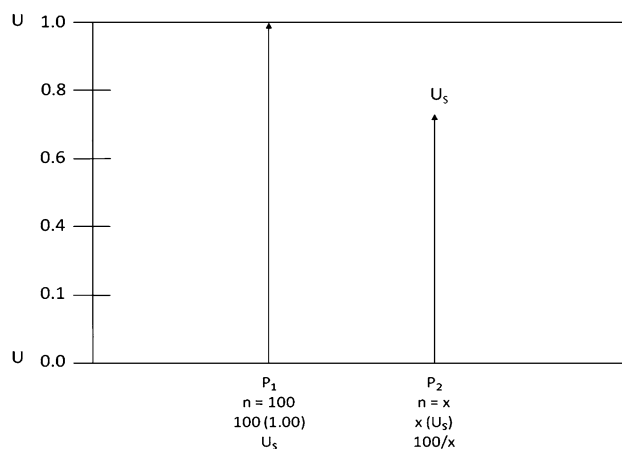
**Fig. 2** Utility (TTO) program value (PTO) and severity

An alternative (over) simplified explanation of the hypothesis is to assume  $PTO_i = TTO_i + S_i$  where  $S$  is the net social utility of a person being taken from death but left in the health state measured by  $TTO_i$ . Consequently  $PTO_2 - PTO_1 = 0.2 + (S_2 - S_1)$  and  $PTO_4 - PTO_3 = 0.2 + (S_4 - S_3)$ . Our hypothesis is that  $(S_2 - S_1) > (S_4 - S_3)$  as a result of diminishing returns to severity as health improves and our data are used to test this.

While the equation above ( $PTO_i = TTO_i + S_i$ ) is numerically correct it implies an over-simplification of the relationship between  $TTO$  and  $PTO$  which measure different quantities—a state versus change—from different perspectives—individual versus social—and with different framing effects—time discounting and distributional preferences. Consequently, the difference between  $PTO$  and  $TTO$  is likely to reflect more than just severity. Our test is that differences in  $PTO$  are a function, *inter alia*, of changes in  $TTO$  and the severity of the worst state in which a person may be left.

The study used health state descriptions and data from the Assessment of Quality of Life (AQoL-II) project [24–28], the descriptive system. The health state descriptions formed by this instrument have a very high level of psychometric integrity and represent a valid and reliable description of critical health states. In the second stage of the AQoL project the instrument was scaled and, as part of the validation process,  $TTO$  and  $PTO$  scores were obtained for 18 multi-attribute health states from respondents selected to represent the Australian SES and demographic profile.  $TTO$  data were collected in the conventional way. Using a slide board as a visual aid subjects were asked to select between 10 years in the relevant health state and a reduced number of years in full health. The latter were ‘flip flopped’ until the subject believed the value of the reduced years was equivalent to the value of 10 years in the health state.  $TTO$  scores were obtained by dividing this number of years by 10.

$PTO$  scores were obtained as shown in Fig. 3. Using a visual aid, subjects were asked to select between two programs,  $P_1$  and  $P_2$ . Program 1 would save the life of 100 patients and return them to full health. Program 2 would



**Fig. 3** Health gain, same start points

save the life of  $x$  patients and leave them in the health state of interest,  $U_s$ . The value of  $x$  was varied until the two programs appeared to be of equal value. ‘Social utility’ scores (i.e. value obtained from a social perspective) were obtained from the equation  $U_s = 100/x$ .

These  $PTO$  data all related, as described above, to programs commencing at death (without treatment). To test the study hypothesis different combinations of programs were compared. This had a twofold advantage. First, it increased the number of observations which could be observed from a single individual. Secondly, it allowed the observations of marginal improvements in health from two alternative programs. As data were not collected for the full 18 health states from all subjects it was not possible to construct all combinations of health state movements. In total, the study constructed 36 comparisons of programs which left patients at different levels of severity. An average of 22.75 observations was obtained for each comparison giving a total of 819 individual observations.

As described, the importance of severity was tested by econometrically regressing individual values for health changes upon the health state improvement measured by the  $TTO$  scores ( $TTO_i - TTO_j$ ) and upon the severity (disutility) of the poorer health state in which a patient would remain without further treatment ( $1 - TTO_j$ ). A power function was used as shown in Eq. 1. This was selected as a flexible functional form with the required property that the equation must pass through the points (0,0) and (1.00, 1.00).

$$PTO_{ij} = (U_i - U_j)^\alpha \cdot DU_j^\beta \tag{1}$$

If social and individual assessments were identical then  $\alpha = 1$  and  $\beta = 0$ . With  $\alpha < 1$ ,  $\beta = 0$  the  $PTO$  would give additional weight to smaller health improvements which would be consistent with the greater scores given by  $PTO$  to life saving programs or programs removing people from severe health states. But it is inconsistent with the

observation that PTO gives smaller, not larger, scores to programs giving incremental improvements to patients in less severe health states. The study hypothesis would result in  $\alpha < 1$ ,  $\beta < 1$ . Small health improvement where health states are severe would have amplified importance. Small health improvement nearer full health would have relatively smaller effects as determined by the relative magnitude of  $\alpha$  and  $\beta$ .

**Results**

Demographic characteristics are summarised in Table 5. A response rate of 41.7% was obtained from the 1,030 possible respondents. Compared with the Australian population a disproportionate number of respondents had a tertiary degree. Otherwise the sample characteristics satisfactorily reflected those of the general population.

Results of the econometric analysis are reported in Eqs. 1 and 2 in Table 6. Both OLS and random effects models were employed with the latter taking account of the clustering of observations on individuals. For both functional forms coefficients on health improvement and severity were significant at the 1 percent level.  $\alpha$  and  $\beta$  coefficients were, as hypothesised, less than unity. Taking account of individual clustering (Eq. 2) the random effects model assigned greater importance to severity and a corresponding reduction in the importance of health improvement.

One criticism of the PTO technique used in this study is that some subjects may treat large numbers metaphorically, not literally (e.g 1,000 means ‘many’; 10,000 means ‘very many’) [29]. Implausibly large values of  $n$  in the calculation of utility ( $100/n$ ) will result in implausibly low utility scores. As a consequence, a second analysis was carried out to test the sensitivity of results to this possible problem. Values that were more than 0.4 below the resulting mean

**Table 5** Scaling survey: respondents and response rate

	PTO-TTO interview
Possible respondents	1,030
Respondents	430
Response rate %	41.7
Respondents	
Sex % male	35
Age % age <25	3
Age % age 60	22
Education	
A Primary %	3
B Secondary %	47
C Tertiary %	50

**Table 6** Model:  $PTO = (U_1 - U_2)^\alpha \cdot (DU_2)^\beta$

No	Eq type	Statistic	$\alpha$	$\beta$	$t\alpha$ $t\beta$ ( $Z\alpha$ $Z\beta$ )	$n$	$R^2$ (wold)
1	OLS	Individual	0.77	0.12	2.52, 2.34	819	0.76
2	RE	Individual	0.55	0.33	13.4, 4.96	819	(1,059)
3	OLS	Individual <sup>(2)</sup>	0.71	0.23	10.97, 2.64	237	0.83
4	RE*	Individual <sup>(2)</sup>	0.47	0.46	5.85, 4.06	237	(440)

Key (1) Random effects model; U = TTO

(2) Censored data;  $DU = 1 - TTO$

utility score for a health state were deleted and every other observation for that individual was deleted. The severity of this criterion reduced the number of individual observations to 237.

Results of the econometric analyses of the censored dataset are reported in equations 3 and 4 of Table 6.

Contrary to expectation, the results were not particularly sensitive to the editing of data. With and without censoring, both the improvement in health (the treatment effect),  $U_1 - U_2$ , and the initial severity of the health state,  $DU_2$ , were significant, with the co-efficient on the former falling marginally with data censoring and the co-efficient on the latter increasing. Wold and  $R^2$  summary statistics indicate that in the reduced dataset results also have high explanatory power.

Results from equation 4, which is the theoretically most reliable result, were used to generate value scores for a range of health gains and initial severity levels, as indicated in Table 7. The importance of severity may be seen by reading down the columns. For example, from column 1 the health gain for point 0.2 has a social value of 0.22 if the initial severity ( $DU$ ) is 0.2. The same health gain is worth 0.31 with an initial severity of 0.4 and 0.47 if the initial severity is 1.0 (i.e. the patient would otherwise have died).

The effect of severity is highlighted in Table 8, which takes the ratio of the change in social value to the change in utility. Reading down the same column, a gain of 0.2 due to

**Table 7** Social value by utility gain and severity

$(U_1 - U_0)$ DU severity	0.2	0.4	0.6	0.8	1.0
0.2	0.22				
0.4	0.31	0.42			
0.6	0.37	0.51	0.62		
0.8	0.42	0.59	0.73	0.82	
1.0	0.47	0.65	0.79	0.9	1.0

Social value:  $PTO = (U_1 - U_0)^{0.47} (DU)^{0.46}$



**Table 8** Ratio change in social value to utility gain

(U <sub>1</sub> -U <sub>0</sub> ) DU severity	0.2	0.4	0.6	0.8	1.0
0.2	1.1				
0.4	1.6	1.1			
0.6	1.9	1.3	1.0		
0.8	2.1	1.5	1.2	1.0	
1.0	2.4	1.6	1.3	1.0	1.0

Value (D<sub>ij</sub>)/Utility gain (U<sub>1</sub>-U<sub>0</sub>)

health improvement has rapidly increasingly social value as the severity of the initial health state increases.

## Discussion

The results of the present study indicate that differences in the social valuation of programs as measured by the PTO are related to the value of the poorer health state in the comparison as measured by the TTO. One plausible explanation of this consistent with observations elsewhere is that there is a social disutility associated with patients remaining in severe health states and for reasons discussed in the literature the effect is additional to the effect of health improvement per se judged by the individual. The hypothesis implies that there will be a social benefit if an individual is taken out of a severe health state and a social loss if they are left in a severe health state. The present study tests this second implication. If the severity effect is subject to diminishing returns with respect to health then the severity effect will rise disproportionately as the health state a person is left in deteriorates. Therefore when a pair of programs are compared, the difference between the social and the personal valuation rises as health declines. The present study uses this result to test the hypothesis. The differences in the social valuation of pairs of programs are compared with the level of health. If the severity effect is subject to diminishing returns with respect to health then after standardising for the magnitude of the change programs offering similar health gain will be more highly valued when health is poor.

Despite the different form of the test the fundamental hypothesis is the same as in other studies: *people do not wish others to be in a severe health state*. The results here suggest that it is not the reason for their being in the severe state that is of importance—whether the severe state occurs before or after a health program. Rather it is people ‘not being in the severe state’ which is valued.

Like the results of any statistical analysis our hypothesis is only one of several which might explain the data. Despite the longevity of their use, the properties of the TTO and PTO are not fully documented. Results may be

due to some unknown property of the particular PTO instrument used or to a framing effect or to systematic misunderstanding of the question. However the latter explanation, in particular, is unconvincing. We applied an extremely stringent criterion for accepting data in the second analysis. Experienced interviewers reported a satisfactory level of comprehension by those interviewed.

A further possible problem is described by Nord et al. [30] who argue that changes in health states, as measured by conventional instruments like the TTO may not have the interval property that is claimed for them. In particular, two changes in health may be valued differently from the value implied by the conventional numbers because people with unequal health potential may have varying aspirations and reference points. This argument may well be true as it is very likely that people’s interest in health improvement will vary with their circumstances. As argued in Richardson and McKie [31] and more explicitly in Richardson et al. [32] orthodox theory implicitly assumes that disembodied health benefits may be redistributed to compensate losers (the Kaldor-Hicks criterion). As this is untrue the social or individual benefits of health gains that are embodied in a particular individual become subject to a range of other considerations. However, in the present context, the pertinent point is that individual utility is currently measured by the TTO and SG. The present results suggest that, as measured, health improvement must be adjusted by the severity of health states to obtain correct social valuations. The interpretation of this for policy is, of course, open to discussion.

A third possible interpretation of the results sees a more significant role for respondents’ preferences regarding potential: that is, respondents believe it is important for others to realise their potential for health improvement, even if this is limited. The idea that it is unfair to give a lower priority to those with a limited capacity for improvement—e.g. the disabled and chronically ill—is well known in the philosophical literature [33–35], and empirical studies of public preferences provide further support for it, both in the context of life-saving and life-enhancing treatment [36–41]. What the present study shows, on this interpretation, is that the importance of potential is subject to diminishing marginal returns, as severity decreases, conversely, the value of observing others achieve their potential increases disproportionately as the health state of the person deteriorates. While this possibility has been raised in other studies [37, 41], it has never been clearly demonstrated.

As with severity, it would be a mistake to think that the social benefit of potential is simply added to the private benefit of health improvement (and therefore explains why PTO exceeds TTO scores). This would be inconsistent with the data. As both TTO and PTO scales vary from 0.00

(death) to 1.00 (best health), the greater PTO values in the region of poor health imply a smaller, not greater, benefit from programs that treat patients in the regions of good health. Consistent with this, improvements in PTO scores are typically smaller than improvements in TTO utilities for programs treating patients in moderate to good health states.

A synthesis of these two interpretations is that PTO social values are a function of TTO individual utilities and the effect of potential and severity. One point in favour of this interpretation is that respondents in the PTO exercise were presented with the comparisons in Fig. 3 and the patients who will benefit from program  $P_2$ —the program being evaluated—have less capacity to benefit in comparison with the invariant program,  $P_1$ . It was not explained to respondents, and is irrelevant for present purposes, whether the patients who will benefit from  $P_2$  are already in state  $U_s$  and will be returned to this state after life-saving treatment, or whether  $U_s$  will begin after treatment [on this distinction see, 38]. On this interpretation, the results reported here are a result of the combined effect of potential and severity: people derive value from seeing others achieve their potential, but this varies with the severity of the health state involved. The importance of our results, however, is not diminished by the alternative explanation, as both the ‘severity hypothesis’ and ‘potential hypothesis’ are in conflict with the usual view that the value of a program can be measured by health improvement alone.

The results add further weight to the evidence already reported that members of the public have an aversion to the health inequalities implied by individuals remaining in serious health states. The results are consistent with widely recognised ethical perceptions. Indeed, Rawls has been criticised for not allowing natural inequalities, such as those arising from health status, to be a factor in determining who is worst off and therefore more deserving of compensation. For example, Kymlicka comments: ‘According to Rawls, people born into a disadvantaged class or race not only should not be denied social benefits, but also have a claim to compensation because of that disadvantage. Why treat people born with natural handicaps any differently? Why should they not also have a claim to compensation for their disadvantage?’ [42, pp. 72–73]. Green makes a similar point, arguing that health care is in fact a *social* good: ‘Access to health care is not only a social primary good, but possibly one of the most important such goods ... [because] disease and ill health interfere with our happiness and undermine our self-confidence and self-respect’ [43, p. 117]. As we have seen, a growing body of evidence indicates widespread support for the ethical principle that, other things being equal, extra weight should be given to treatments that benefit those who are more severely ill.

Neither the results of the present study, nor those surveyed earlier, mandate that severity must be incorporated into health policy. Even leaving aside the few studies that find little support for severity, to assume that because something *is* the case (because a number of studies have detected widespread support for a policy or principle) that it *ought* to be the case (that the policy or principle should be implemented) is to commit the ‘naturalistic fallacy’ [44]. It is to assume that normative conclusions can be derived directly from factual premises. On the contrary, the results of public consultation exercises must always remain open to ethical criticism, and no normative conclusions can be deduced from purely empirical enquiries.

This being said, the results of empirical inquiries are relevant, and a strong case can be made for applying a severity-based equity weighting to the value of health state improvements for the purposes of setting priorities in the health sector [45]. The present results suggest that results of well-conducted studies will neither give overriding importance to severity nor neglect it entirely. It will represent a trade-off between health maximisation (efficiency) and giving priority to the more severely ill (equity). In liberal democracies, this can be seen as a socially acceptable compromise given the range of moral views in the community, some of which will tend to support more efficient policies others more egalitarian ones. This does not gainsay the impossibility of deriving ethical conclusions from empirical data. Rather, it represents a reasonable compromise given the persistence of ethical disagreement.

## Conclusion

There is a growing body of empirical evidence concerning severity of illness, most of which indicates that it is a significant factor for many people when allocating limited health care. When asked to judge for others—when asked to adopt a social perspective—many respondents systematically re-weight individual patient preferences to take account of the severity of a patient’s health state. As a result, higher priority will be accorded to programs which allow patients to avoid being in severe conditions. This implies higher priority for some programs than would be obtained from an individual’s assessment of health improvement alone. This reflects a social judgement about the distribution of health benefits—about what is fair. Our results therefore suggest that the notion of ‘effectiveness’ in cost-effectiveness analyses should be extended to encompass not just magnitude of health improvement but a weighting of expected benefits to reflect the severity of the condition. In practice this implies additional priority to programs treating patients in severe health states.

Similarly, the values attached to health states by multi-attribute utility instruments (MAUs) need to be modified to account for severity of pre-treatment condition [46]. Our results are not general which implies that 'further empirical research is needed to determine the appropriate transformation functions with reasonable precision' [47]. However our analysis and the results are robust and plausible, and increase the case for independently including the importance of severity in economic evaluation studies.

## References

- Callahan, D.: Setting mental health priorities: problems and possibilities. *Milbank Q.* **72**, 451–470 (1994)
- Cohen, B.J.: Utility measurement and the allocation of health care resources. *Med. Decis. Making* **15**, 287–288 (1995)
- Campbell, A., Gillett, G.: Justice and the right to health care. In: *Ethical Issues in Defining Core Services*. The National Advisory Committee on Core Health and Disability Support Services, Wellington (1993)
- Dutch Committee on Choices in Health Care. *Choices in Health Care*. Ministry of Welfare, Health and Cultural Affairs, Rijswijk (1992)
- Swedish Health Care and Medical Priorities Commission: *No Easy Choices: The Difficulties of Health Care*. Sveriges offentliga utredningar, 1993. The Ministry of Health and Social Affairs, Stockholm (1993)
- Rawls, J.: *A Theory of Justice*. Harvard University Press, Cambridge (1971)
- Daniels, N.: *Just Health Care*. Cambridge University Press, Cambridge (1985)
- Ubel, P.A., Arnold, R.M., et al.: Rationing failure: the ethical lessons of the retransplantation of scarce vital organs. *JAMA* **270**, 2469–2474 (1993)
- Wikler, D.: Equity, efficiency, and the point system for transplant recipient selection. *Transplant. Proc.* **21**, 3437–3439 (1989)
- Nord, E.: Helsepolitikere ønsker ikke mest mulig helse per krone (Health politicians do not wish to maximize health benefits). *J. Norwegian Med. Assoc.* **113**, 1171–1173 (1993)
- Dolan, P., Green, C.: Using the person trade-off approach to examine differences between individual and social values. *Health Econ.* **7**, 307–312 (1998)
- Ubel, P.A.: How stable are people's preferences for giving priority to severely ill patients? *Soc. Sci. Med.* **49**, 895–903 (1999)
- Dolan, P., Tsuchiya, A.: Health priorities and public preferences: the relative importance of past health experience and future health prospects. *J. Health Econ.* **24**, 703–714 (2005)
- Nord, E.: Severity of illness and priority setting: worrisome lack of discussion of surprising finding: discussion. *J. Health Econ.* **25**, 170–172 (2006)
- Dolan, P., Tsuchiya, A.: Severity of illness and priority setting: worrisome criticism of inconvenient finding? A reply to Erik Nord. *J. Health Econ.* **25**, 173–174 (2006)
- Nord, E.: The validity of a visual analogue scale in determining social utility weights for health states. *Int. J. Health Plann. Manage.* **6**, 234–242 (1991)
- Nord, E.: The trade-off between severity of illness and treatment effect in cost-value analysis of health care. *Health Policy* **24**, 227–238 (1993)
- Nord, E., Richardson, J., et al.: Social evaluation of health care versus personal evaluation of health states: evidence on the validity of four health state scaling instruments using Norwegian and Australian Surveys. *Int. J. Technol. Assess. Health Care* **9**, 463–478 (1993)
- Ubel, P.A., Spranca, M.D., et al.: Public preferences for prevention versus cure: what if an ounce of prevention is worth only an ounce of cure. *Med. Decis. Making* **18**, 141–148 (1998)
- Prades, J.-L.P.: Is the person trade-off a valid method for allocating health care resources? *Health Econ.* **6**, 71–81 (1997)
- Ubel, P.A., Loewenstein, G., et al.: Individual utilities are inconsistent with rationing choices: a partial explanation of why Oregon's cost-effectiveness list failed. *Med. Decis. Making* **16**, 108–116 (1996)
- Richardson, J.: Critique and some recent contributions to the theory of cost utility analysis. In: *Working Paper 77*. Centre for Health Program Evaluation, Monash University, Melbourne (1997)
- Nord, E.: *Cost-Value Analysis in Health Care*. Cambridge University Press, New York (1999)
- Richardson, J., Day, N.A., et al.: The assessment of quality of life (AQoL) II instrument: overview of the assessment of quality of life mark 2 project. In: *Working Paper 144*. Centre for Health Program Evaluation, Monash University, Melbourne (2004a)
- Richardson, J., Day, N.A., et al.: The assessment of quality of life (AQoL) II instrument: derivation of the scaling weights using a multiplicative model and econometric second stage correction. In: *Working Paper 142*. Centre for Health Program Evaluation, Monash University, Melbourne (2004b)
- Richardson, J., Day, N.A., et al.: Conceptualising the assessment of quality of life instrument mark 2 (AQoL 2): methodological innovations and the development of the AQoL descriptive system. In: *Working Paper 141*. Centre for Health Program Evaluation, Monash University, Melbourne (2004c)
- Peacock, S., Richardson, J., et al.: The assessment of quality of life (AQoL) II instrument. The effect of deliberation and alternative utility weights in a multi-attribute utility instrument. In: *Working Paper 143*. Centre for Health Program Evaluation, Monash University, Melbourne (2004)
- Richardson, J., Day, N., et al.: Measurement of the Quality of Life for Economic Evaluation and the Assessment of Quality of Life (AQoL) Mark 2 Instrument. *Aus. Econ. Rev.* **37**, 62–88 (2004)
- Schwarzinger, M., Lanoe, J.-L., et al.: Lack of multiplicative transitivity in person trade-off responses. *Health Econ.* **13**, 171–181 (2004)
- Nord, E., Undrum Enge, A., et al.: QALYs: is the value of treatment proportional to the size of the health gain? *Health Econ.* doi:10.1002/hec.1497, 10.1002/hec.1497 (2009)
- Richardson, J., McKie, J.: Economic evaluation of services for a national health scheme: the case for a fairness-based framework. *J. Health Econ.* **26**, 785–799 (2007)
- Richardson, J., McKie, J., et al.: A critique of efficiency focussed economic evaluation in the context of a NHS: the case for empirical ethics, Research Paper 34. Centre for Health Economics, Monash University, Melbourne (2009)
- Harris, J.: Qualifying the value of life. *J. Med. Ethics* **13**, 117–123 (1987)
- Daniels, N.: Rationing fairly: programmatic considerations. *Bioethics* **7**, 224–233 (1993)
- Scanlon, T.M.: Preference and urgency. *J. Philos.* **LXXII**, 655–669 (1975)
- Nord, E., Richardson, J., et al.: Maximizing health benefits vs egalitarianism: an Australian survey of health issues. *Soc. Sci. Med.* **41**, 1429–1437 (1995)
- Abellan-Perpiñán, J.-M., Prades, J.-L.P.: Health state after treatment: a reason for discrimination? *Health Econ.* **8**, 701–707 (1999)

38. Ubel, P.A., Richardson, J., et al.: Exploring the role of order effects in person trade-off elicitation. *Health Policy* **61**, 189–199 (2002)
39. Dolan, P., Cookson, R.: A qualitative study of the extent to which health gain matters when choosing between groups of patients. *Health Policy* **51**, 19–30 (2000)
40. Nord, E.: QALYS: is the value of treatment proportional to the size of the health gain? *Health Econ.* (2009) doi: [10.1002/hec.1497](https://doi.org/10.1002/hec.1497). Accessed 3 July 2009
41. Nord, E.: The relevance of health state after treatment in prioritising between different patients. *J. Med. Ethics* **19**, 37–43 (1993)
42. Kymlicka, W.: *Contemporary Moral Philosophy*. Oxford University Press, Oxford (1990)
43. Green, R.M.: Health care and justice in contract theory perspective. In: Branson, R. (ed.) *Ethics and Health Policy*, pp. 111–126. Ballinger Publishing Co, Cambridge (1976)
44. Pigden, C.: Logic and the autonomy of ethics. *Australas. J. Philos.* **67**, 127–151 (1989)
45. Nord, E., Pinto-Prades, J.L., et al.: Incorporating societal concerns for fairness in numerical valuations of health programmes. *Health Econ.* **8**, 25–39 (1999)
46. Nord, E.: Health state values from multiattribute instruments need correction. *Ann. Med.* **33**, 371–374 (2001)
47. Nord, E.: Severity of illness versus expected benefit in societal evaluation of health care interventions. *Expert Rev. Pharmacoecon. Outcomes Res.* **1**, 85–92 (2001)