HEALTH STATUS INDEX MODELS FOR USE IN RESOURCE ALLOCATION DECISIONS

A Critical Review in the Light of Observed Preferences for Social Choice

Erik Nord
National Institute of Public Health, Norway

Abstract
In the last two decades a number of health status index models have been developed for assessing the value of health outcomes in terms of quality-adjusted life years. The models can be tested by comparing their implications with direct observations of how societies think resources should be distributed across patient groups. This paper reviews empirical evidence of this kind from various countries and summarizes the evidence in three rules of thumb for selecting values for health states. Nine different models are judged relative to these rules of thumb. Eight of the models underestimate the strength of social preferences for treating the severely ill before the less severely ill. The ninth has a strong bias against states associated with emotional distress. As a consequence, none of the models can be seen as sufficient stand-alone instruments for valuing health outcomes. Instead, the models may be seen as complementary and adjustable parts of a tool kit that should also include the rules of thumb suggested in this paper.

In the last two decades a number of models have been developed for determining the value of health states at a numerical (cardinal) level of measurement (11;35;36;38;40;41). The values provided by the models may be used as inputs in calculations of quality-adjusted life years (QALYs). One of the purposes of such calculations is to serve as aids in setting priorities and allocating resources.

Unfortunately, different health status index models yield different values for health states and hence different estimates of the value of health outcomes. Until recently, there were hardly any efforts to compare these different estimates and to decide which models are more valid than others (9;26). An important reason is that the models were viewed as tools for estimating health outcomes in terms of gained quality of life. Since there is no gold standard for measuring quality of life, there was no way of judging objectively which models estimated such gains better than others.

However, rather than interpreting health status index numbers as measurements of quality of life, one may view them as expressing trade-offs that people make between quality and quantity of life (14;16;25;44). The trade-offs may be in terms of either individual utility or social value. The former refers to individuals' preferences
between different health scenarios for themselves. The latter refers to society's preferences among health care programs for different categories of patients. Both these kinds of preferences may be observed directly, which means that the validity of the health status index numbers may be tested empirically (19). Information on health status index models, based on comparing their implications for trade-offs with directly observed preferences, is now starting to emerge (28;37).

From the ways in which most health status index models are constructed, one would expect them primarily to capture trade-offs in terms of individual utility. The numbers yielded by the models have nonetheless largely been marketed as also expressing trade-offs that society would make between competing health care programs (6;11;46). This paper addresses this de facto use of the models. Its purpose is to evaluate existing health status index models as aids in determining the social value of increasing or reducing treatment capacity in different areas of health care. It reviews studies in which the general public's preferences for resource allocation in health care have been measured directly. It then compares the implications of existing health status index models with these observed preferences.

ANALYTICAL FRAMEWORK

By the social value of a health outcome, I mean the degree to which society appreciates it. The degree of societal appreciation is a function of: (a) the utility that the individual concerned derives from the outcome; and (b) equity considerations (15). The degree of appreciation translates into society's strength of preference for the outcome. This can be expressed in terms of the sacrifices that society is prepared to make to obtain the outcome. The willingness to sacrifice can be expressed in terms of person trade-offs (19). Consider, for instance, three health improvements: A, B, and C. Assume that society is prepared to sacrifice obtaining A in 100 people to obtain B in 50 people or C in 25 people. These person trade-offs imply that C is valued twice as much as B and four times as much as A.

The above operational definition of social value implies a straightforward way of testing models that purport to express social value. The test is as follows: if a model assigns the value V to outcome X, and a value N \times V to outcome Y, then it predicts society to be indifferent (more or less) between a project resulting in outcome Y for 100 people and a project resulting in outcome X for N \times 100 people. The more society deviates from this view when asked directly, the less valid is the model. This is referred to as the person trade-off test (19;21).

A number of factors determine society's appreciation of a health care program. In QALY calculations, health status index models purport to capture two of these factors, namely the severity of the initial state and the size of the intervention effect. As a preamble to the judgment of existing models, I shall review empirical evidence from various countries regarding social concerns for these two factors.

SOCIAL CONCERNS FOR SEVERITY

In 1986 a group of health care politicians, health administrators, health care personnel, and representatives of patients was commissioned by the Norwegian government to set out guidelines for setting priorities in the Norwegian National Health Service (30). One of the main conclusions of the commission was that severity of illness should continue to be the most important criterion for prioritizing between patients, although this criterion should be considered together with the effectiveness.
of treatment. Later, similar positions were adopted by government-appointed commissions in several other countries (4;7;31;39).

Studies of population preferences support these official government positions. In a study among 150 Norwegian politicians involved in health care decision making at the county level (18), the subjects were presented with the following problem:

Imagine an illness A that gives severe health problems and an illness B that gives moderate problems. Treatment will help patients with illness A little, while it will help patients with illness B considerably. The cost of treatment is the same in both cases. There is insufficient treatment capacity for both illnesses, and an increase in funding is suggested. Three different views are then conceivable:
1. Most of the increase should be allocated to treatments for illness B, since the effects of these are greater.
2. Most of the increase should be allocated to treatments for illness A, since these patients are more severely ill.
3. The increase should be divided evenly between the two groups.
Which of these views comes closest to your own?

Forty-five percent chose the third view (divide evenly), while 38% chose the second one (main emphasis on severity). Only 11% placed the main emphasis on the size of the treatment effect (option 1). Interestingly, this fits quite well with a view put forward recently by a leading American bioethicist (2):

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.

Nord (25) asked subjects in various convenience samples and small population samples to compare programs for curing severely ill and less severely ill patients in terms of person trade-offs. The preference for treating the severely ill was very strong. To understand this, consider the following states:

- State A: Unable to work, unable to pursue family and leisure activities, strong pain, depressed;
- State B: Unable to work, moderate pain; and
- State C: Moderate pain.

Typically, restoring one person in state A to full health was considered equally valuable as curing 50 people in state B and 100 people in state C. Similar preferences were later observed in a series of workshops with health politicians, planners, professionals, and patients in different counties in Norway (20;24), as well as in personal interviews with members of the above-mentioned Priority Commission (Nord, unpublished data).

In England, Rosser and Kind (36) asked a convenience sample of patients, doctors, nurses, and health care workers to compare various states of illness with a light reference state (“mild distress, no physical disability”). For each state X, the subjects were asked to indicate by a numerical factor “how much more ill a patient in state X is than a patient in the reference state.” They were asked to indicate such numbers with the additional information that the numbers would be interpreted in terms of person trade-offs and as having implications for resource allocation. The results
Table 1. Responses to the Question, "How Much More Ill Are You in State X Than in State (1, 2)?" (Rounded Medians) (36)

<table>
<thead>
<tr>
<th>Disability</th>
<th>None</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Slight social</td>
<td>2</td>
<td>3</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Severe social/slight work</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>Severe work</td>
<td>7</td>
<td>9</td>
<td>12</td>
<td>26</td>
</tr>
<tr>
<td>Unable to work</td>
<td>11</td>
<td>13</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>Confined to wheelchair</td>
<td>25</td>
<td>31</td>
<td>64</td>
<td>200</td>
</tr>
<tr>
<td>Confined to bed</td>
<td>65</td>
<td>87</td>
<td>200</td>
<td>500</td>
</tr>
</tbody>
</table>

(medians) are shown in Table 1. The numbers may be interpreted as the relative values that the subjects would place on cures for patients in the different states. The values increase very rapidly with increasing severity.

The social concern for saving lives was addressed in a joint Norwegian–Australian study (28). Subjects were asked how they, thinking of themselves as members of Parliament, would evaluate two equally expensive proposed special units A and B. Unit A would save 10 people per year from dying and restore them to full health. Unit B would restore to full health a number of people in the following state:

Sitting in a wheelchair, pain most of the time, unable to work.

The question put to the subjects was: how many patients must be treated in unit B per year in order that you would find it just as valuable to spend the money on unit B as on unit A? The median responses in Norway and Australia were 50 and 40, respectively. In another group of subjects, the condition treated in unit B was described instead as follows:

Uses crutches for walking, light pain intermittently, unable to work.

The median responses were then 110 and 85. These results highlight the great value that the societies in question assign to life-saving as opposed to health-improving interventions.

An exercise similar to the one above was given to members of the Norwegian Priority Commission. In this case, saving lives was compared with curing people in the following condition:

Severely reduced mobility; can sit, but need help to move about, both at home and outdoors.

The person trade-offs expressed by the members of the Committee varied from 1 to 1,000. Nine out of 11 chose a number between 10 and 100. The median was 50 (versus 10 lives saved). This fits well with the results of the Norwegian/Australian study.

In Rosser and Kind's study, subjects were asked to add a number for being dead after having provided numbers for all the states in Table 1. The median number for being dead was 200. Again, this may be interpreted as the relative value placed on saving lives. The value is three times as high as the value placed on curing people confined to a wheelchair in a state of moderate distress. It is 30 times as high as the
value of curing people whose "performance at work is very severely limited" due to their illness.

Saving lives was also valued highly in a preference study in a group of Spanish doctors and students (33). For instance, a person trade-off of 1:10 was found between a program that would save a life and a program that would cure people in the following state:

Some problems with walking about and washing or dressing, unable to perform usual activities, moderate pain or discomfort, moderately anxious or depressed.

Results similar to that of the Spanish study were obtained in a recent study in the United States (Dr. Peter Ubel, University of Pittsburgh, personal communication).

**RULES OF THUMB CONCERNING SEVERITY**

The above evidence is scattered and heterogeneous. Nevertheless, it suggests the order of magnitude by which severity of illness seems to be emphasized by people in a number of industrialized countries. To picture this order of magnitude, consider four classes of outcomes, exemplified as follows:

• Class A: Saving the life of a healthy person.
• Class B: Curing a person with a severe problem, for example, a person who sits in a wheelchair, has pain most of the time, and is unable to work.
• Class C: Curing a person with a considerable problem, for example, a person who uses crutches for walking, has light pain intermittently, and is unable to work.
• Class D: Curing a person with a moderate problem, for example, a person who has difficulties in moving about outdoors and has slight discomfort, but is able to do some work and has only minor difficulties at home.

In countries like Norway, England, Spain, and Australia, the social appreciation of class A outcomes seems to be approximately 3–6 times as high as that of class B outcomes, 10–15 times as high as that of class C outcomes, and 50–200 times as high as that of class D outcomes. These numbers pertain to valuations of outcomes in decisions about future treatment capacity. Forthcoming data suggest that the United States does not seem to deviate significantly from this valuation pattern. Quantitative models that purport to be useful for estimating the social value of health care activities in these countries, as well as in other countries with similar values, must reflect this structure of concern.

To do this, health status index models must assign values in the following order of magnitude:

• Severe problem (outcome class B): 0.65–0.85
• Considerable problem (outcome class C): 0.50–0.94
• Moderate problem (outcome class D): 0.98–0.995

For instance, if a health state scores 0.98–0.995, then the implication is that the value of curing a person in that state is 1/50–1/200 of the value of saving a person from dying to a life as healthy (0.02–0.005 versus 1.0). This required value structure is referred to as one that compresses health states to the upper end of the scale (27).
CONCERNS FOR TREATMENT EFFECT

The previous section shows a strong social preference for treating the severely ill before the less severely ill. But what if two groups suffer equally and more can be done for one than the other? To what extent does the size of the health improvements achieved in different patient groups affect their appreciation by society?

In health economics theory, it is a basic assumption that the social value of an intervention is proportional to the size of the health improvement (43;47). This is called a utilitarian view. The Norwegian Priority Committee (30), on the other hand, argued that all patients have an equal right to realize their potential for health, whether small or large. One interpretation of this claim is that as long as the patient's potential for improvement is realized, the size of the improvement is less important, i.e., the degree of social appreciation of a moderate improvement and a large improvement should be the same. Ethicists in England, Sweden, the United States and New Zealand have argued similarly (2;3;5;10;39). This is considered an egalitarian view. Results from public preference measurements vary on this issue.

Patrick et al. (32) asked graduate students and health leaders in New York to value in person trade-off terms saving the lives of people in different states of illness in relation to saving the lives of healthy people. Priority was clearly given to the latter category. For instance, saving the life of a healthy person was considered approximately equivalent to saving the lives of three people sitting in a wheelchair unable to work. One interpretation of this result is that saving life was valued more highly “the more health that was saved,” i.e., a utilitarian view.

Later events suggest that these results may not be generally valid in the United States today. One of the main reasons for the rejection by the Bush Administration of the Oregon prioritization plan in 1992 was that it would discriminate against the permanently disabled and chronically ill, in whom the potential for health tends to be lower than in other people.

Data in Norway tend to support the egalitarian rather than the utilitarian view. In a small-scale study using the person trade-off technique, Nord (23) found that subjects were in general not willing to give priority to treating people with a greater potential for benefiting if this meant treating fewer people. The egalitarian view was particularly strong in people with an education below the college level. The subjects rationalized their responses by pointing to equality between individuals in value of life and entitlement to treatment, irrespective of differences in potential for health.

In the study mentioned earlier among 150 Norwegian politicians involved in health care decision making at the county level (18), the subjects were asked the following:

Imagine two illnesses A and B. They are equally common and cause the same degree of suffering. Treatment will help patients with illness A a little, while it will help patients with illness B considerably. The cost of treatment is the same in both cases. There is insufficient treatment capacity for both illnesses, and an increase in funding is suggested. Two different views are then conceivable:

1. Most of the increase should be allocated to treatments for illness B, since the effects of these are greater.

2. The increase should be divided evenly between the two groups, on the grounds that they suffer equally and are equally entitled to treatment.

Which of these views comes closest to your own?
Seventy-two chose the second view (egalitarian), while 24% chose the first one (utilitarian). The preference for the egalitarian view was particularly strong among women, older people, those with less than a college-level education, and members of political parties to the left.

In the workshops mentioned earlier with health politicians, planners, professionals, and patients in different counties in Norway (2), subjects were asked to compare various pairs of projects, where in each pair the patients' initial function levels were the same while the level resulting from treatment was higher in one project than in the other. The subjects were placed in the role of health planners; that is, they were asked to distribute resources between groups of other people. Given this role, they generally expressed some degree of willingness to give priority to those who benefit more. However, it was hypothesized that the subjects might take a different view if they were to choose between different sets of rules for prioritizing that could have consequences for themselves as potential future patients. They were therefore also presented with the following problem (21):

Assuming that you do not know what kind of illnesses you yourself might get in the future, which of two hospitals, A and B, would you rather belong to—one that gives equal priority to patients with equal initial severity of illness as long as the treatment effect is substantial in either case (hospital A), or one that gives priority to those with a greater potential for improvement (hospital B)?

The question placed the subjects behind a Rawlsian veil of ignorance (34) and, therefore, in a sense elicited unbiased preferences. Of 51 subjects, 31 preferred to belong to hospital A (equal priority), while 20 preferred to belong to hospital B (priority to better outcome). This means that a majority of those who gave priority to the better outcome option in the role of health planners did not themselves wish to belong to a hospital that followed such a practice. In ensuing discussions, it appeared that most people believed that their valuation of receiving treatment would primarily be determined by how ill they were and not as much by how much they could be helped as long as the help would be substantial.

A similar pattern of responses was later observed in personal interviews with members of the Norwegian Priority Commission (Nord, unpublished data).

In a survey in Australia (29), subjects were given a choice between the following options:

1. Among patients who are suffering equally, some priority should be given to those who will be helped most from treatment.
2. Among patients who are suffering equally, those who can become a little better should have the same priority as those who can become much better.

Fifty-three percent chose option 1, while 47% chose option 2 ($n = 551$). The preference for option 2 was stronger among women than men. There was a slight tendency for the former to find their choice more difficult. It should also be noted that option 1 used the expression "some priority," which in itself is an expression of weak preference. On the other hand, option 2 compared those who can become only "a little better" with those who can become "much better." This description of a quite large difference in outcome did not make option 2 an obvious choice. Nonetheless, 47% chose option 2. Altogether, the observed distribution suggests that
the size of the potential health improvement is of moderate importance in Australians' valuations of interventions for different patient groups.

In summary, the above evidence suggests that, in valuing interventions for different patient groups, the degree to which people emphasize treatment effect varies. It does seem clear, however, that in a country like Norway, the social valuation of health outcomes is much less tightly related to their size than is suggested in conventional health economics theory. People tend to feel that equally ill people have (or should have) equal rights to treatment. This view seems to be shared by a considerable portion of the Australian population, and has been expressed strongly in bioethical debate in England, New Zealand, Sweden, and the United States.

RULES OF THUMB CONCERNING SEVERITY AND TREATMENT EFFECT

The rules of thumb formulated earlier concerning severity encapsulate to some degree egalitarian views regarding treatment effect. Consider four planned services, A, B, C, and D, for patients in life-threatening conditions. A will restore the patients to full health, while B, C, and D will leave the patients with moderate, considerable, and severe problems, respectively. With health state values as in the rules of thumb, there will be small differences between the valuations of services A, B, and C (1.0, 0.98–0.995, and 0.90–0.94, respectively), and even the outcomes of service D will score 0.65–0.85 of the outcomes of service A. Similarly, the value of helping patients with severe or considerable problems will vary only moderately with the degree to which they are helped.

In the following, therefore, the value structure suggested in the rules of thumb should be regarded as encapsulating not only social concerns for severity, but also, to a reasonable degree, concerns for equity between patient groups with different potentials for benefiting from treatment. Quantitative models that purport to be useful for estimating the social value of health care activities must, at least roughly, reflect this structure of concern. In the following, nine models will be briefly reviewed in the light of this requirement.

THE MODELS

Main features of the models are given in Table 2. All nine models have a multidimensional, descriptive system by which any state of illness can be categorized. The number of dimensions ranges from 2 to 15. With some models, different dimensions have been valued separately, and additive or multiplicative formulas are applied for calculating values for composite states. With other models, composite states have been valued directly and are presented in tables. Valuation techniques vary and include standard gamble, time trade-off, rating scales, and magnitude estimation.

As noted previously, most of the models were constructed primarily to capture trade-offs in terms of individual utility. The numbers yielded by the models are nonetheless largely being marketed also as expressing trade-offs that society would make between competing health care programs. The following test addresses this de facto use of the models.

TESTING OF THE MODELS

Each of the states described in the rules of thumb above (severe, considerable, and moderate problems) was scored using each of the nine models. Some of the scores were taken from the literature (references below). Others were obtained ad hoc by
<table>
<thead>
<tr>
<th>Model (reference)</th>
<th>Dimensions</th>
<th>Valuation technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Quality of Well-being Scale (12)</td>
<td>Mobility, social act, physical act, 25 symptoms</td>
<td>Rating scale on single dimensions; additive formula</td>
</tr>
<tr>
<td>The Health Utilities Index Mark I (41;42)</td>
<td>Physical, role, social-emotional, health problem</td>
<td>Rating scale on single dimensions, time-trade off on composite states; multiplicative formula</td>
</tr>
<tr>
<td>The Health Utilities Index Mark II (42)</td>
<td>Sensory, mobility, emotion, cognitive, self care, pain, fertility</td>
<td>Rating scale on single dimensions, rating scale and standard gamble on composite states; multiplicative formula</td>
</tr>
<tr>
<td>The EuroQol Instrument (1;40)</td>
<td>Mobility, self care, usual activity, pain, anxiety/depression</td>
<td>Rating scale on composite states</td>
</tr>
<tr>
<td>The York EuroQol Time Trade-off Tariff (45)</td>
<td>Same</td>
<td>Time trade-off on composite states</td>
</tr>
<tr>
<td>The Index of Health-related Quality of Life (simple) (35)</td>
<td>Disability, physical discomfort, emotional distress</td>
<td>Standard gamble on composite states</td>
</tr>
<tr>
<td>The Index of Health-related Quality of Life (complex) (35)</td>
<td>Dependency, disharmony, dysfunction, pain/discomfort, symptoms, dysphoria, fulfillment</td>
<td>Multistep rating scale procedure; additive formula</td>
</tr>
<tr>
<td>The 15 D (38)</td>
<td>Mobility, vision, hearing, breathing, sleeping, eating, speech, bladder and bowel function, usual activities, mental, discomfort, depression, distress, vitality, sexual activity</td>
<td>Multistep rating scale procedure; additive formula</td>
</tr>
<tr>
<td>The Rosser/Kind Index (13;36)</td>
<td>Disability, pain/distress</td>
<td>Magnitude estimation on composite states</td>
</tr>
</tbody>
</table>
Table 3. Health State Scores According to Rules of Thumb and Different Health Status Index Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Severe</th>
<th>Considerable</th>
<th>Moderate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rules of thumb</td>
<td>0.65-0.85</td>
<td>0.50-0.94</td>
<td>0.98-0.995</td>
</tr>
<tr>
<td>QWB&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.45-0.55</td>
<td>0.65-0.70</td>
<td>&lt;0.80</td>
</tr>
<tr>
<td>HUIM1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.10-0.20</td>
<td>0.30-0.40</td>
<td>&lt;0.85</td>
</tr>
<tr>
<td>HUIM2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.40</td>
<td>0.70</td>
<td>0.90-0.94</td>
</tr>
<tr>
<td>EuroQol</td>
<td>0.20&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.60&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.70&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>York EuroQol (TTO)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.20-0.25</td>
<td>0.40-0.50</td>
<td>0.80</td>
</tr>
<tr>
<td>IHQL (3D)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.90-0.75</td>
<td>0.80-0.90</td>
<td>0.90-0.94</td>
</tr>
<tr>
<td>IHQL (complex)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.75&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.86&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.91-0.93&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>15 D</td>
<td>0.68&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.94&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.97-0.98&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Values taken from Nord et al. (28).
<sup>b</sup> Ad hoc judgments (Methods section).
<sup>c</sup> Values taken from Sintonen (37).

having two independent judges (the author and a medical doctor) apply the various instruments to the states in question (scores in the literature are based on a similar approach). In a few cases where there was substantial disagreement between the two judges, we discussed the cases and reached a closer agreement. Results are presented as intervals or in terms of order of magnitude to reflect the uncertainty involved in the judgment process.

RESULTS

The results are summarized in Table 3. They suggest that eight out of nine existing health status index models do not have sufficient upper-end compression to capture the strength of social preferences for treating the severely ill before the less severely ill. In five models, the bias is so strong that the models are completely unusable, at least as stand-alone aids, in decisions concerning treatment capacity across patient groups (the QWB, HUIM1, MUIM2, and EuroQol raw scores). Two models (IHQL-complex, 15 D) assign adequate values to curing people with severe problems relative to life-saving interventions, but overestimate society's appreciation of cures for moderate conditions relative to cures for severe ones.

DISCUSSION

The preference measurements reviewed in this paper are scattered and heterogeneous and do not allow any precise estimates of the trade-offs that modern western societies wish to make between health benefits for different patient groups. However, the measurements do allow us to judge whether numbers produced by existing health status index models are at all plausible as expressions of social value. The answer is that they mostly are not.

The insufficient upper end compression in the QWB, HUIM1, and EuroQol has been shown in a previous study (28). The bias in the QWB has also been demonstrated in practice: many of the counterintuitive rankings of interventions produced by this model in the first attempt of the Oregon Health Commission to set priorities within Medicaid were clearly a result of the instrument's inappropriate value structure (8,27).

40  INTL. J. OF TECHNOLOGY ASSESSMENT IN HEALTH CARE 12:1, 1996
The 15D performs relatively well. However, it should be noted that this instrument has a "floor problem." Because of its additive structure, states that are quite unbearable may score in the medium range of the 1-0 scale. For instance, as long as there are no problems with vision, hearing, breathing, eating, speech, bladder and bowel functions, and mental function, a state will score above 0.40, irrespective of the degree of immobility, pain, and depression. This does not seem reasonable.

The Rosser/Kind Index is the only model that seems to perform satisfactorily on a person trade-off test over a wide range of illness severity. The reason for this good performance is, of course, that the index values derive from responses to questions that focused on person trade-off implications.

None of the other models are based on preference measurements that included this resource allocation perspective. They all used valuation techniques that aim at capturing the individual utility associated with different health states (standard gamble, time trade-off, category rating, visual analog scale). The issue of equity, which we must assume is essential in explaining social concerns for the most severely ill, is simply not addressed by these techniques.

From this one might infer that, while all six models have serious shortcomings as instruments for estimating social value, some of them might be useful for quantifying the individual utility of health outcomes. This issue lies beyond the scope of the present paper. However, the concept of individual utility is in itself quite problematic. In particular, it is not obviously meaningful to compare in finite, numerical terms the satisfaction that individuals get from life-saving procedures with the satisfaction they get from improvements in health (22).

Returning to the issue of social evaluation, it may be tempting to draw the conclusion that the Rosser/Kind Index should be recommended as the only usable model for analysts and decision makers who wish to assess numerically the social value of different health outcomes and, hence, the value of increasing or reducing treatment capacity for different diagnostic groups. However, it is not so simple.

First, the index has a crude descriptive system. There are only two dimensions (disability and distress), and the labels for the different levels of the two dimensions are not very informative. Example states mentioned in the original publication shed additional light on only a few of these levels. This means that in practical applications some rough judgment will often be necessary to decide where in the value matrix different states of ill health should be located. This amounts to saying that estimates based on the Rosser/Kind Index will probably be associated with considerable random error.

Second, the index does seem to have an important bias—not with respect to the weight attached to severity of illness, but to the disutility attached to disability relative to distress. The latter includes anxiety and depression, and these conditions by definition imply a loss of quality of life. The same is not true of disability: a person may be confined to a wheelchair and still be happy. In the light of this, it is counterintuitive that the index assigns the same value to "no disability, moderate distress" as to "slight disability, no distress." Similarly, it is surprising that "no disability, severe distress" scores higher than "severe work disability, no distress." As noted elsewhere (19), the explanation may be that the valuations were elicited by asking subjects, "How much more ill you are in state X than in state Y?" The subjects may have regarded the concept of illness as more relevant to disability than to distress.

Interestingly, the bias against psychological distress is less present in the three-dimensional IHQI, which, although related to the Rosser/Kind Index, used a different valuation technique (standard gamble). The IHQI-3D, for instance, assigns
Nord

a lower score to “severe emotional distress” than it does to “confined to bed” (0.79 versus 0.84). But it also assigns a score of 0.64 to “extremely depressed and actively suicidal.” For a state that by definition cannot be much better than being dead, this is much too high.

CONCLUSIONS

Given the shortcomings of the Rosser/Kind Index, none of the models can be seen as sufficient stand-alone instruments for valuing health outcomes. Instead, the models may be seen as complementary and adjustable parts of a tool kit that should also include the rules of thumb suggested in this paper (severe problems = 0.65–0.85; considerable problems = 0.90–0.94; moderate problems = 0.983–0.995). In some comparisons of competing projects or budget proposals, single models may be chosen from this tool kit and applied directly as they stand. For example, the IHQL and the 15 D may be valid in comparisons of life-saving programs versus programs for people with severe chronic illnesses, and the Rosser/Kind Index seems reasonably valid for comparisons of cures for people with different degrees of physical disability. In other situations, one may have to choose a model that has a suitable descriptive system and a plausible weighting of different health problems and then adjust the health state values so that they satisfy the rules of thumb regarding concerns for severity and treatment effect. For instance, consider a choice between increasing treatment capacity for people with moderate depression versus increasing treatment capacity for people with limitations in moving about and working. The Rosser/Kind Index is not a good valuation model here because of its bias against distress-type illnesses. The IHQL and the 15 D do not have this problem (at least not to the same extent), but they do have insufficient upper-end compression in their value structure. Analysts interested in evaluating the two competing programs may then use the rules of thumb pertaining to classes C and D outcomes to adjust the value structures of the IHQL and/or the 15 D and calculate the value of the two programs on the basis of either of these adjusted models.

These conclusions pertain to the short term. In the longer term, ideally some of the present models will be revised to fit better with prevailing social values. Work is also in progress on alternative models (17; 24; 20).

REFERENCES