



Severity as an independent determinant of the Social Value of a Health Service

Professor Jeff Richardson

Foundation Director, Centre for Health Economics,
Monash University

Dr John McKie

Research Fellow, Centre for Health Economics,
Monash University

Dr Stuart Peacock

Senior Scientist, Centre for Health Economics in Cancer, British Columbia Cancer Agency
Honorary Senior Research Fellow, Centre for Health Economics, Monash University

Mr Angelo Iezzi

Research Fellow, Centre for Health Economics,
Monash University

July, 2007

Correspondence:

Prof. Jeff Richardson
Centre for Health Economics
Faculty of Business and Economics
Building 75
Monash University Victoria 3800
Australia

Phone: +61 (0)3 9905 0754, Fax: +61 (0)3 9905 8344
Jeff.richardson@buseco.monash.edu.au

TABLE OF CONTENTS

| | |
|---|----|
| Introduction | 1 |
| Background..... | 2 |
| Methodology of the 2004 Australian Study..... | 11 |
| Results | 13 |
| Discussion..... | 15 |
| Conclusion | 15 |
| References..... | 16 |
| Appendix..... | 18 |

List of Tables

| | |
|--|----|
| Table 1. Studies of Social Preferences for Severity of Illness | 1 |
| Table 2. Comparison of Norwegian and Australian Results: Distribution in Percentages, Median Values, and Confidence Intervals (CI) | 3 |
| Table 3. Numbers of Different Outcomes That May Be Considered Equivalent in Social Value..... | 4 |
| Table 4. Values of the Health States Using Different Preference Elicitation Methods..... | 6 |
| Table 5. Mean Utilities for the Health States as Measured by the Three Utility-elicitation Methods | 7 |
| Table 6. Rationing Choices for Health States Between Which Subjects Should Be Indifferent..... | 7 |
| Table 7. Discrepancies Between Indifference Points of Rationing Choices and Those Predicted by Utility-survey Responses | 8 |
| Table 8. Subjects' Justifications for their Allocation Preferences | 9 |
| Table 9. Rules of Thumb Concerning Severity (after Nord 1999)..... | 10 |
| Table 10. Scaling Surveys: Respondents and Response Rates | 13 |
| Table 11. Model: $PTO = (U1 - U2) \alpha \cdot (DU2) \beta$ | 13 |
| Table 12: Summary Statistics | 18 |
| Table 13: Descriptive Statistics | 18 |

ABSTRACT

The measure of benefit in cost utility analysis (CUA) is the increase in utility which is attributable to a health service. This paper reviews the evidence that the severity of an illness – the health state before receipt of the health service – may be independently important for social (as distinct from individual) preferences for different services. An earlier 1997 Australian study is summarised.

Data from a 2004 survey are used to quantify the apparent importance of severity. Person trade off (PTO) scores are used to measure social preferences and time trade off (TTO) scores to measure individual preferences. Econometric results suggest the severity may more than double the index of social value of a health service.

Severity as an independent determinant of the Social Value of a Health Service:

Introduction

In conventional Cost-Utility Analysis (CUA) – based, for example, on the cost-per- QALY prioritising rule - the initial health state of a person is only of importance to the extent that *health improvement* depends upon 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, people generally express a strong preference for allocating resources to those with the worst initial health state. This result has been independently derived in Norway, Australia, the USA, Canada and Spain (see Table 1). It is true that moderately ill patients can only benefit moderately from treatment, whereas severely ill patients can benefit more substantially. However, when patients are expected to derive the same benefit, and all else is equal, conventional CUA provides no basis for distinguishing between them. In fact, the cost-per-QALY prioritising rule will “discriminate” against more severely ill patients when prioritising less severe illnesses produces more QALYs.¹ There is no value associated with the severity of the initial condition itself. The cost-per-QALY prioritising rule disregards entirely the following sort of sentiment: “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” (Callahan 1994, p. 463).

Table 1. Studies of Social Preferences for Severity of Illness

| Study | Country | Number of Subjects | Type of Subjects | Elicitation Method |
|--------------------------------|------------------|--------------------|---|---------------------|
| (Nord 1991) | Norway | 1,141 | Doctors/Bio ^a /GenPop ^b | PTO ^d |
| (Nord 1993b) | Norway | 150 | Politicians | DCM ^e |
| (Nord, Richardson et al. 1993) | Norway/Australia | 486 | GenPop/Students/Nurses | PTO |
| (Nord 1993a) | Norway | 10 | NIPH ^c | PTO |
| (Abelson, Lomas et al. 1995) | Canada | 280 | GenPop/Health Officials | DelPol ^f |
| (Baron, Wu et al. 2000) | USA | 42 | Economics Students | PTO |
| (Prades 1997) | Spain | 30 | Economics Students | PTO |
| (Baron, Wu et al. 2000) | USA | 289 | Prospective Jurors | DCM |
| (Baron, Wu et al. 2000) | USA | 479 | Prospective Jurors | DCM |

a Bio = Bioengineers

b GenPop = General Population

c NIPH = Staff at the National Institute of Public Health in Oslo

d PTO = Person Trade-Off

e DCM = Discrete Choice Method

f DelPol = Deliberative Polling

¹. It would be wrong to construe such discrimination as intentional. It is an unrecognised implication of the exclusive focus on health-maximisation. Savulescu calls such unintentional discrimination “*de facto* discrimination” (Savulescu, J. (1998). “Consequentialism, Reasons, Value and Justice.” *Bioethics* 12(3): 212-235. There is no intention to disadvantage those who are more severely ill, or any presumption that they are less worthy of concern. In this respect it differs from racism, sexism, ageism, to which it is sometimes likened, underlying which there is an intention to disadvantage some groups *based on* race, sex, age, and so on.

The idea that the worse off – e.g. the more severely ill - have a moral claim for special consideration has strong intuitive appeal. It can be found in official government guidelines in several countries, and in reports of government-appointed commissions (Dutch Committee on Choices in Health Care 1992; Campbell and Gillett 1993; Swedish Health Care and Medical Priorities Commission 1993). 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 (Rawls 1971, pp. 302-3). Even philosophers within the utilitarian tradition – from which CUA derives its credentials - have attempted to account for the intuition that the worst off deserve special consideration. For example, Wolf argues from a utilitarian perspective that actions which reduce (or minimise) misery are *prima facie* obligatory, whereas actions that increase well-being, although good, are not obligatory (Nord and Wolfson 1999). This gives lexical priority to helping those who are worst off – e.g. the more severely ill.

The importance of severity is not a purely theoretical matter. In the USA 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 "perverse" situation where the relatively healthy must wait until their health has deteriorated sufficiently for them to satisfy the severity criterion (Wikler 1989; Baron, Wu et al. 2000). However, this policy can only be described as "perverse" if maximising health gains is the overriding social objective. In the present case, health production is explicitly of secondary importance to severity.²

Background

In an early study 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 (Nord 1991; Nord and Wolfson 1999, p. 33). Similarly, returning one person to full health from the first state was considered as valuable as returning 100 people to full health from a state of "moderate pain", even though the utility associated with "moderate pain" implied that curing one person in the more severe state should be equivalent to curing five people in "moderate pain". This indicates that the societal value of treating the more severely ill is much higher than would be expected from taking account only of the utility scores of patients.

In another study of 150 Norwegian politicians involved in health-care decision making, Nord found that 38 per cent of subjects 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" (Nord 1993b). Another 45 per cent would divide any increase in

². 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.

funding evenly between the two, leaving only 11 per cent who would follow the health maximisation strategy of conventional, QALY-based CUA.

Table 2. Comparison of Norwegian and Australian Results: Distribution in Percentages, Median Values, and Confidence Intervals (CI)

| Number cured in state <i>i</i> equivalent to 10 lives saved | State | | | |
|---|----------------|-----------|----------------|-----------|
| | W ^a | | Z ^b | |
| | Norway | Australia | Norway | Australia |
| 0-9 | 4.1 | 0.0 | 5.9 | 6.2 |
| 10 | 8.8 | 9.3 | 23.5 | 21.5 |
| 11-19 | 4.1 | 2.3 | 5.9 | 7.7 |
| 20-39 | 8.3 | 6.9 | 11.8 | 10.8 |
| 40-50 | 4.1 | 20.9 | 11.8 | 20.0 |
| 51-100 | 20.8 | 16.3 | 23.5 | 16.9 |
| 101-999 | 8.3 | 25.7 | 5.9 | 6.1 |
| 1000 up | 41.7 | 18.6 | 11.8 | 10.8 |
| Total | 100.0 | 100.0 | 100.0 | 100.0 |
| Number | 24 | 43 | 17 | 65 |
| Median | 110 | 85 | 50 | 40 |
| 90% CI | 100 | 50 | 25 | 30 |
| | 200 | 200 | 100 | 50 |

Source: (Nord, Richardson et al. 1993)

a State W = Uses crutches for walking. Light pain intermittently. Unable to work.

b State Z = Sits in a wheelchair. Pain most of the time. Unable to work.

In a joint Norwegian-Australian study, Nord, Richardson et al. surveyed members of the general public (in Norway) and students and nurses (in Australia) using the PTO. 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 (W or Z). Participants were asked to indicate how many patients treated in unit B per year they consider equivalent to saving the ten patients in unit A. Table 2 shows that the median response in Norway was 110 for state W and 50 for state Z, and in Australia was 85 for state W and 40 for state Z. These numbers are higher than would be expected from the individual utility scores for these states. As Nord, Richardson et al. note, the utility scores seriously underestimate the *social value* placed upon the health states when the alternative is death (Nord, Richardson et al. 1993, p. 46). The authors also comment, however, that subjects' responses to the PTO exercise were highly dispersed, "indicating the likelihood of a high sampling error for the median values" (Nord, Richardson et al. 1993, 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 (Nord 1993a). 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 3). 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 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). According to Nord, these findings “support the hypothesis that the conventional QALY model, which attributes all social appreciation to treatment effect, is misconceived” (Nord 1993a, p. 233). Moreover, because the levels on the disability scale were judged approximately equal by the subjects themselves, it is hard 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.

Table 3. Numbers of Different Outcomes That May Be Considered Equivalent in Social Value

| | | Severity with intervention | | | | | | |
|-------------------------------|------|----------------------------|----|----|----|----|----|-----|
| | | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| Severity without intervention | 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: (Nord 1993a)

Seven Step Severity Scale:

1. No problems with walking.
2. Can move about without difficulty anywhere, but has difficulties with walking more than a kilometer.
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.

General support for the importance of severity was also detected in a Canadian study by Abelson, Lomas, and colleagues. Participants in this study were drawn from five different groups: randomly selected citizens, attendees at town-hall meetings, appointees to district health councils, elected officials and experts in health care and social services. All five groups thought that information about *needs* was most important for health-care and social-service decision making (Abelson, Lomas et al. 1995). With the exception of elected officials, all groups thought that information about *benefits* was next in importance, followed by *costs*, and finally *preferences*. Elected officials ranked costs second. It is possible to identify need with potential to benefit - those who are in greatest need are those who can benefit most from treatment – and in this way make need as an allocation criterion compatible with the cost-per-QALY prioritising rule. However benefits were separately listed in this study and it is difficult not to equate "need" with those who are most severely ill (Edgar, Salek et al. 1998, pp. 68-71). The subjects in this study appear to have adopted the latter interpretation, and ranked information on benefits as *less important* than information on need (severity). In this way they confirmed Lockwood's view that “the QALY arithmetic is inherently insensitive to differences in degree of need, except in so far as they happen to correlate with the degree of benefit per unit cost that treatment can confer” (Lockwood 1988, p. 45-46).

Ubel, Spranca and colleagues, using the same seven-step scale as Nord in the previous study (1993a), found that the observed preference for more severely ill patients extends to preventative

interventions (Baron, Wu et al. 2000). Using prospective jurors as subjects, Ubel 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 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 SG and (three forms of) the PTO for the purposes of priority setting (Prades 1997). The study used the following four EuroQoL health states:

12121

No problems with walking about.
Some problems washing or dressing self.
No problems with performing usual activities (e.g. work, study, housework, family or leisure activities).
Moderate pain or discomfort.
Not anxious or depressed.

21312

Some problems with walking about.
No problems with self care.
Unable to perform usual activities.
No pain or discomfort.
Moderately anxious or depressed.

23232

Some problems with walking about.
Unable to wash or dress self.
Some problems with performing usual activities.
Extreme pain or discomfort.
Moderately anxious or depressed.

32331

Confined to bed.
Some problems washing or dressing self.
Unable to perform usual activities.
Extreme pain or discomfort.
Not anxious or depressed.

The values assigned to these health states by the VAS, the SG and the PTO are shown in Table 4. Again, the PTO, which measures social value and incorporates information on severity per se, consistently assigned higher values to these states. Thus, for example, for state 21312 the VAS score was 52, the SG score was 81 and the PTO score was 95. This implies that saving one life would be equivalent to curing two people of condition 21312 according to the VAS ($0.48 \times 2 = 0.96$), five people according to the SG ($0.19 \times 5 = 0.95$), and twenty people according to the PTO ($0.5 \times 20 = 1.0$). Clearly, using the PTO will place a much higher value on life saving than the VAS or the SG.

Table 4. Values of the Health States Using Different Preference Elicitation Methods

| | Mean (SE) | | | | |
|--------------|--------------|--------------|----------------|-----------------|-----------------|
| | VAS | SG | PTO-1 | PTO-2 | PTO-3 |
| 12121 | 75 (2.37) | 95 (1.28) | 98.5 (0.82) | 99.20 (0.29) | 99.14 (0.58) |
| 21312 | 52 (2.19) | 81 (2.92) | 95 (1.31) | 97.05 (0.69) | 93.73 (2.43) |
| 23232 | 29 (2.01) | 71 (3.83) | 84 (3.83) | 90.78 (2.27) | 81.26 (4.5) |
| 32331 | 16 (1.52) | 44 (5.11) | 59 (6.44) | 79.65 (5.32) | 59.25 (6.44) |

Source: (Prades 1997)

In a study in the US, Ubel, Loewenstein and colleagues conducted a study with economics students using the VAS, the SG and TTO. Using these instruments they sought to measure the utility associated with three health states – ganglion cyst of the hand, ligament damage to the knee, and severe headache. The results are given in Table 5. They then tested the same students one to three weeks later using the PTO to elicit their rationing choices for groups of patients with the same three conditions (Baron, Wu et al. 2000). 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 asterices were replaced by the number at which each participant was predicted to be indifferent, based on their answers to the utility elicitation questions. The study showed that people do not agree with the rationing implications of their answers to utility elicitation questions. For example, when asked to compare appendicitis with meningioma, 37 subjects chose to treat the appendicitis patients, 2 chose to treat those with meningioma, and 3 were indifferent (see Table 6.) “This preference to treat more severely ill people was consistent across all six rationing choices and all three methods of elicitation” (Baron, Wu et al. 2000, p. 113). Confirming this, subjects were asked:

B. How many people would have to be cured of meningioma to equal the benefit brought by curing ten people of appendicitis?

Table 5. Mean Utilities for the Health States as Measured by the Three Utility-elicitation Methods

| Health Condition | Analogue Scale | Utilities | |
|------------------|----------------|-----------------|---------------|
| | | Standard Gamble | Time Tradeoff |
| Cyst | 0.92 | 0.91 | 0.99 |
| Knee | 0.63 | 0.83 | 0.94 |
| Meningioma | 0.37 | 0.75 | 0.90 |

Source: (Baron, Wu et al. 2000)

Table 6. Rationing Choices for Health States Between Which Subjects Should Be Indifferent

| Treatment Choice | Analogue Scale | Standard Gamble | Time Tradeoff | Total | Significance of Difference |
|-----------------------------------|----------------|-----------------|---------------|-------|----------------------------|
| Appendicitis vs meningioma | | | | | p < 0.0001 |
| Appendicitis | 15 | 10 | 12 | 37 | |
| Indifferent | 0 | 1 | 2 | 3 | |
| Meningioma | 2 | 0 | 0 | 2 | |
| Appendicitis vs knee | | | | | p < 0.0001 |
| Appendicitis | 17 | 11 | 13 | 41 | |
| Indifferent | 0 | 0 | 0 | 0 | |
| Knee | 0 | 0 | 1 | 1 | |
| Appendicitis vs cyst | | | | | p < 0.0001 |
| Appendicitis | 17 | 11 | 14 | 42 | |
| Indifferent | 0 | 0 | 0 | 0 | |
| Cyst | 0 | 0 | 0 | 0 | |
| Meningioma vs cyst | | | | | p < 0.0001 |
| Meningioma | 17 | 9 | 14 | 40 | |
| Indifferent | 0 | 2 | 0 | 2 | |
| Cyst | 0 | 0 | 0 | 0 | |
| Meningioma vs knee | | | | | p < 0.0001 |
| Worse | 15 | 9 | 13 | 37 | |
| Indifferent | 1 | 2 | 1 | 4 | |
| Better | 1 | 0 | 0 | 1 | |
| Knee vs cyst | | | | | p < 0.0001 |
| Knee | 15 | 10 | 13 | 38 | |
| Indifferent | 1 | 1 | 1 | 3 | |
| Cyst | 1 | 0 | 0 | 1 | |

Source: (Baron, Wu et al. 2000)

The results are given in Table 7, along with five other pair-wise comparisons. The value 8.3 in the second column means that the point at which subjects were indifferent between the two outcomes (cured of appendicitis versus cured of meningioma), calculated by their answers to question B, was 8.3 times greater than the point at which they were predicted to be indifferent based on their answers to the utility elicitation question (using the VAS). "The results show how far the respondents' answers diverged from those predicted by their answers to the utility survey" (Baron, Wu et al. 2000, p. 113).

Table 7. Discrepancies Between Indifference Points of Rationing Choices and Those Predicted by Utility-survey Responses

| Rationing Scenario | Median Discrepancy* | | |
|----------------------------|---------------------|-----------------|---------------|
| | Analogue Scale | Standard Gamble | Time Tradeoff |
| Appendicitis vs meningioma | 8.3 | 100.0 | 81.5 |
| Appendicitis vs knee | 35.7 | 10,000.0 | 720.9 |
| Appendicitis vs cyst | 100.0 | 6.0 EX8 | † |
| Meningioma vs cyst | 10.0 | 1270 | 100.0 |
| Knee vs cyst | 10.0 | 40.0 | 5.0 |
| Meningioma vs knee | 44.3 | 3.3 | 10.0 |

Source: (Baron, Wu et al. 2000)

* A value of 1.0 means that the subject's indifference point in the rationing choice was the same as that predicted by the utility survey response (no discrepancy) and 100 means that the subject's indifference point in the rationing survey was one hundred times greater than that predicted by the utility survey response.

† Could not be calculated because the median lay between a numerical answer and one denoting infinity.

In a subsequent study using prospective jurors as subjects, Ubel repeated the earlier study by Nord (1993b). Like Nord, he found that many people are prepared to sacrifice overall health gains in order to benefit those with the worst initial health state. Of 479 subjects, 9 per cent gave priority to patients with "moderate" health problems, 26 per cent gave priority to those with "severe" health problems, and 64 per cent chose to divide resources equally between the two groups. However, Ubel noted that when subjects are not given the option of dividing resources evenly between patients with "severe" health problems and those with "moderate" health problems, only a small majority favoured giving priority to the former group. This result raises the possibility that subjects in Nord's original study, and perhaps in other studies, were not expressing a preference for severity per se, but "may have simply been unwilling to make a difficult treatment choice" (Baron, Wu et al. 2000, 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 per cent compared with 26 per cent without the reminder), and more gave priority to the moderately ill (21 per cent compared with 9 per cent without the reminder). This does not negate the significance of severity, as Ubel acknowledges (Baron, Wu et al. 2000, p. 902), but it does indicate that caution must be exercised in calculating severity weights for the purposes of priority setting. Table 8 shows subjects' reasons for their allocation choices.

Table 8. Subjects' Justifications for their Allocation Preferences

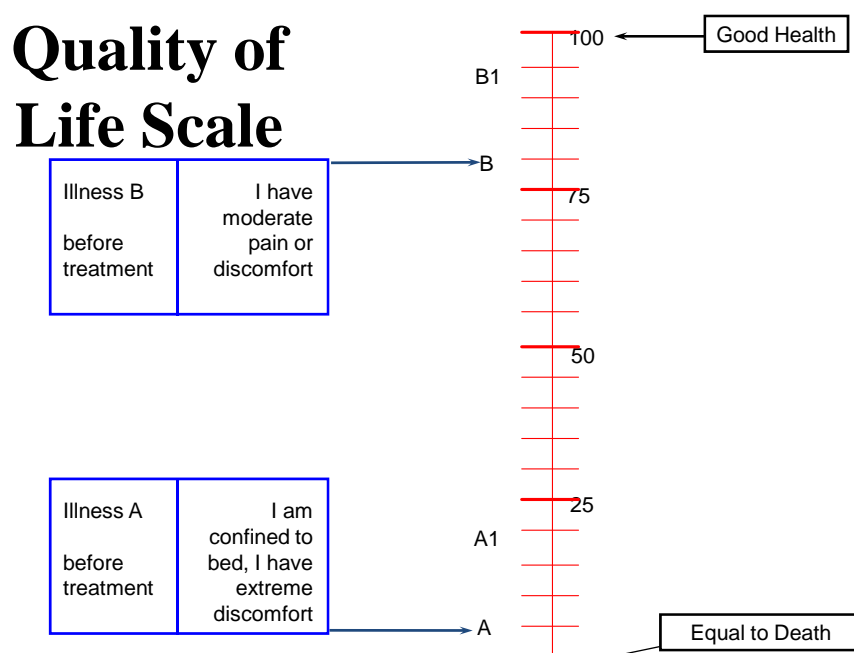
| Justification | Number of subjects providing justification according to allocation preference ^a | | | Total |
|---|--|-------------------------------------|-------------------------|-------|
| | Priority to severely ill patients | Priority to moderately ill patients | Divide resources evenly | |
| Fairness | 2 | 1 | 101 | 104 |
| Maximize health benefits | 0 | 83 | 3 | 86 |
| Severely ill deserve priority | 52 | 0 | 2 | 54 |
| Future research benefits | 11 | 1 | 10 | 22 |
| Prevent decline in health | 2 | 11 | 8 | 21 |
| Future economic benefits | 4 | 11 | 2 | 17 |
| Severe illness more urgent | 8 | 0 | 1 | 9 |
| Do not discriminate according to treatment benefits | 0 | 0 | 7 | 7 |
| Level the playing field | 5 | 0 | 1 | 6 |

Source: (Baron, Wu et al. 2000)

a Some subjects provided more than one justification.

In an unpublished study of 78 Melbourne students in 1997 Richardson sought to ensure that the value of health improvement to individuals at different levels of severity was perceived as being the same by informing subjects that they would be prepared to pay \$30,000 for either the treatment of illness A or illness B; alternatively they would consider health improvement from these treatments to be sufficiently valuable that they would sacrifice one year of their life to achieve them in both cases. This information was provided in writing next to the scale shown in Figure 1 which indicated the severity of the health state before the treatment was provided. Subjects were asked to adopt a social perspective by imagining that they were on a health committee of Parliament and had to prioritise the two treatments.

Figure 1



Higher priority was given to illness A and illness B by 57 and 16% of respondents respectively and equal priority (social = individual value) by 28%. When asked to nominate the number of people who 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 as highly as the value of the less serious illness despite individuals having the same personal value for the treatments.

Taking into consideration the information on public preferences for severity revealed in the preceding studies, Nord divides health states into three classes – “severe”, “considerable” and “moderate” - and assigns them values consistent with the above empirical findings (Nord and Wolfson 1999, pp. 37-38). See Table 9. 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, 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 societal 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” (Nord and Wolfson 1999, p. 38). Notice that, 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. From the societal perspective, conventional CUA underestimates the value of curing severe health problems, including life-saving treatments. The value structure encapsulated in Table 9 therefore “compresses health states to the upper end of the scale” (Nord and Wolfson 1999, p. 38).

Table 9. 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 |

Measuring the severity of different health states is not without its problems. First, there is disagreement about the best way to measure HRQoL. Different techniques, such as the SG, the TTO, and VAS, produce different results, and have their own advantages and disadvantages (Froberg and Kane 1989; Nord 1992; Richardson 1994; Dolan, Gudex et al. 1996; Prades 1997). As Nord observes, “this is naturally a serious problem for potential users of economic evaluations” (Nord and Wolfson 1999, p. 91). Second, there is disagreement about whose preferences should be used to value health states in the first place. There is evidence that different groups – patients, health professionals, “well-informed” members of the general public – rate health states differently (Boyd, Sutherland et al. 1990; Rothman 1991; Hurst, Jobanputra et al. 1994; Glick, Polsky et al. 1999; Menzel, Dolan et al. 2002). The studies cited earlier show something quite different however: that there is a consistent societal preferences for giving

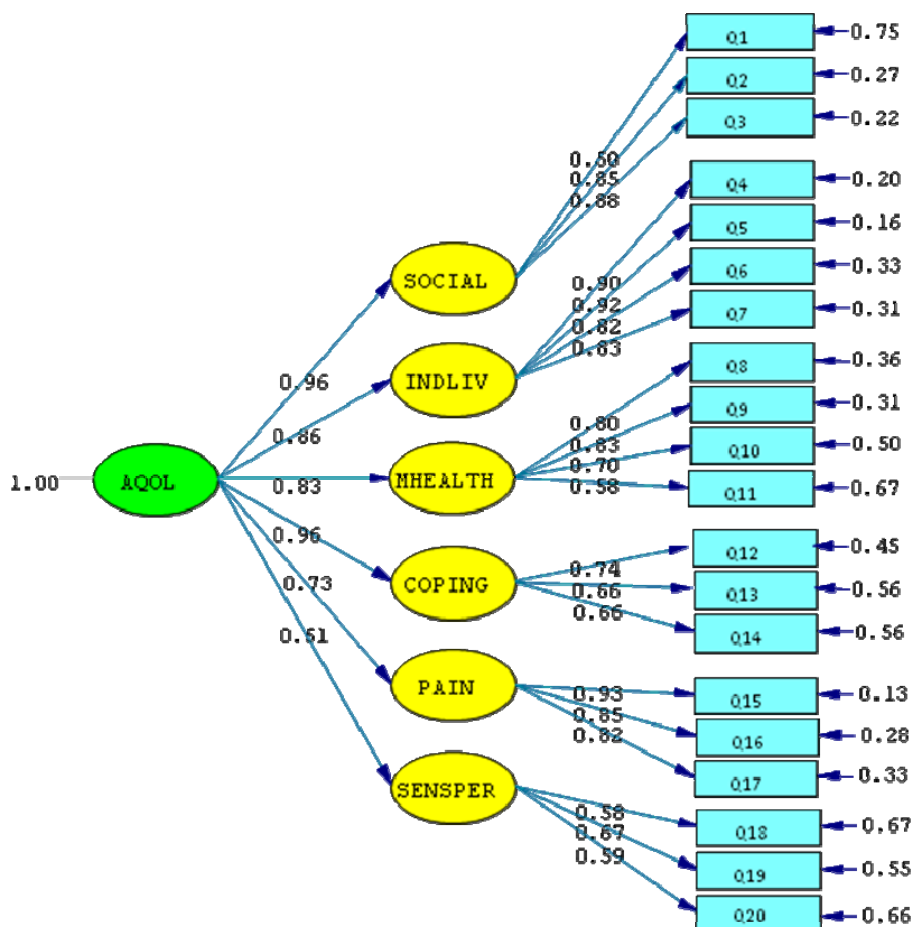
priority to the more severely ill, regardless of how severity is measured and regardless of whose preferences are used to value health states.

Methodology of the 2004 Australian Study

The study reported below was based upon data collected for the assessment of Quality of Life (AQoL) project which is described in more detail in Richardson et al 2003. Data were collected from the general population in Melbourne stratified by the socio-economic status of the location. Interviews were conducted either at the subject's home or at a convenient central location. Data collected included time trade off (TTO) scores for the scaling of the AQoL, and of greater relevance here, both time trade off and person trade off (PTO) data for 18 health states constructed from the AQoL descriptive system (see Figure 2).

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.

Figure 2.

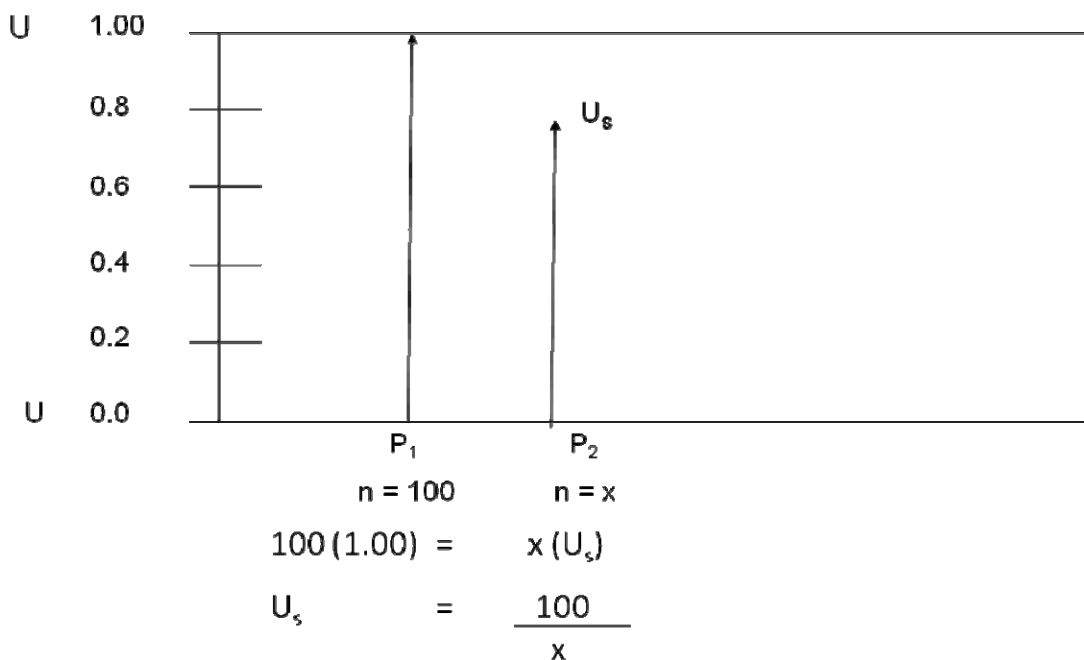


Chi Square = 460.73, df = 164, P – value = 0.00000, RMSEA = 0.054, CFI

PTO scores were obtained as shown in Figure 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 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 obtain the value of programs which commenced at some other health state calculations were carried out as shown in Figure 4. In this P_{01} and P_{02} represent 2 programs which take patients from death to U_1 and U_2 respectively. Subtracting the PTO score for the former from the latter produces the score for P_{12} a program which takes patients from U_1 to U_2 . Similarly by subtracting P_{01} and U_{02} from a third program P_{03} PTO scores may be obtained for P_{13} and P_{23} which are programs taking patients from U_1 to U_3 and U_2 to U_3 respectively. 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 "pseudo PTO" scores involving a movement from a health lower greater than death.

Figure 3.



The importance of severity was tested by econometrically regressing the values of these 36 health states obtained from different subjects against the health state improvement defined by the TTO scores ($TTO_i - TTO_j$) and the severity (disutility) of the initial health state ($1 - TTO_j$). A power function was used as shown in Equation 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)

Basic Severity Equation

$$\text{Value} = (TTO_1 - TTO_2)^\alpha (DU(AQoL_0))^\beta \dots (1)$$

Results

Survey characteristics are summarised in Table 10. A response rate of 41.7% was obtained from the 1030 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.

Table 10. Scaling Surveys: Respondents and Response Rates

| | | PTO-TTO Interview |
|----------------------|--|-------------------|
| Possible respondents | | 1030 |
| 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% |

PTO values included a large number which were implausibly low. After consultation with the interviewers it became apparent that some subjects had difficulty with the treatment of numbers greater than 100 in a way which maintained an interval property (i.e., the difference between 200 and 210 means the same as the difference between 3003 and 310). Implausibly large values of n in the calculation of utility ($100/n$) resulted in implausibly low utility scores. As a consequence analysis was carried out with two sets of data. In the first there was no censoring of data and implausible results were included. In the second very severe data editing was carried out. Initially, obviously low observations were deleted. Secondly, values which were more than 0.4 below the resulting mean were deleted. As entire individuals were thereby removed from the data bank as their understanding of the procedure was in question the resulting observations were drastically reduced from 819 to 237.

Summary statistics are given in an appendix.

Results of the econometric analysis are reported in Table 11. Both OLS and random effects models were employed with the latter taking account of the clustering of observations on individuals. Equations 1 and 2 were for the uncensored data and equations 3 and 4 for the restricted data set.

Table 11. Model: $PTO = (U1 - U2) \alpha + (DU2) \beta$

| No | Eq Type | Statistic | α | β | $t\alpha$ $t\beta$ (Z_α Z_β) | 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 | (1059) |
| 3 | OLS | individual ⁽²⁾ | 0.71 | 0.23 | 10.97, 2.64 | 237 | 0.83 |
| 4 | RE* | Individual ⁽²⁾ | 0.47 | 0.46 | 5.85, 4.06 | 237 | (440) |
| 5 | OLS | median | 0.73 | 1.23 | 2.82, 2.53 | 36 | 0.93 |

Key (1) Random effects model; $U = TTO$
(2) Censored data $DU = 1-TTO$

The table indicates that, contrary to expectation, the results were not particularly sensitive to the editing of data. Both 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 coefficient on the latter falling marginally with data censoring and the co-efficient on the latter increasing. Wold and R^2 summary statistics indicate that in all cases the results have very 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 in Table 12. 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.

Figure 4.

Social value by utility gain and severity
Social value: $PTO = (U_1 - U_0)^{0.47} (DU)^{0.46}$

| | Health Gain | | | | |
|--------------------------------------|-------------|------|------|------|-----|
| $(U_1 - U_0)$ DU Severity y | .2 | .4 | .6 | .8 | 1.0 |
| .2 | 0.22 | | | | |
| .4 | 0.31 | 0.42 | | | |
| .6 | 0.37 | 0.51 | 0.62 | | |
| .8 | 0.42 | 0.59 | 0.73 | 0.82 | |
| 1.0 | 0.47 | 0.65 | 0.79 | 0.9 | 1.0 |

The effect of severity is highlighted in Figure 13 which takes the ration of the change in social value to the change in utility. Reading down the same column a gain of 0.2 due to health improvement has rapidly increasingly social value as the severity of the initial health state increases.

Figure 5.

Ratio change in social value to utility gain
Value (D_i) / Utility Gain $(U_1 - U_0)$

| $(U_1 - U_0)$ DU Severity | .2 | .4 | .6 | .8 | 1.0 |
|---------------------------------|-----|-----|-----|-----|-----|
| .2 | 1.1 | | | | |
| .4 | 1.6 | 1.1 | | | |
| .6 | 1.9 | 1.3 | 1.0 | | |
| .8 | 2.1 | 1.5 | 1.2 | 1.0 | |
| 1.0 | 2.4 | 1.6 | 1.3 | 1.0 | 1.0 |

Discussion

Results reported here are robust. Data were elicited very carefully as part of a larger study and, as a form of "sensitivity analysis", respondents were eliminated when there was any suggestion that they did not understand the task. The exclusion criteria were very severe but it did not significantly alter the conclusions.

Results do not, of course, indicate that we must base social policy upon the social values of the population. While this may seem to be a natural conclusion to draw, there is no automatic link between empirically determined values and the ethically acceptable conclusion ("the naturalistic fallacy"), and not all ethicists have taken health disadvantage into account. 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?" (Kymlicka 1990, 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" (Green 1976, p. 117). As we have seen, empirical studies show that many people agree with Kymlicka and Green that inequalities arising from health status – even when they are "natural" rather than "social" - are relevant to prioritising health care, to the extent that people are willing to give priority to the more severely ill even if this means reducing aggregate health gains.

Conclusion

There is a growing body of persuasive empirical evidence indicating that severity of illness is a significant factor for many people when allocating limited health care resources, to the extent that sacrificing overall health production to help the more severely ill is a justifiable option for many people. Our results contribute very significantly to the confidence in these results. There is reason to think that the notion of "effectiveness" used in cost-effectiveness analyses should be expanded to encompass not just potential to benefit from treatment, but a weighting of health gains to reflect the severity of the initial condition of patients. Despite the caveat in the previous section, it is a reasonable rule of thumb that in the absence of ethical or other social-political objections, population values should strongly influence policy.

Empirical studies show that people reject the idea that social value in the health sector is determined solely by increases in length of life and quality of life. When asked to judge for others – when asked to adopt a social perspective – respondents systematically re-weight individual patient preferences according to the severity of the initial health state. As a result, the priority accorded to patients with a severe condition does not reflect the health benefits those patients would personally derive from treatment. Rather, it reflects a social judgement about the distribution of health benefits. As Cohen notes: "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" (Cohen 1995, p. 287).

References

- Abelson, J., J. Lomas, et al. (1995). "Does the Community Want Devolved Authority? Results of Deliberative Polling in Ontario." Canadian Medical Association Journal **153**(4): 403-412.
- Baron, J., Z. Wu, et al. (2000). Analog scale, magnitude estimation, and person trade-off as measures of health utility: biases and their correction: 17.
- Boyd, N. F., H. J. Sutherland, et al. (1990). "Whose Utilities for Decision Analysis?" Medical Decision Making **10**(1): 58-67.
- Callahan, D. (1994). "Setting Mental Health Priorities: Problems and Possibilities." The Milbank Quarterly **72**(3): 451-470.
- Campbell, A. and G. Gillett (1993). Justice and the right to health care. Ethical Issues in Defining Core Services. Wellington, The National Advisory Committee on Core Health and Disability Support Services.
- Cohen, B. J. (1995). "Utility measurement and the allocation of health care resources." Medical Decision Making **15**: 287-288.
- Dolan, P., C. Gudex, et al. (1996). "Valuing health states: A comparison of methods." Journal of Health Economics **15**(2): 209-231.
- Dutch Committee on Choices in Health Care (1992). Choices in Health Care. Rijswijk, Ministry of Welfare, Health and Cultural Affairs.
- Edgar, A., S. Salek, et al. (1998). The Ethical QALY: Ethical Issues in Healthcare Resource Allocations. Haslemere, UK, Euromed Communications.
- Froberg, D. G. and R. L. Kane (1989). "Methodology for measuring health-state preferences." Journal of Clinical Epidemiology **42**(6): 345-354, 459-471, 585-592, 675-685.
- Glick, H. A., D. Polsky, et al. (1999). "A comparison of preference assessment instruments used in a clinical trial." EuroQol EQ-5D and Health Utilities Index **19**(3): 265-275.
- Green, R. M. (1976). Health care and justice in contract theory perspective. Ethics and Health Policy. R. Branson. Cambridge, Massachusetts, Ballinger Publishing Co: 111-126.
- Hurst, N. P., P. Jobanputra, et al. (1994). "Validity of Euroqol - a Generic Health Status Instrument - in Patients with Rheumatoid Arthritis." British Journal of Rheumatology **33**(7): 655-662.
- Kymlicka, W. (1990). Contemporary Moral Philosophy. Oxford, Oxford University Press.
- Lockwood, M. (1988). Quality of Life and Resource Allocation. Philosophy and Medical Welfare. J. M. Bell and S. Mendus. Cambridge, Cambridge University Press: 33-55.
- Menzel, P., P. Dolan, et al. (2002). "The Role of Adaptation to Disability and Disease in Health State Valuation: A Preliminary Normative Analysis." Social Science and Medicine **55**(12): 2149-2158.
- Nord, E. (1991). The use of EuroQol values in QALY calculations. EuroQol Conference, Lund.
- Nord, E. (1992). "Methods for quality adjustment of life years." Social Science and Medicine **34**(5): 559-569.
- Nord, E. (1993a). "The trade-off between severity of illness and treatment effect in cost-value analysis of health care." Health Policy **24**: 227-238.

-
- Nord, E. (1993b). "Helsepolitikere ønsker ikke mest mulig helse per krone (Health politicians do not wish to maximize health benefits)." Journal of the Norwegian Medical Association **113**: 1171-1173.
- Nord, E., J. Richardson, et al. (1993). "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." International Journal of Technology Assessment in Health Care **9**(4): 463-478.
- Nord, E. and M. Wolfson (1999). "Multi-attribute Health State Valuations: Ambiguities in Meaning." Quality of Life Newsletter **21**: 5-6.
- Prades, J.-L. P. (1997). "Is the person trade-off a valid method for allocating health care resources?" Health Economics **6**(1): 71-81.
- Rawls, J. (1971). A Theory of Justice. Cambridge, Harvard University Press.
- Richardson, J. (1994). "Cost Utility Analysis: What Should Be Measured?" Social Science and Medicine **39**(1): 7-21.
- Rothman, M. L. (1991). "The Validity of Proxy-Generated Scores as Measures of Patient Health Status." Medical Care **29**(2): 115-124.
- Savulescu, J. (1998). "Consequentialism, Reasons, Value and Justice." Bioethics **12**(3): 212-235.
- Swedish Health Care and Medical Priorities Commission (1993). No Easy Choices: The Difficulties of Health Care. Sveriges offentlige utredningar, 1993. Stockholm, The Ministry of Health and Social Affairs.
- Wikler, D. (1989). "Equity, efficiency, and the point system for transplant recipient selection." Transplantation Proceedings **21**(3): 3437-3439.

Appendix

Table 12: Summary Statistics

| 18 e values | | 36 change scores | |
|-------------|------|-----------------------|-------|
| Mean TTO | 0.70 | Mean TTO | 0.106 |
| Mean PTO | 0.77 | Mean PTO | 0.062 |
| Median PTO | 0.08 | Median PTO | 0.052 |
| PTO > TTO | 14 | PTO > TTO | 7 |
| PTO < TTO | 3 | PTO < TTO | 19 |
| PTO = TTO | 1 | PTO = TTO (± 2) | 10 |

Table 13: Descriptive Statistics

| Health State | Mean TTO | | PTO uncensored | | | PTO censored | | | |
|--------------|----------|------|----------------|------|------|--------------|------|-----|----------------|
| | n | mean | n | mean | min | n | mean | min | Point censored |
| H1 | 65 | .82 | 65 | .50 | .01 | 17 | .84 | .59 | .50 |
| H2 | 65 | .88 | 65 | .51 | .02 | 17 | .90 | .71 | .67 |
| H3 | 65 | .86 | 65 | .50 | .01 | 17 | .82 | .5 | .50 |
| H4 | 52 | .64 | 50 | .36 | .005 | 14 | .71 | .95 | .29 |
| H5 | 52 | .72 | 50 | .44 | .005 | 14 | .77 | .57 | .57 |
| H6 | 51 | .50 | 50 | .30 | .005 | 14 | .66 | .33 | .33 |
| H7 | 50 | .79 | 54 | .46 | .02 | 20 | .68 | .5 | .40 |
| H8 | 50 | .74 | 54 | .47 | .02 | 20 | .75 | .5 | .40 |
| H9 | 50 | .83 | 54 | .51 | .02 | 20 | .77 | .57 | .56 |
| H10 | 51 | .54 | 50 | .32 | .01 | 7 | .70 | .5 | .50 |
| H11 | 51 | .83 | 51 | .60 | .02 | 7 | .91 | .83 | .71 |
| H12 | 51 | .70 | 51 | .51 | .01 | 7 | .89 | .83 | .56 |
| H13 | 44 | .70 | 41 | .41 | .01 | 11 | .76 | .4 | .36 |
| H14 | 44 | .61 | 41 | .39 | .01 | 11 | .79 | .5 | .33 |
| H15 | 44 | .82 | 41 | .51 | .04 | 11 | .87 | .77 | .71 |
| H16 | 43 | .47 | 44 | .31 | .01 | 10 | .64 | .3 | .33 |
| H17 | 43 | .60 | 45 | .34 | .01 | 10 | .65 | .4 | .33 |
| H18 | 43 | .6 | 45 | .40 | .08 | 10 | .69 | .5 | .40 |