

---

ECONOMIC EVALUATION

---

# INCORPORATING SOCIETAL CONCERNS FOR FAIRNESS IN NUMERICAL VALUATIONS OF HEALTH PROGRAMMES

ERIK NORD<sup>a,\*</sup>, JOSE LUIS PINTO<sup>b</sup>, JEFF RICHARDSON<sup>c</sup>, PAUL MENZEL<sup>d</sup> AND PETER UBEL<sup>e</sup>

<sup>a</sup> *National Institute of Public Health, Oslo, Norway*

<sup>b</sup> *Department of Economics, Universitat Pompeu Fabra, Barcelona, Spain*

<sup>c</sup> *Centre for Health Program Evaluation, Melbourne, Australia*

<sup>d</sup> *Department of Philosophy, Pacific Lutheran University, Tacoma, Washington, USA*

<sup>e</sup> *Center for Bioethics, University of Pennsylvania, Philadelphia, Pennsylvania, USA*

## SUMMARY

The paper addresses some limitations of the QALY approach and outlines a valuation procedure that may overcome these limitations. In particular, we focus on the following issues: the distinction between assessing individual utility and assessing societal value of health care; the need to incorporate concerns for severity of illness as an independent factor in a numerical model of societal valuations of health outcomes; similarly, the need to incorporate reluctance to discriminate against patients that happen to have lesser potentials for health than others; and finally, the need to combine measurements of health-related quality of life obtained from actual patients (or former patients) with measurements of distributive preferences in the general population when estimating societal value. We show how equity weights may serve to incorporate concerns for severity and potentials for health in QALY calculations. We also suggest that for chronically ill or disabled people a life year gained should count as one and no less than one as long as the year is considered preferable to being dead by the person concerned. We call our approach 'cost-value analysis'. Copyright © 1999 John Wiley & Sons, Ltd.

**KEY WORDS**—QALY; cost-utility analysis; health programme evaluation; fairness; severity of illness; potential for health; person trade-off; cost-value analysis

## INTRODUCTION

QALYs are a measure of the volume of health output. As noted by many [1–8], society's overall valuation of health output is a function not only of total output, but also of the distribution of health output across individuals. More specifically, society may be prepared to make some sacrifices in the total production of health in order to secure a fair or equitable distribution of health. To encapsulate such distributive concerns, economists have proposed to assign *equity weights*

to QALYs according to characteristics of their recipients [9–11]. The aim of resource allocation in health care would then be to maximize the sum of equity weighted QALYs rather than an unweighted sum [10,12,13].

The term *health-related societal value* may be used to designate the overall value that society assigns to different health outcomes and programmes when concerns for both efficiency and equity are taken into account. Equity weighted QALYs are thus measures of health-related societal value.

---

\* Correspondence to: National Institute of Public Health, P.O. Box 4404 Torshov, N-0403 Oslo, Norway. Tel.: +47 22 042342; fax: +47 22 042595

One possible approach to establishing equity weights is to measure the strength of society's aversion to inequality in health outcomes. Wagstaff [14] specifies a social welfare function that includes a parameter reflecting the strength of this aversion. In principle, the parameter may be estimated by asking members of society to express preferences between different combinations of total health production and health distribution. Results from a pilot study of such preferences were reported by Dolan [15].

Williams [16] argues that the most salient ethical basis for the introduction of equity weights is the *fair innings argument*, i.e. the general sentiment that everyone is entitled to a 'normal' life-time around 70–75 years and that anyone failing to achieve this has in some sense been cheated, while anyone getting more than this is living on 'borrowed time' (for a more detailed argument, see [17]). Williams addresses the fact that there is a significant difference between social classes in the UK with respect to quality adjusted life expectancy (QALE) at birth. Adopting a social welfare function of the kind suggested by Wagstaff, Williams gives a hypothetical example of how observations of people's willingness to trade-off mean QALE for equality in QALE could be used to estimate a parameter for the strength of aversion to inequality. He goes on to present a hypothetical table of equity weights for QALYs that could be used in comparisons of programmes for socioeconomic groups that differ with respect to QALE at birth.

While the fair innings argument may provide a clear ethical basis for introducing equity weights in QALY calculations, there are several other possible bases for such weights. The purpose of the present paper is to address two other concerns for fairness that perhaps indicate even more strongly a need to weight QALY gains according to characteristics of their recipients: severity of illness and limitations in potential for health. We outline a methodology for establishing equity weights to reflect these concerns that represents a possible alternative to the direct efficiency–equity trade-off questions suggested by Wagstaff, Dolan and Williams.

Although throughout the paper we focus on QALYs, our points also apply to the Healthy Year Equivalents (HYEs) procedure, which assigns values to health scenarios rather than health states [18]. HYE are just like QALYs based on

individual (personal) utility assessments by means of the standard gamble or time trade-off, none of which allow respondents to take into account concerns for equity across individuals.

### AN EXAMPLE WITH THREE HYPOTHETICAL PROGRAMMES

Consider Figure 1. A, B and C are three groups of patients who on average are alike in all respects except that they have different illnesses. (Note for the argument below that their likeness includes being equal (on average) with respect to preferences for health states.) Their health-related quality of life (HRQOL) with and without treatment is measured at an interval level of measurement by means of the standard gamble or the time trade-off. A scale from zero to unity is used to express this HRQOL, often referred to as health-related utility (hereafter mostly 'utility' for brevity). The bottom end points of the three vertical lines to the right of the scale indicate the utility of patients in the three groups when untreated. The top end points indicate their utility after treatment. (In the following we will for brevity generally refer to the bottom end points as 'start points' and the top end points simply as 'end points'.)

Assume that the three groups have the same life expectancy without treatment and that none of the treatments have any effect on life expectancy. Thus the health improvements are pure health-related quality of life improvements. Assume also that treatments costs per patient are the same in all three groups.

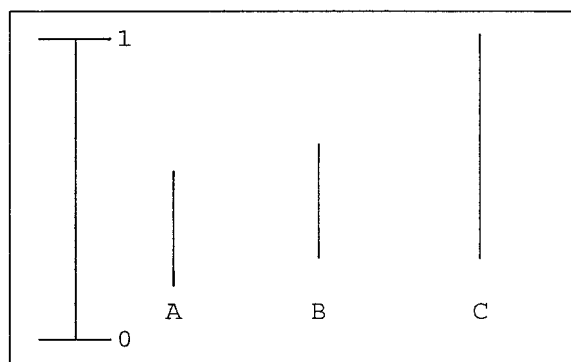


Figure 1. Three improvements in health on a 0–1 scale

Given these assumptions, Figure 1 tells us first that patients A are *more severely ill* than patients B and C (in the sense that their health-related utility is lower) when no treatment is provided. Second, it is indicated that the same amount of resources will produce the same benefit—in terms of individually assessed utility—in A and B. Third, while patients in B and C experience the same low level of health-related utility when untreated, a given amount of resources will produce a larger benefit (individual utility gain) in C than in B (the reason being that the groups have different illnesses with different treatability). For brevity we may say that group C has a greater potential for health than group B.

If the goal is to maximize QALYs, group C will be given first priority, while A and B will share second place.

There are two concerns for fairness that we believe are particularly salient in priority setting in health care, that run counter to this rank ordering of groups A, B and C.

#### *Severity of pre-treatment condition*

First, group A is worse off without treatment than groups B and C and may, therefore, be considered to have a stronger claim to being helped than groups B and C, even if the potential utility gains in the latter groups are equally high (B) or higher (C). Rawls [19] argued this point forcefully and is supported, for instance, by Callahan [20]: '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'. In the following we shall for brevity refer to this concern for fairness as the *severity argument*.

Some may argue that the extent to which people have a concern for severity will be captured in their initial personal utility assessments of health states. This is only partly true. In utility assessments, for instance by means of the standard gamble or the time trade-off, subjects are not asked about distributive concerns or societal value. They are asked to quantify the disutility they personally would feel with different states of illness, and thus—indirectly—to quantify the utility gain they personally would derive from treatment. On the basis of responses to such

questions it is possible only to say which of different programmes provides the greatest sum of individual utility gains. The respondents have not expressed any opinion about priority setting. If one knew that their view on priority setting was that the sum of utility gains should be maximized, then one could infer from their utility responses that, for instance in the above example, patients C should have priority over patients A and B. But that doesn't have to be their view. It is perfectly conceivable that a person could think as follows: 'I accept that patients in groups A and B seem to value their respective treatments equally much for themselves (in terms of willingness to sacrifice life expectancy). Nonetheless, if I have to make a choice, I would give priority to group A over group B, since they are worse off, and perhaps even over group C, in spite of the greater potential benefit in the latter group'. In fact, Callahan demonstrates this way of thinking in the above quotation. We shall see below that he is far from the only one.

#### *Realization of potential for health*

The other concern for fairness we wish to draw attention to, has to do with the comparison between B and C. In terms of QALYs gained, C scores better than B. However, it may be seen as unfair to hold against patients in group B that they happen to have a lesser degree of treatability than group C. Their potential for health improvement is still substantial and important to themselves, and they are just as ill as patients in group C.

Medical ethicists have addressed an analogous problem—referred to as 'the fair chances/best outcomes problem'—in so-called 'microrationing' (i.e. rationing at the individual level) for instance in the allocation of scarce organs for transplantation [17,21–24]. Assume for instance that two people, A and B, are the same age, have waited in queue the same length time and that each will live only 1 week without a heart transplant. With the transplant, however, A is expected to live 2 years and B 20 years. Who should get the transplant? Daniels [24] and Harris [25] discuss a similar problem in 'macrorationing': an intervention A preserves life in a given group of people and restores these to full health, while another service B preserves life in a different group of people but leaves these in a state of disability. All else equal,

should A have priority? In both the micro- and the macrorationing case one may ask why those with a lesser outcome should give up their chances of receiving something valuable to them just because somebody else can benefit even more. Harris [25] argues strongly against such discrimination, while Daniels [24], when reviewing the ethical debate in the literature, concludes that in neither of the above contexts is it clear what would be a fair allocation rule. He reaches this conclusion even after considering arguments invoking people's self-interest behind a veil of ignorance. In a response to Harris [25], Williams [26] seems to concede that the issue has no clear logical answer, which leads him to conclude that 'at the end of the day, we simply have to stand up and be counted as to which set of principles we wish to have underpin the way the health care system works'.

Interestingly, Williams [16] notes that equity weights could be used to secure fair innings—in terms of QALYs—not only for disadvantaged social classes, but also for the permanently disabled. For the latter, the fair innings argument coincides with the argument that it should not be held against the permanently disabled that life years provided to them bring less QALYS. Hence, their QALYs need to be given extra weight.

Altogether, ethical reflection seems to suggest that society may not want to discriminate as strongly against those with a lesser potential as pure QALY maximisation would suggest. We shall refer to this as the *realization-of-potential argument*.

#### *Aversion to inequalities in health*

There is a third aspect of fairness which needs to be noted here, even if it is beyond the scope of this paper to examine it in detail. Assume that only one of the groups A, B and C in Figure 1 can be treated, and that group C is given priority over groups A and B. Then inequality in health between the three groups, i.e. the differences between the resulting end states, will be greater than if group B or particularly group A were given priority. Such increased inequality may also be seen as undesirable.

Note that aversion to inequality may occur as an independent concern. While it will necessarily work in the same direction as the realization-of-potential argument, it may both support and run counter to the severity argument, depending on

the health outcomes in question. What we above called the severity argument thus refers to the concern for the worst off *per se*.

Aversion to inequality between end states is not a theme of this paper, the reason being that it has not been much explored quantitatively. It is, however, certainly an issue for further research.

#### WHAT DO PREFERENCE DATA SAY ABOUT SEVERITY AND POTENTIAL?

A number of researchers have followed Williams' suggestion that people be asked to stand up and have their vote counted concerning health care allocation principles. Their message regarding the severity argument is clear. In a number of studies using the person trade-off technique [27,28] the argument consistently receives strong support [29–34]. We also note a recent direct study of the efficiency–equity trade-off by Dolan [35]. He found in a group of economics students that subjects preferred moving one individual from utility level 0.2 to 0.4 to moving another individual from 0.4 to 0.6, and that the latter movement would rather have to be from 0.4 to 0.8 to be equally preferable as the movement from 0.2 to 0.4. This preference may derive from an aversion to inequality between end states. But it may also reflect concerns for the severity of the pre-treatment state.

The experimental evidence regarding the concern for realization of potential is less extensive and also more ambiguous [36–39]. A fair summary seems to be that of a study in Spain by Pinto and Perpinan [40], namely that capacity to benefit is viewed as relevant by most people, but not as strongly so as the QALY maximization approach assumes. This accords with recent results from a series of focus group discussions in North Yorkshire in England, from which the conclusion was drawn that 'there should be priority accorded to those in urgent need of medical attention. Whilst capacity to benefit does matter, it is a secondary consideration' [41]. It is also consistent with actual policy making in the US: the proposal to determine a priority ranking of medical procedures in Medicaid in the state of Oregon on the basis of cost-per-QALY calculations was rejected partly because of the discriminatory effect such a ranking was thought to have on the permanently disabled.

Table 1. Societal values for health improvements

To problem level: From problem level:	7	6	5	4	3	2	1
1. None (healthy)							0.0001
2. Slight						0.009	0.01
3. Moderate					0.07	0.079	0.08
4. Considerable				0.12	0.19	0.199	0.20
5. Severe			0.15	0.27	0.34	0.349	0.35
6. Very severe		0.25	0.40	0.52	0.59	0.599	0.60
7. Completely disabled	0.40	0.65	0.80	0.92	0.99	0.999	1.00
8. Dead							

Examples at levels 2–7:

2. Can move about anywhere, but has difficulties with walking more than 2 km.
3. Can move about without difficulty at home, but has difficulties in stairs and outdoors.
4. Moves about with difficulty at home. Needs assistance in stairs and outdoors.
5. Can sit. Needs help 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. Permanently bedridden.

Source: [43].

Altogether we suggest that the realization-of-potential argument has too much public support to be ignored in a model of societal preferences for resource allocation.

## MODELLING HITHERTO

Societal concerns for severity and realization of potential may be quantified by asking people to value different improvements in health across individuals directly relative to each other. In this approach, the object of valuation is in other words *changes in health* rather than health states as such. The value of restoring a person from a life-threatening condition to full health (which many will think of as the largest possible individual health outcome) is used as numeraire [42]. The feasibility of this approach was demonstrated in a pilot study which resulted in a two-dimensional table that indicated the societal value of different health improvements as a function of their start and their end points [37]. A revised value table, based on an overall judgement of the various preference data referred to in the preceding section, was later published in the *Journal of the Norwegian Medical Association* [43] and is reproduced in Table 1. We emphasize that the method of synthesis was informal, the ambition being only to indicate quite roughly what seems to be a widespread societal structure of concern.

The severity scale in Table 1 is a modified version of a scale constructed by Sintonen [44]. The descriptions were chosen with a view to making each step up on the scale appear equally significant in terms of individual utility. With a few exceptions, subjects involved in the pilot study [37] said that they perceived the intervals as quite equal in this sense. The states described at each level were also mapped into two different health status indexes, one based on magnitude estimation and category rating, the other on standard gamble and time trade-off. Both mappings supported the impression that the 7-point severity scale has fairly equal intervals in terms of individual utility (see [37] for details).

The concern for severity comes clearly through in the upper diagonal in Table 1: one step up on the scale is valued more highly (and much more so) the lower the start point. The concern for realization of potential comes through in each horizontal line: a movement from any given start point scores better the higher the end point, but marginal value decreases significantly with increasing treatment effect. For instance, a person with a potential to go from level 7 to level 4 will score almost as much as a person with a potential to go all the way from level 7 to level 1.

Table 1 corresponds to a set of health state values as shown in Table 2. If we accept that the 8-point scale approximates an equal-interval one in terms of individual utility, Table 2 shows decreasing marginal societal value of utility gains.

The numbers correspond to a curve for societal health state values as a function of individual utilities that is convex to the  $y$ -axis.

The reader will note that the states in question are very strongly compressed towards the upper end of the 0–1 scale—more so than is normal for comparable states in main stream cost-utility analysis. The reason is of course that the valuation perspective and procedure are different ones. Many readers will have seen this ‘upper end compression’ earlier in the Rosser and Kind Matrix [29], in which most combinations of disability and distress are assigned values above 0.95. A distribution of disability weights with somewhat similar upper end compression is suggested for calculating DALYs in the Global Burden of Disease Project [45].

The technical point to note here is that if one wishes to incorporate concerns for severity and health potential in the assignment of values to health outcomes, one does not need to have recourse to a two-dimensional table of changes in health like Table 1. The same effect may be obtained by adopting convexity (decreasing marginal value) and strong upper end compression in the assignment of values to health states. As shown elsewhere [32,46–48], the ‘failure of cost-effectiveness analysis in Oregon’ [49] may largely be attributed to the lack of upper end compression in the health state values used to calculate QALYs. If values with convexity and strong upper end compression are used to weight life years, there is no need to add separate equity weights to account for concerns for severity and realization of potential. Instead, QALYs themselves become a comprehensive measure of societal value rather than a measure of the simple sum of individual health gains.

Table 2. Health state values encapsulating concerns for severity and realisation of potential

Problem level	Value
1. Healthy	1.00
2. Slight problem	0.9999
3. Moderate problem	0.99
4. Considerable problem	0.92
5. Severe problem	0.80
6. Very severe problem	0.65
7. Completely disabled	0.40
8. Dead	0.00

The comprehensive approach to valuing movements between health states is not without weaknesses. In particular, it may be seen as unfortunate to try to encapsulate concerns for several different aspects of response to health care—in this case initial severity, potential for health and the actual health gain—in one single set of numbers, as health state valuations with convexity and strong upper end compression purport to do. Some may prefer to make the nature and the extent of the efficiency–equity trade-off explicit by adopting a decomposed approach, in which separate equity weights are introduced for distributive concerns (see for instance Dolan [50]).

The decomposed approach leads to a multifactorial model that looks quite complicated and, therefore, may not be more attractive to potential users than a model using comprehensive health state values with strong upper end compression. We return to this issue in the Discussion section. Like Dolan, we would encourage further research efforts looking at both approaches. In the following we explore the feasibility of the decomposed approach. We outline a procedure for establishing one set of equity weights that would encapsulate concerns for severity of illness and one set that would encapsulate concerns for realization of potential. These weights do not purport to replace the fair innings based weights suggested by Williams [16]. They serve different purposes. As noted above, they are also based on a different method for preference measurement.

## A TWO STEP VALUATION PROCEDURE

One of the controversial issues in the debate about QALYs has been: from whom should utilities for health states be elicited? The approach adopted by most researchers today, for instance in the construction of multi-attribute health scaling instruments, is to describe a number of hypothetical states of illness and to ask a random sample of the general population to express how good or bad they think it would be for themselves to be in each of those states. This approach has been criticized on the grounds that the best judges of the disutility of states of illness must be people who are actually in those states.

We believe the debate about whom to ask has become unduly difficult to settle because of a conflation of two issues. One is the measurement

of quality of life associated with different health problems. The other is the measurement of distributive preferences in resource allocation. When QALYs are thought of as a measure of societal value—as has often been the case historically (see above)—utilities have direct implications for resource allocation across diagnostic groups. The case for eliciting utilities from the general population then becomes strong, since they are all potential patients with an interest in how resources are distributed. The problem looks different, however, if it is accepted that QALYs are one thing, societal value another. In societal valuation of a health programme, a utility gain of a given size may receive different weighting depending on characteristics of the patients who get it.

One can then envisage a two step procedure for constructing a societal value model. The first step consists of measuring the severity of different health states in terms of utility. On the basis of such measurements, the utility gains associated with different health interventions may be estimated. The second step is to assign weights to different utility gains, taking into account for instance societal concerns for the severity of the patients' initial condition, the patients' potential for health, their age or whatever other factors the public might consider to be of importance in an overall judgement of societal value. This second step is essentially about measuring distributive preferences. As noted above, there is a strong case for eliciting these from the general public. It is also possible to do the first step (utility measurement) by asking members of the general public to evaluate hypothetical health states. But with the separation of utility measurement from the measurement of distributive preferences, the case for asking the general population to assign utilities to states of illness is less clear. An important qualification for judging the individual disutility of different states of illness (that is without taking on board distributive concerns) must in our view be personal experience with those states. While there is much such experience in the general population, it is spread quite thinly across illnesses. Consequently, when a general population sample is asked to evaluate a particular hypothetical state, in most cases only a small share of the respondents will have had personal experience with that state. Arguably, more valid judgements of the disutility of a state of illness may be obtained by placing greater emphasis on data collected from people who are or have been in that state.

We are aware that there is much more to be said about the 'whom to ask' issue (for example concerning adaptation phenomena and coping mechanisms). However, it is beyond the scope of the present paper to discuss these other arguments in full, let alone to attempt to 'resolve' the issue once and for all. Instead we make a choice that we believe is of theoretical interest. The procedure we outline in the following uses patients (or former patients) and disabled people as subjects in utility measurement and representatives of the general population as subjects in the measurement of concerns for severity and realization of potential. In the latter step, subjects are asked to express their distributive preferences on the basis of information about the disutility associated with different health problems as reported by people who have or have had those problems. The subjects are thus meant to be informed members of the general population. This is in line with official recommendations in the US [51] (p. 106).

#### A MULTIPLICATIVE MODEL OF SOCIETAL VALUE

Consider four treatment programmes A, B, C and D aimed at persons with different illnesses and/or conditions. Assume that each programme represents the best one can do for these patients given the state of art in medical knowledge and technology. In other words, each programme realizes the potential for health in the patients in question. The programmes could for instance address persons with heart disease, arthritis, asthma and lower back pain, respectively, and the interventions could be anything from drug treatments to surgical procedures or rehabilitation programmes. Using patients' self-judgements, the initial and end state utilities, and thus the utility gains of treatment, are estimated for the average patients as in columns 2–4 of Table 3 (we return to the remaining columns).

Assume that these four programmes are equal with respect to all other aspects (duration, age, number of people treated etc.) than the patients' utilities before and after treatment. According to conventional utility analysis, programme D is then the most valuable one. C is the second most valuable, and A and B share third place. As shown above, concerns for severity and realization of potential may, however, lead to a different rank ordering in a societal priority setting context.

Table 3. Severity, effect and societal value in four programmes

Program	Utility			<i>SW</i>	<i>RPR</i>	<i>PW</i>	Societal value
	Initial ( $U_1$ )	End ( $U_2$ )	Gain				
A	0.60	0.70	0.10	0.20	1/4	2.5	0.05
B	0.70	0.80	0.10	0.10	1/3	2.0	0.02
C	0.85	1.00	0.15	0.03	1/1	1.0	0.0045
D	0.60	0.80	0.20	0.20	1/2	1.5	0.06

In order to encapsulate these two concerns for fairness, the societal value of an improvement from utility level  $U_1$  to utility level  $U_2$  could be modelled mathematically in a number of ways. Here we restrict ourselves to indicating the feasibility of a multiplicative model, which would be in keeping with the multiplicative approach used to combine gains in life years and quality of life in the conventional QALY model. Our multiplicative model is:

$$SV = dU \times SW \times PW$$

where  $SV$  stands for societal value,  $dU$  represents the utility gain ( $U_2 - U_1$ ),  $SW$  is a weight determined by the severity of the initial condition and  $PW$  is a weight determined by the potential for health.

$SV$  may be measured on 0–1 scale just like conventional individual utility. To achieve this, the most practical arrangement is to set  $SW$  at 1 for  $U_1 = 0$  (on the verge of dying), to let it decrease with increasing values of  $U_1$  and to let it be 0 for  $U_1 = 1$  (healthy). Table 4 gives an illustration.

The factor  $PW$  needs some explanation. Representing the realization-of-potential argument, the

purpose of  $PW$  is to attenuate the discrimination between patients with different potentials for health (different capacities to benefit) implied by the conventional QALY maximization approach. To achieve this, one may either devalue utility gains for patients with relatively large potentials, or revalue utility gains for patients with relatively small potentials. Here we choose to do the latter, since it makes it easier to keep the measurement of societal value ( $SV$ ) within the conventional 0–1 range. We focus on the ratio  $(U_2 - U_1)/(1 - U_1)$ , which expresses the ratio between the actual potential in a given patient group with initial utility level  $U_1$  (their maximum capacity to benefit given the state of art in medical technology) and the theoretically greatest possible health gain in patients with that initial utility level. The theoretically greatest possible health gain is of course the movement from the initial level all the way up to full health. We may call the ratio above the relative potential ratio ( $RPR$ ).

To keep  $SV$  within the conventional 0–1 range,  $PW$  needs to be set at 1 when  $U_1 = 0$  and  $U_2 = 1$  (since  $dU$  and  $SW$  then both equal 1).  $RPR$  in this case equals 1, so we have  $PW = 1$  when  $RPR = 1$ . As noted above, we want to revalue (upgrade) utility gains in patients groups with lesser potentials.  $PW$  must, therefore, increase with falling values of the relative potential ratio. Table 5 gives an illustration.

Columns 4–8 in Table 3 shows how all this might work for programmes A–D. The severity and potential weights are illustrative (although they are chosen so as to be roughly consistent with the societal value structure indicated in Table 1). The relative potential ratios follow from columns 2 ( $U_1$ ) and 3 ( $U_2$ ). With the chosen weights, programme C falls to the bottom of the societal value ranking list, and B falls below A. Both these changes are due to the incorporation in the valuation model of an independent societal

Table 4. Severity weights for different levels of initial utility (illustration)

Initial utility	Severity weight
0.0	1.00
0.1	0.80
0.2	0.65
0.3	0.50
0.4	0.40
0.5	0.30
0.6	0.20
0.7	0.10
0.8	0.05
0.9	0.01



Table 5. Potential weights for different relative potential ratios (illustration)

<i>RPR</i>	Potential weight
1.0	1.00
0.9	1.05
0.8	1.15
0.7	1.30
0.6	1.45
0.5	1.60
0.4	1.80
0.3	2.00
0.2	2.50
0.1	4.00

preference for giving priority to the worst off. Furthermore, while D remains above A, the difference between A and D is greatly reduced. This change stems from the incorporation in the model of societal concerns for realization of potential. The consequences for resource allocation across patient groups are substantial. In terms of utility gains (column 4), ten people treated in programme D would be equivalent to 20 people treated in programme A and B and 13 people treated in programme C. According to the societal value estimates in the final column, the correct equivalence numbers would rather be 12, 30 and 133, respectively.

The question then, is: how might severity weights and potential weights such as those exemplified in Tables 4 and 5 be established empirically?

### SOCIETAL VALUE JUDGEMENTS

There are a number of ways in which one may ask people to value different movements on a health scale relative to each other with a view to capturing concerns for fairness. One, suggested by Dolan [35], was noted above: subjects are for instance asked to compare an improvement in utility from 0.2 to 0.4 for one group of patients with an improvement from 0.4 to some level  $X$  for an other group of patients. What must  $X$  be for the two improvements to be valued equally in a resource allocation decision? If, for instance,  $X$  on average is set at 0.8, then one might infer that the subjects on average assign a severity weight to utility level 0.2 which is twice that of utility level 0.4.

While Dolan's direct focus on 'equally valuable movements on a utility scale' is feasible and interesting in the case of severity weights, the reader can easily verify that it will not work with potential weights (for instance in comparing a movement from 0.2 to 0.4 with a movement from 0.2 to 0.8). The relative value of different movements must then be expressed on an independent scale. Here we shall mention three possibilities. One is magnitude estimation: how much more valuable is for instance the movement in programme C in Table 4 than the movement in programme A? Another is willingness to pay: as a tax payer, how much would you be willing to contribute to the treatment of say 100 patients in programme C compared to 100 patients in programme A? (We are aware that this is a rather unusual framing of a willingness to pay question.) A third is person trade-off: if, for a given amount of money, either 100 people can be treated in programme C or  $X$  people can be treated in programme A, what number must  $X$  be for you to consider the two programmes equally deserving of funding?

We hold a preference for the person trade-off technique, based on two concerns.

The first is for clarity of meaning. Magnitude estimation begs a question: if somebody for instance says that the movement in programme C is 50% more valuable than the movement in programme A, what does that imply? The willingness to pay question and the person trade-off do not have this problem. In fact, the answer to the question about the meaning of magnitude estimation might be that the person is willing to pay 50% more (in tax contribution) for the movement in programme C than the movement in programme A, or that the subject considers treating 150 people in programme A equivalent to treating 100 people in programme C. Both these are perfectly meaningful statements. But to achieve this degree of meaningfulness, it would be necessary to ask the willingness to pay question or the person trade-off question rather than the magnitude estimation question.

Our second concern is for directness of measurement. We are assuming that—all else being equal—the value of a programme is a product of (a) the value of the movement that occurs in that programme and (b) the number of people that get to enjoy it. There is then a trade-off between these two factors. For example, if movement A is

regarded as only half as valuable as movement C, then that would be compensated for in programme A if it benefited twice as many patients. The willingness to pay question establishes this person trade-off indirectly: if a subject says that he is willing to pay twice as much for each movement C than for each movement A, that is taken to mean that he would be indifferent between treating 200 people in programme A and 100 people in programme B. However, that is an inference, the validity of which in principle would need to be verified. The way to do that would be to ask a direct person trade-off question. Arguably, then, one may as well ask a direct person trade-off question in the first place.

A parallel here would be two approaches to finding out which of two persons is taller. One can either measure them both with the same yard stick. Or one can judge by letting them stand back to back. Arguably, the risk of random measurement error is greater when two measurements are involved rather than one.

We do not wish to rule out the willingness to pay approach or Dolan's direct measurement of the severity/health gain trade-off as potentially useful ways of eliciting societal valuations. However, in the following we have chosen to outline a strategy for establishing societal concerns for fairness by means of person trade-off questions.

### SEVERITY WEIGHTS

Severity weights may be estimated by asking a representative sample of the general population to make person trade-offs between movements that are equal in terms of utility gains, but different in terms of start points. Paired comparisons would be chosen so as to cover the whole range of the 0–1 utility scale. So for instance, one would compare a utility movement from 0.9 to 1.0 (in the following referred to as '0.9 → 1.0') with a utility movement from 0.8 to 0.9 (0.8 → 0.9); '0.5 → 0.7' with '0.7 → 0.9'; '0.2 → 0.3' with '0.4 → 0.5' etc. At least one of the paired comparisons would include a movement that has zero as a start point (the patient dies if not treated), for instance '0.0 → 0.5' compared to '0.5 → 1.0'.

To see the mathematics of the severity weights, consider for instance the following three movements on the utility scale:

Movement	Utility
<i>X</i>	0.0 → 0.3
<i>Y</i>	0.3 → 0.6
<i>Z</i>	0.6 → 0.9

Subjects would be asked person trade-off questions of the following kind: in judging different areas in which to increase treatment capacity, how many *Y*'s would be equivalent to 10 *X*'s? How many *Z*'s would be equivalent to 10 *X*'s? Assume that the median answers to these questions were: 10 *X* = 20 *Y*; 10 *X* = 50 *Z*. As noted above, the severity weight for *X* (life saving) would be set at 1. Then the severity weights' for movement *Y* relative to *X* would be  $1 \times (10:20) = 0.50$ . The severity weight for movement *Z* would be  $1 \times (10:50) = 0.20$  (e.g. programme A and D in Table 3).

As noted earlier, our proposal is based on the value judgement that societal preferences for resource allocation in health care should be elicited from an *informed* public. To achieve this, the description of movements in the paired comparisons above would not be in terms of abstract utilities, but rather in as rich verbal terms as possible without overloading the subjects cognitively. For instance, the description of each movement would specify a certain level of health problems in terms of functioning, physical discomfort and psychological distress before and after treatment. The description would furthermore include the level of satisfaction and happiness associated with those levels of health problems (as suggested by data collected from people who have experience with those problems).

### POTENTIAL WEIGHTS

Potential weights may be estimated in a similar way. A representative sample of the general population could be asked to make person trade-offs between movements that are equal in terms of severity (start point), but different in terms of utility gains. Again, paired comparisons would be chosen so as to cover the whole range of possible utility gains (0–1). So for instance, one would compare '0.8 → 1.0' with '0.8 → 0.9'; '0.5 → 0.9'; with '0.5 → 0.7'; '0.2 → 1.0' with '0.2 → 0.5' etc.

Some of the paired comparisons would include movements that have zero as a start point (the patient dies if not treated), for instance '0.0 → 1.0' compared to '0.0 → 0.5'.

To see the mathematics of the potential weights, consider for instance the following three movements on the utility scale:

Movement	Utility	RPR
<i>X</i>	0.4 → 0.6	1/3
<i>Y</i>	0.4 → 0.8	2/3
<i>Z</i>	0.4 → 1.0	1/1

Subjects would be asked person trade-off questions of the following kind: in judging different areas in which to increase treatment capacity, how many *Y*'s would be equivalent to 10 *Z*'s? How many *X*'s would be equivalent to 10 *Z*'s? Assume that the median answers to these questions were: 10 *Z* = 13 *Y*; 10 *Z* = 20 *X*. As noted above, the potential weight for *Z* would be set at 1. The potential weight (*PW*) for movement *Y*—for which *RPR* equals 2/3—would then be given by

$$13 \times (0.8 - 0.4) \times PW_Y = 10 \times (1.0 - 0.4) \times PW_Z$$

where  $PW_Z = 1$ , such that  $PW_Y = 1.15$ . This could then be used as a potential weight for movements for which *RPR* equals 2/3. Similarly, the potential weight for movement *X*—for which *RPR* equals 1/3—would be given by:

$$20 \times (0.6 - 0.4) \times PW_X = 10 \times (1.0 - 0.4) \times PW_Z$$

$$PW_X = 1.5$$

This could then be used as a potential weight for movements for which *RPR* equals 1/3.

## PRESENTATION FOR USERS

Severity weights and potential weights based on population preference measurements could be made available for analysts in standard tables that would be combined with utility data in order to allow calculations of societal value (as was done in Table 3). Tables 4 and 5 are illustrations.

## DISCUSSION

The procedure outlined above has two main virtues. One is that it separates the measurement of utility from the measurement of societal value. By doing this, a clearer case is made for eliciting utilities directly from patients and disabled people. This accommodates a concern about QALYs that has been expressed by many, including the organizations of patients and the disabled, who understandably are sceptical about having healthy people make judgements about quality of life on their behalf.

The other main virtue is that the procedure incorporates in the quantification of societal value of health care two concerns for fairness that many regard as no less important than the maximization of total utility, namely the preference for giving priority to those who are worst off, and the reluctance to discriminate strongly against patients who happen to have lesser potentials for health than others.

Admittedly, measuring the strength of these concerns for fairness will not be easy. However, estimating utility in conventional cost-effectiveness analysis is in itself a difficult task—so difficult that health economists after several decades of work are still unable to make clear recommendations regarding which estimation techniques to use [51], even when these techniques produce quite different results [31]. We suggest that the methodological difficulties involved in quantifying concerns for severity and potentials for health are not prohibitively greater than those associated with measuring utility. Given the documented, high relevance of these concerns for policy decisions we, therefore, see no reason for health economists to leave them out of economic evaluations of health care outcomes.

We emphasize that the purpose of severity weights and potential weights is to allow better estimates of society's appreciation (valuation) of the spending of resources on different patient groups. The weights are not relevant in evaluations of alternative ways of spending a given amount of resources on a particular patient or patient group. In the latter context, a conventional QALY model seems more adequate.

There are a number of problems with the approach we have outlined above that need to be addressed in future research.

First, the societal preference basis for potential weights as suggested in our model is that people do not want to discriminate too strongly between programmes that provide the best health improvement possible for different groups of patients, even if some groups have 'higher ceilings' than others. This means that our model, when its use includes potential weights, pertains to comparisons of programmes for different patient groups that represent the best one can do for those patients. While this restricts the applicability of the model somewhat, we believe it will still be relevant for comparing a wide range of medical technologies and health programmes. Possibly 'second or third best programmes' could be valued by using some other kind of potential weights. Further reflection is needed to clarify in what form such weights would have to be.

Second, the concept of differences in potential has two somewhat different aspects. If two patients A and B are equally ill, and health care can improve functioning more in A than in B, then A has a greater potential for health than B. This is the way we have used the concept of potential in the present paper. We suggest that society wishes to strike a balance between producing as large functional improvements as possible and helping each individual realize his or her highest possible level of functioning. The purpose of potential weights is to encapsulate this balanced view with regard to interventions for non life-threatening conditions. However, differences in potential is also an issue in the valuation of life extending interventions for people with different permanent levels of functioning. The permanently disabled have, by definition, a lower potential for health than the healthy. If the logic of conventional QALY analysis is directly translated into policy, life years gained by disabled people will be regarded as less valuable than life years gained in healthy people. The introduction of potential weights as in Table 5 or—alternatively—health state values with strong upper end compression as in Table 2, would eliminate much of this devaluation of life extending programmes for the disabled. However, with life extending programmes, people may not want to strike a balance between health maximization and realization of individual potentials. Rather, we suspect they want the healthy and the disabled to be treated on completely equal terms on the grounds that people's interest in, and entitlement to, continued life is

largely independent of their health ([17,38]; cfr. also the public and political reactions to the first cost-per-QALY based priority list in the state of Oregon in the USA).

If this is correct, it calls for yet another refinement of the conventional QALY model in assessments of societal value. Fortunately, it is not difficult to achieve technically. A preference for non-discrimination in matters of life-saving may be encapsulated in QALY calculations simply by saying that for states of chronic illness or disability that are preferred to death, all saved life years count as one. Values for health states—as for instance in Table 2—will then not apply to life extending procedures for the chronically ill or the disabled. They will only be relevant in estimating the value of health improvements for people with non-fatal conditions—relative to each other and to the prevention of premature death.

Some may find it illogical that there should not be a devaluation of additional life years provided to people with illness when most people do seem to be willing to make a trade-off between length of life and quality of life when asked standard gamble and time trade-off questions. We do not see a problem with this, for two reasons. First, individuals' real willingness to trade-off own life expectancy for quality of life is probably not as widespread as suggested by preference studies in healthy people that are asked to imagine themselves in states of illness. When standard gamble or time trade-off questions are asked in real patients, people with moderate illness or disability tend to be unwilling to sacrifice life expectancy to become healthy [54,55]. Second, one cannot always expect consistency across decision contexts. The fact that an individual says that he personally would be willing to sacrifice years at the end of life or chances of survival in order to gain quality of life, does not necessarily mean that he supports a public policy that discriminates between the healthy and the disabled in the provision of reliable life extending procedures.

Altogether we believe that restricting the use of health state values less than 1 to health improving programmes is a simple and valid way to eliminate a discriminatory effect of the conventional QALY model that potential weights alone do not fully prevent.

Third, when we compared the outcomes in Figure 1, we focussed on two particular aspects of fairness: the concern for severity and the concern

for realization of potential. We noted aversion to inequality in end states as a third possible (and conceptually independent) aspect. There is to our knowledge little empirical evidence on the strength of aversion to inequalities in end states (with Dolan [35] as a notable exception) and we have, therefore, not addressed this potential factor in our modelling. This is, however, certainly an issue for further research.

Fourth, while the psychometric properties of person trade-off questions have been explored to some extent [28,32,33,52], much work remains to be done. It is clear that the reliability at the individual level is low. However, average responses in large groups of subjects are likely to have good reliability [45]. Further research in this area is underway in a European Burden of Disease project [53].

Fifth, according to our multiplicative model, a programme will have value only if the time trade-off value of the initial state is less than unity. As noted above, people who actually experience moderate illness or disability tend to be unwilling to sacrifice life expectancy to become healthy. In other words, their time trade-off utility is 1. Yet they may be experiencing disadvantage and loss of quality of life, and society may feel that also these people have some claim on resources. The model above would not capture this. To do so, utility measurements may have to be supplemented by other, more sensitive quality of life measurements. (Some may also argue that the possibility of cardinal measurements of utility is unclear, in spite of the impression created by techniques like the time trade-off and the standard gamble of yielding responses at a cardinal level (see, for instance, [55]). It is not inconceivable that a model of societal valuations of health improvements for use in cost-effectiveness analysis could be based on ordinal measurements of patients' health-related quality of life rather than on utility measurements. However, this is a much larger issue that will not be dealt with at this stage.)

Sixth, we have described a procedure for estimating equity weights one by one. In doing so, we are disregarding possible interactions between the various arguments in a societal value function. This is a potential weakness of the procedure. Ultimately it must be checked that a model in which equity weights are used in combination with each other and with utilities, assigns relative

values to different health programmes that accord with people's overall assessments of those programmes (the proof of the pudding lies in the eating). Most likely this kind of test will reveal a number of inaccuracies, biases and/or inconsistencies in the initial valuation model. Some of these may presumably be eliminated through an iterative process of testing and adjusting by which an initial set of parameter estimates is gradually improved.

Last, but not least, there is a problem at the psychological level with the use of numbers in valuing health outcomes. Quite a few people already have difficulties with the quantification of quality of life in QALY calculations [56,57]. Even more will become uneasy when told that a numerical valuation of an improvement in health may be obtained by multiplying a utility score by a severity weight and a potential weight. There is a danger that the whole operation will start to look exceedingly complicated, far-fetched and academic.

As noted earlier, a multifactorial valuation model that encapsulates societal concerns for severity and realization of potential can be translated into a corresponding set of health state values with strong upper end compression—as in Table 2. The latter way of presenting numbers for use in calculating equity weighted QALYs is more simple (one set of weights rather than three), but also less explicit with respect to the efficiency–equity trade-off. It is a matter for further research to examine which tool is more attractive to potential users.

## CONCLUSION

It is not the goal of health care to maximize total health-related utility gains. There is a preference for giving priority to those who are worst off, and a reluctance to discriminate strongly against patients who happen to have lesser potentials for health than others. We suggest that these concerns be accounted for in economic evaluations either by adding severity weights and potential weights to the conventional QALY model or by assigning values to health states such that they obey rules of convexity (decreasing marginal value of increases in utility) and strong upper end compression. We also suggest that health state values may be relevant primarily in calculations of QALY gains

from health improving programmes. QALY gains from life extending programmes may need to be based on the principle that, in comparisons across patient groups with chronic illness or disability, a life year gained should count as one and no less than one as long as the year is considered preferable to being dead by the persons concerned.

In sum these suggestions, which are all motivated by concerns for equity, represent very significant departures from value measurement in conventional cost-utility analysis, which focuses on efficiency only. It would, therefore, be reasonable and wise to give economic evaluation based on these suggestions a different name. The term 'cost-value analysis' has been proposed earlier for what we above referred to as the 'comprehensive approach' [37]. We think the name 'cost-value analysis' is suitable in general for numerical evaluation models that aim at capturing societal concerns for both efficiency and equity in health care.

## ACKNOWLEDGEMENTS

We are indebted to the Rockefeller Foundation, who made it possible for us to embark on this research as a team during a residency at the Foundation's Conference Center in Bellagio, Italy, in July 1997. We also thank Fundacion BBV for its financial support. We thank Marthe Gold for valuable contributions during the preparation of the paper and an anonymous referee for helpful comments. Alan Williams made a number of extremely constructive suggestions concerning the first submitted version of the paper and contributed significantly in improving its clarity. For this we are very grateful.

## REFERENCES

1. Weinstein, M.C. and Stason, W.B. Foundations of cost-effectiveness analysis for health analysis and medical practices. *New England Journal of Medicine* 1977; **296**: 716–721.
2. Williams, A. Welfare economics and health status measurement. In: van der Gaag, J. and Perlman, M. (eds.) *Health, economics and health economics*. Amsterdam: North-Holland, 1981.
3. Broome, J. Goodness, fairness and QALYs. In: Bell, J.M. and Mendus, S. (eds.) *Philosophy and medical welfare*. New York: Cambridge University Press, 1988.
4. Lockwood, M. Quality of life and resource allocation. In: Bell, J.M. and Mendus, S. (eds.) *Philosophy and medical welfare*. New York: Cambridge University Press, 1988.
5. Nord, E. The significance of contextual factors in valuing health states. *Health Policy* 1989; **13**: 189–198.
6. Menzel, P. *Medical costs, moral choices*. New Haven: Yale University Press, 1983.
7. Mooney, G. and Olsen, J.A. QALYs: where next? In: McGuire, A. et al. (eds.) *Providing health care: the economics of alternative systems of finance and delivery*. London: Oxford University Press, 1991.
8. Richardson, J. Cost-utility analysis: what should be measured? *Social Science and Medicine* 1994; **39**: 7–21.
9. Williams, A. Ethics and efficiency in the provision of health care. In: Bell, J.M. and Mendus, S. (eds.) *Philosophy and medical welfare*. New York: Cambridge University Press, 1988.
10. Culyer, A.J. The normative economics of health care finance and provision. *Oxford Review of Economic Policy* 1989; **5**: 34–58.
11. Broome, J. *Weighing goods*. Oxford: Blackwell, 1991.
12. Williams, A. *Economics, QALYs and medical ethics*. Discussion Paper No. 121. York: Centre for Health Economics, University of York, 1994.
13. Williams, A. QALYs and ethics: a health economist's perspective. *Social Science and Medicine* 1996; **43**: 1795–1804.
14. Wagstaff, A. QALYs and the equity–efficiency trade-off. *Journal of Health Economics* 1991; **10**: 21–41.
15. Dolan, P. *The allocation of benefits in health care: Does equity matter?* Paper for the First World Conference of the International Health Economics Association, Vancouver, May 1996.
16. Williams, A. Intergenerational equity: an exploration of the 'fair innings' argument. *Health Economics* 1997; **6**: 117–132.
17. Harris, J. *The value of life*. London: Routledge and Kegan Paul, 1985.
18. Mehrez, A. and Gafni, A. Quality-adjusted life years, utility theory and healthy year equivalents. *Medical Decision Making* 1989; **9**: 142–149.
19. Rawls, J. *A theory of justice*. Cambridge: Harvard University Press, 1971.
20. Callahan, D. Setting mental health priorities: problems and possibilities. *The Milbank Quarterly* 1994; **72**: 451–470.
21. Brock, D. Ethical issues in recipient selection for organ transplantation. In: Mathieu, D. (ed.) *Organ substitution technology: ethical, legal and public policy issues*. Boulder: Westview, 1988.
22. Kamm, F. The report of the US Task Force on Organ Transplantation: criticisms and alternatives. *Mount Sinai Journal of Medicine* 1989; **56**: 207–220.
23. Menzel, P. *Strong medicine*. New York: Oxford University Press, 1990.

24. Daniels, N. *Just health care*. Cambridge, MA: Harvard University Press, 1985.
25. Harris, J. QALYfying the value of life. *Journal of Medical Ethics* 1987; **13**: 117–123.
26. Williams, A. Response: QALYfying the value of life. *Journal of Medical Ethics* 1987; **13**: 123.
27. Patrick, D., Bush, J. and Chen, M. Methods for measuring levels of well-being for a health status index. *Health Services Research* 1973; **8**: 228–245.
28. Nord, E. The person trade-off approach to valuing health care programs. *Medical Decision Making* 1995; **15**: 201–208.
29. Rosser, R. and Kind, P. A scale of valuations of states of illness: is there a social consensus? *International Journal of Epidemiology* 1978; **7**: 347–358.
30. Nord, E., Richardson, J. and Macarounas-Kirchmann, K. 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* 1993; **9**: 463–478.
31. Nord, E. Health status index models for use in resource allocation decisions. A critical review in the light of observed preferences for social choice. *International Journal of Technology Assessment in Health Care* 1996; **12**: 31–44.
32. Ubel, P.A., Loewenstein, G., Scanlon, D. and Kamlet, M. Individual utilities are inconsistent with rationing choices. *Medical Decision Making* 1996; **16**: 108–116.
33. Pinto, J.L. Is the person trade-off a valid method for allocating health care resources. *Health Economics*, 1997; **6**: 71–81.
34. Richardson, J. *Critique and some recent contributions to the theory of cost utility analysis*. Working Paper No. 77/ 1997. Melbourne: Centre for Health Program Evaluation, 1997.
35. Dolan, P. The measurement of individual utility and social welfare. *Journal of Health Economics* 1998; **17**: 39–52.
36. Nord, E. Helsepolitikere ønsker ikke mest mulig helse per krone. (Health politicians do not wish to maximize health benefits.) *Journal of the Norwegian Medical Association* 1993; **113**: 1171–1173.
37. Nord, E. The trade-off between severity of illness and treatment effect in cost-value analysis of health care. *Health Policy* 1993; **24**: 227–238.
38. Nord, E. The relevance of health state after treatment in prioritising between patients. *Journal of Medical Ethics* 1993; **19**: 37–42.
39. Nord, E., Richardson, J., Street, A., Singer, P. and Kuhse, H. Maximising health benefits versus egalitarianism: an Australian survey of health issues. *Social Science and Medicine* 1995; **41**: 1429–1437.
40. Pinto JL, Perpinan JMA, Health state after treatment: a reason for discrimination? (Submitted to *Health Economics*.)
41. Dolan, P. and Cookson, R. *Measuring preferences over the distribution of health benefits*. Mimeo. University of York: Centre for Health Economics. 1998.
42. Nord, E. An alternative to QALYs: The Saved Young Life Equivalent (SAVE). *British Medical Journal* 1992; **305**: 875–877.
43. Nord, E. A table of values for cost-effectiveness analysis in health care. *Journal of the Norwegian Medical Association* 1996; **27**: 3246–3249 (text in Norwegian).
44. Sintonen, H. An approach to measuring and valuing health states. *Social Science and Medicine* 1981; **15c**: 55–65.
45. Murray, C. and Lopez, A. *The Global Burden of Disease*. New York: WHO/Harvard University Press, 1996.
46. Eddy, D. Oregon's methods: did cost-effectiveness analysis fail? *Journal of the American Medical Association* 1991; **266**: 2135–2141.
47. Nord, E. Unjustified use of the Quality of Well-being Scale in Oregon. *Health Policy* 1993; **24**: 45–53.
48. Pinto, J.L. *The Oregon experience: scales, numbers and their meaning*. Paper for the 14th Spanish Meeting of Health Economics. Santiago de Compostela, June 1994.
49. Hadorn, D. Setting health care priorities in Oregon. *Journal of the American Medical Association* 1991; **265**: 2218–2225.
50. Dolan, P. Aggregating health state valuations: A reply to Erik Nord's comment. *Journal of Health Services Research and Policy* 1997; **2**: 203–204.
51. Gold, M.R., Siegel, J.E., Russell, L.B. and Weinstein, M.C. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996.
52. Richardson, J. and Nord, E. The importance of perspective in the measurement of quality adjusted life years. *Medical Decision Making* 1997; **17**: 33–41.
53. Van der Maas, P.J. *Disability weights for diseases in Europe*. Project Program. Mimeo. Rotterdam: Erasmus University, 1997.
54. O'Leary, J.F., Fairclough, D.L., Jankowski, M.K. and Weeks, J.C. Comparison of time trade-off utilities and rating scale values of cancer patients and their relatives. *Medical Decision Making* 1995; **15**: 132–137.
55. Fowler, F.J., Cleary, P.D., Massagli, M.P., Weissman, J. and Epstein, A. The role of reluctance to give up life in the measurement of the value of health states. *Medical Decision Making* 1995; **15**: 195–200.
56. Mulkay, M., Ashmore, M. and Pinch, T. Measuring the quality of life: a sociological invention concerning the application of economics to health care. *Sociology* 1987; **21**: 541–564.
57. Klein, R. The role of health economics. *British Medical Journal* 1989; **299**: 275–276.