VALUING HEALTH: A NEW PROPOSAL

DANIEL M. HAUSMAN*

Department of Philosophy, University of Wisconsin, Madison, WI, USA

SUMMARY
After criticizing existing systems of health measurement for their unargued commitment to evaluating health states in terms of preferences or well-being, this essay argues that public rather than private values of health states should help guide the allocation of health-related resources. Private evaluation of health states is relative to a prior individual choice of specific activities and goals, while public evaluation is relative to the whole range of important activities and goals. Public evaluation is concerned with securing a wide range of choices as well as with success given one’s choice. A reasonable simplification from the public perspective is to focus on just two features of health states: the subjective feelings attached to health states and the limitations that health states imply on the range of important activities that individuals can pursue. Focusing on just these two dimensions permits the construction of a parsimonious classification of health states with regard to what matters most from the public perspective. This classification, which resembles those in the HALex and the Rosser and Kind Disability and Distress Index, might best be built on top of existing health-state classifications, by mapping the health states they define to activity-limitation/feeling pairs.

To assign values to these pairs, I propose relying on deliberative groups to make comparisons among the pairs with respect to the relation ‘is a more serious limitation on the range of objectives and good lives available to members of the population’. A ranking according to this property, is not a preference ranking, because it is not a ranking in terms of everything that matters to individuals.

Working back from the weights attached to the activity-limitation/feeling pairs, one can impute weights for the health states in other classification systems that were mapped to those pairs. If those weights coincide roughly with current weights, then one legitimates current weights and provides a vehicle for their public discussion and possible revision. If those weights do not coincide, then one has both an argument for revising current views of the cost effectiveness of treatments and policies and a method to carry out such a revision. Copyright © 2009 John Wiley & Sons, Ltd.

Received 8 May 2007; Revised 6 November 2008; Accepted 23 January 2009

KEY WORDS: QALYs; health measurement; values; preferences

1. INTRODUCTION
This paper sketches a new proposal for health measurement for the purpose of guiding allocations of health-related resources. Such a proposal might appear to be timed badly. Enormous resources have already been devoted to devising systems of generic health measurement, several of which are already in use.1 One might reasonably conclude that efforts should go toward consolidating this work and comparing the results of measurements employing existing indexes rather than toward developing another way to measure health. Nevertheless, I shall sketch a new proposal – though one that

*Correspondence to: Department of Philosophy, University of Wisconsin, Madison, WI 53706, USA. E-mail: dhausman@wisc.edu

1Among the most important are the SF36, the EQ5D, the Health Utilities Index, the Quality of Well-Being Scale, the Health and Activity Limitation Scale, and the World Health Organization’s techniques for estimating Disability Adjusted Life Years.
piggybacks on existing work – that aims to make possible major improvements conceptually, ethically, and empirically.

The second of this essay’s seven sections clarifies the purpose of health measurement with which the essay is concerned and sketches how health has been measured for that purpose. Section 3 criticizes preference-based methods of measuring health. Section 4 articulates constraints on the principles that should govern policy in liberal democracies and sketches abstractly how to carry out a public evaluation of health states. Section 5 explores the classification of health states suggested in Section 4, while Section 6 sketches a method for quantifying what matters about health. Section 7 concludes.

2. HOW HEALTH HAS BEEN MEASURED

Generic health measurements serve at least three purposes. First, they can help to evaluate treatment options, whose effects on a patient’s health are rarely limited to their efficacy in coping with a particular ailment. A measure such as the SF36 provides detailed information about different aspects of health, which is useful in the context of treatment. A second purpose of generic health measurements is to identify populations with a distinctive health status in order to help isolate the causes. Measures such as the SF36 may be useful for this purpose as well.

The third purpose for measuring health, the one I am concerned with, is to help guide the allocation of health-related resources by providing cost-effectiveness information. I assume that cost-effectiveness information is relevant to decisions concerning health policy, not that health policy should be entirely determined by it, nor that health policy should aim to maximize population health. Policies that would maximize health may be unjust and for that reason unacceptable. I assume that it is nevertheless important to compare the costs and health consequences of alternative policies. When alternatives are equally fair, the more cost-effective policy provides greater health improvements, and when the alternatives are not equally fair, cost-effectiveness information tells policy makers how much health they must sacrifice in order to be fair.

To provide quantitative information to help guide health policy, a health measure must be at least intervally significant. Since death (or full health) is a natural zero, a ratio scale seems even more appropriate. In addition, health measures must be interpersonally comparable, or else measures of total health or differences in population health would be arbitrary. Apart from the SF36, which in its original form did not aim to provide a scalar measure, all the other major systems of health measurement specify values for death and ‘full health’, with \( \mu(\text{death}) = 0 \) and \( \mu(\text{full health}) = 1 \) or, in the case of the World Health Organization’s disability adjusted life years, full health has the disability weight of 0, and death is valued at 1. ‘Full health’ is some level of good, but not necessarily perfect health. The health measure defended in this essay can be scaled on a 0–1 scale.

Furthermore, in common with most health measurement schemes, I shall suppose that people’s health can be described by the sequence of their momentary ‘health states’ along with the amount of time that they spend in these health states. Diseases, syndromes, injuries, or illnesses map on to temporally extended sequences of health states, not on to health states themselves. Health states are individuated symptomatically by locating them along multiple dimensions of bodily or mental functioning. For example, in the EQ5D, the five dimensions are mobility, self-care, usual activity, pain/discomfort, and anxiety/depression. Along each dimension, individuals either have no limitation, some

---

2As originally designed, the SF36 did not provide a scalar measure of health, which is not needed to assess how well treatment is going. Recently, Brazier et al. (2002) have shown how to construct a scalar measure from multidimensional SF36 scores.

3Otherwise measures of differences between health states that health policies cause will be arbitrary. An intervally significant health measure must be unique up to a positive affine transformation – that is, if \( \mu(H) \) and \( \mu'(H) \) are two satisfactory measures of the health state \( H \), then \( \mu'(H) = a\mu(H) + b \), where \( a \) and \( b \) are real numbers and \( a \) is greater than zero. Once one specifies the value of any two health states, the values of all other health states are fixed.

limitation, or a severe limitation. So there are 3^5 (243) health states or 244 including death. The Health Utilities Index (HUI) (mark 3) has eight dimensions: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain, and it distinguishes five or six levels along each dimension, for a total of 972,000 distinguishable health states. The separate dimensions, even in the case of the HUI, are typically amalgams of different aspects of health. The proposal defended here depends on a coarse classification of health states in terms of their consequences in particular environments. Finer-grained classifications of health states in terms of the more specific functional limitations they involve are needed for other purposes, which this proposal will not serve.

To guide resource allocation, a health measurement system must then assign numbers – quality or disability ‘weights’ – to the health states in its classification. In common with other systems of generic health measurement, the units of health resulting from this new proposal will be something like ‘health-adjusted life years’, ‘quality-adjusted life-years’, or ‘disability-adjusted life-years’. In assuming that the impact of diminished health is the product of amount of time an individual is in a diminished health state and a measure of the health state, one assumes that the same number can be assigned to a health state no matter how long it lasts and no matter what other health states precede or succeed it. This assumption is defensible only as a reasonable simplification.

How can numbers be assigned to health states? The approach that measurement theory suggests would be to define a two-place empirically detectable relation \( R \) (‘is at least as healthy as’), which is complete, reflexive, and transitive and which can thus be represented by a function \( \mu \) assigning real numbers to health states such that \( \mu(H_1) \geq \mu(H_2) \) if and only if someone in \( H_1 \) is at least as healthy as someone in \( H_2 \). If the relation \( R \) is complete, reflexive, and transitive, then the function \( \mu \) is order preserving – it is unique up to a positive monotone transformation. \( R \) needs to satisfy further conditions in order that it has a representation that is intervally significant. In reality, however, \( R \) is incomplete and fails even to define an ordering. When a health state \( H_1 \) is above another health state \( H_2 \) along some of the dimensions in a health-classification system and below \( H_2 \) along other dimensions, there seems to be no basis to compare health states by the relation \( R \) ‘is at least as healthy as’.

The conclusion most of those concerned with health measurement draw, in which I concur, is that the task is to value health states rather than to measure health itself. A further reason to shift from measuring health itself to measuring the value of health is that what matters to health policy is not health as conceived of by medical theory (see, for example, Boorse, 1977) but the value of health.

In judging how good or bad health states are, both individuals and policy makers appropriately take into consideration a great deal besides the health states themselves. The physical, technological, and social environment matters, too. For example, although paraplegia is the same state of health, regardless of the environment, it is worse when transportation and public buildings are not accessible to wheelchairs than when accommodations have been made (Allotey et al., 2003). Second, the significance of health states depends on individual activities and interests. An elbow injury that would be merely annoying for most people might be devastating for a tennis player. However, as argued below, the public value cannot be sensitive to all the details that are relevant to individual evaluation.

A quantitative measure of the value of health states presupposes that there is among health states an observationally detectable relation \( G \) (‘is at least as good as’) that is complete, reflexive, and transitive. To define an interval or ratio scale, further conditions must be met. Most of those concerned with health state evaluation interpret ‘Health state \( H_1 \) is at least as good as \( H_2 \)’ as meaning either (i) that people weakly prefer \( H_1 \) to \( H_2 \), or (ii) that people are at least as well-off with respect to their health in \( H_1 \) as in \( H_2 \). Preferences are crucial for both interpretations, since those who take the value of a health state to be its consequences for well-being usually take preferences as indicating well-being. For important recent exception, see Dolan and Kahneman (2007).

Consider for example, the wide diversity of different deficiencies in vision or ‘emotion’.

For an important recent exception, see Dolan and Kahneman (2007).
When the evaluation of health states is conceived of as the measurement of preferences, the task can be portrayed as in Figure 1. In Figure 1 the preferences of individuals among health states are the starting point for social evaluation, but they could not possibly be the starting point for individual evaluation. To have preferences among health states, individuals must have already judged their value. In evaluating health states by measuring preferences, health economists derive the social evaluation of health states from individual evaluations of health states. One might thus call the schema in Figure 1 ‘the social choice model’ of health-state evaluation. As in social choice theory, social values depend on individual values.

The two boxes on the left-hand side of Figure 1 and the first two arrows are rarely explicitly modeled. In the social choice model, the starting points for the measurement of health are taken to be preferences, whose values must be elicited and then aggregated. However, in deriving social values from individual values, one can distinguish at least three different possible starting points:

1. **Judgments of personal value**: how good or bad an individual judges health states to be.
2. **Judgments of well-being**: judgments of how health states bear on well-being.
3. **Individual preferences** among health states.

The social choice model either conflates these three, or, more charitably, takes preferences as indicating both judgments concerning well-being and personal value. However, as argued below in Section 3, judgments of personal value, judgments of well-being, and preferences will often diverge.

If $G$ (‘at least as good as’) is interpreted as preference, then those who are concerned to measure health can help themselves to cardinal representation theorems proven by economists and decision theorists. If $H_1$ is full health and $H_2$ is death and one normalizes by setting $V(H_1) = 1$ and $V(H_2) = 0$, then the value of a lottery that results in full health with probability $p$ and death with probability $1-p$, which is called a ‘standard gamble’, is $p$. If (i) individual preferences among health states satisfy all the axioms and (ii) the value of health states is indicated by preferences among health states, and (iii) it is correct to assign the same values to the full health and death of different individuals, then the so-called standard gamble is guaranteed to measure correctly the value of health states. Other technical results partly justify health economists in inferring health state values from trade-offs between health state quality and time (Buckingham and Devlin, 2006). The assignment of the same value for all individuals to states of full health and, respectively, death can be interpreted as an evaluative commitment to weighting everyone’s health equally.

---

In doing so, health economists might regard preferences as evidence of individual evaluations, and those judgments as evidence concerning the value of health states. Social evaluation could then be conceived of as a process of weighing the (fallible) evidence provided by individual preferences. Economists more commonly regard the derivation of a social evaluation from individual preferences instead as a process of aggregating individual values or compromising among them. I am indebted to Sarah Marchand here.
In fact, people’s preferences among health states do not satisfy all the axioms, and neither the standard gamble nor time trade offs correctly value health states. For example, people often refuse to incur any risk of death or any curtailment of their life span in order to be rid of mild health deficits. None of the methods for quantifying health are perfect, and the choice among methods of preference measurement is accordingly empirical and pragmatic (see for example Parkin and Devlin, 2006). Because there are too many health states in health classifications such as the HUI or even the EQ5D for respondents to value them all directly, health economists measure preferences among only a small number of alternatives. They then estimate a multi-attribute utility function from which the values of the remaining health states can be calculated. There are well-known further difficulties concerning whom to ask and how much information to provide to respondents (see Gold et al., 1996). Precision cannot be expected.

3. WHY THE VALUE OF HEALTH STATES SHOULD NOT BE AN INDEX OF PREFERENCES

Health states can be valued in many ways. They need not be evaluated by their bearing on well-being or in terms of preferences. Insofar, as there is any argument for relying on preferences to evaluate health states – and little has been given – it rests on two main assumptions, neither of which is defensible:

(1) People’s preferences among health states coincide with their judgments concerning the value of health states.
(2) People’s judgments of the value of health states are at least as accurate as any alternative.

Those who would measure preferences because they hold that health states should be evaluated in terms of their bearing on well-being must defend two other assumptions: that preferences accurately indicate well-being and that health states should be evaluated by their bearing on well-being.

An individual’s preference ranking of health states takes into account everything that would be relevant to a real or hypothetical choice among health states. This is the standard interpretation of ‘preference’ in economics and decision theory (Hausman and McPherson, 2006, Chapter 4). Preferences among health states will thus be influenced by many things. Among these will be judgments concerning how health states bear on well-being and judgments concerning how good or bad health states are. However, preferences, judgments of personal value, and judgments of well-being will not coincide. How health states affect well-being is only one factor among many influencing the value and choice-worthiness of health states. My judgment whether \( H_1 \) is better for me than \( H_2 \) will not coincide with my beliefs about whether \( H_1 \) is a state of better health than \( H_2 \), and so my preferences cannot match both. For example, although a worse health state, a minor disability might be better for me, because it disqualifies me from military service. Judgments concerning the personal value of health states and their bearing on well-being may diverge from preferences both because of the divergence between personal value and well-being and because preferences depend on other things. For example, I may be and judge myself to be both in better health and better off over all when taking medication controlling bipolar disease, but I might believe that without the manic states the disease brings I will be unable to

---

7The technical problems with preference elicitation mechanisms are not the basis for my rejection of preference-based measurement. The mechanism Section 6 suggests for assigning quantitative values will face analogous difficulties. My reasons for rejecting preference-based measurement are conceptual. They are sketched in the next section and developed in Hausman (2006).

8One might also argue that democratic norms imply that health-related resources ought to be distributed in accordance with the preferences of members of the relevant population. However, democratic norms say only how decisions should be made, not what determines the value of health states. This democratic sovereignty argument should be distinguished from the anti-paternalist view that each individual should be the judge of his or her own good. According to the defender of popular sovereignty, whether social policy should be paternalistic should be up to the public. It is questionable whether respect for popular sovereignty demands deference to popular views on the value of health states. In one survey, respondents themselves rejected reliance on preferences elicited from the public (Richardson and Olsen, 2005).
accomplish some project that matters to me more than my health or welfare. For data showing the influence on preferences of factors that are not relevant to the value of health states, see Baker and Robinson (2004). People’s preferences will not coincide with their judgments of how good or bad health states are nor with their judgments concerning how health affects well-being.

The second assumption that people’s judgments (as expressed by their preferences) are reasonably accurate is also dubious. As is well known, preferences are distorted by false beliefs and other cognitive and affective factors. Moreover, there are specific reasons to question the accuracy of people’s evaluations of health states, as these are reflected in their preferences. Evaluating health states is difficult. If health professionals cannot figure out how to do this, why suppose that individual respondents, who know less and have had less time to reflect on the task have succeeded? In contrast to typical applications of consumer choice theory, in which the consumer’s evaluation of commodities derives from repeated choices, with opportunities for learning and correcting errors, the preferences expressed in health measurement surveys concern alternatives about which individuals know comparatively little. For example, what do most people know about the career possibilities, the social life, or the main recreations of those who are blind, and without this information, how can people construct their preferences among health states? Systematic differences in individual evaluations, especially concerning the impact of disabilities, apparently derive from differences in the information and experience of respondents and provide strong evidence that preferences among health states are often uninformed or distorted (Ubel et al., 2003). Since individual evaluations of health states are not well informed, preferences would be a poor basis for a social evaluation, even if they matched individual evaluations.

Furthermore, even if people’s preferences among health states matched their judgments of the value of health states, and those judgments were well-informed, undistorted by cognitive limitations, and the result of careful reflection, preferences would still only be a guide to the personal value of health states, not to what I call ‘the public value of health states’ – that is, the valuation of health states that should guide resource allocation. To illustrate the contrast between public and personal value, consider the well-known fact that many of the disabled claim that their quality of life is comparable to the quality of life of individuals without disabilities (Ubel et al., 2003). For example, some of the deaf maintain that they would prefer not to have their hearing restored (Lane, 2002). Since people can often construct excellent lives for themselves even if their disabilities preclude many valuable human experiences, one must take seriously the claim that some disabilities do not limit how well off people can be.

If personal values determine public values, then the more successfully people cope with disabilities such as paraplegia or blindness, the less serious those disabilities are and the smaller the benefits of measures that cure blindness or paraplegia. This implication is hard to accept. One reason not to evaluate health (for the purposes of guiding resource allocation) by measuring preferences is that one can then recognize that some of the disabled may reasonably prefer not to be rid of their disabilities without thereby denying that they are in fact disabled.

This section has argued that are three reasons not to measure the public value of health states by surveying members of the population.

1. Preferences among health states do not coincide with judgments of the personal value of health states.
2. Judgments of the personal value of health states are unreliable, because they require knowledge, experience, and time for reflection that individuals do not have.

---

\footnote{Surveys, such as Brickman et al. (1978), Schulz and Decker (1985), and Wu (2001), that establish the power of adaptation also show that on average those with disabilities rate their health and quality of life as somewhat worse. Nevertheless, the significance of a disability from a public policy perspective cannot be identified with its consequences for the well-being of those who are disabled.}
3. The public value of health states is not an aggregation of personal values. Even if the first two problems could be solved, the value of health states to individuals, which is influenced by both the limitations and inconveniences associated with health states and the extent to which people nevertheless build rich human lives, is not the same as the value of health states that is relevant to social policy.

4. PUBLIC EVALUATION

The gap between personal and public evaluation is not peculiar to the assessment of health. Scanlon (1975) has pointed to a general distinction between what individuals care about and the factors that should influence social policy. He argues that the weighing of benefits and costs to individuals that is relevant to questions of justice depends on their ‘urgency’ rather than on their subjective importance. He gives the following memorable example, ‘The fact that someone would be willing to forego a decent diet in order to build a monument to his god does not mean that his claim on others for aid in his project has the same strength as a claim for aid in obtaining enough to eat’ (1975, pp. 659–660). Why?

Scanlon’s answer emphasizes two points: first, judgments concerning the justice of social policies turn on how they bear on central human interests and, second, subjective benefit is in part a matter of individual responsibility. The first point does not help with Scanlon’s example, since spiritual pursuits are central human interests, while the second point, that individuals are responsible for their preferences, risks collapsing into the observation that strength of preference need not reflect the ethical importance of a claim, which is what Scanlon is attempting to explain.

Scanlon’s discussion is nevertheless fruitful. The central implication I draw from it is that public values are not an aggregation of personal values, because personal values depend on individual choices of how to live and what objectives to pursue. State action in contrast (at least from a liberal political perspective) should be neutral among competing grounds for individual choices of how to live. The central responsibility of government is to create an environment that secures fundamental interests, including the fundamental interest in being able to pursue one’s personal interests. The state’s role is to secure justice, to expand opportunities, and to insure individuals against catastrophes. Social policy should aim to enlarge the possibilities for the successful pursuit of individual goals, rather than to achieve those goals for individuals. The choice of objectives and their pursuit should then be up to the individual.

This division of labor does not assume that the public evaluation is more accurate than personal evaluations. It does not assume that the food is more valuable than the temple. The differences between public and personal evaluation lie in the questions they address rather than in the accuracy of the answers. Individuals appraise actions and states of affairs, including health states, from the perspective of their own specific objectives. From such a perspective a temple may be more valuable than basic nutrition. The state, on the other hand, evaluates policies without committing itself to any personal objectives. Its goal is to expand and secure the range of alternatives that are accessible to individuals. Basic nutrition, unlike a temple, helps make possible a wide range of different pursuits. In Sen’s terminology (1992), though both capabilities – the set of different functionings available to an individual – and successful functioning matter, an individual’s main, though not exclusive, concern is with successful functioning, while the state’s main concern is with capabilities. Government serves as a referee, a protector, a facilitator, and an insurer, not as a big brother. Even if a majority of the population preferred to devote their resources to building the temple, a liberal government should not undertake that project. These are not exactly Scanlon’s conclusions, but they are in the spirit of his analysis.
The goal of social policy is not, however, to expand the range of activities that are available to individuals without reference to the importance of the activities. Providing individuals with a wider choice of religions is more important than providing them with a wider selection of breakfast cereals. Some activities are more important and important to more people than others, and the state cannot be neutral about which these are. Prioritizing which alternatives government policy should be particularly concerned to expand depends on social consensus and philosophical argument. Within single societies, even societies as diverse as the United States, there is considerable agreement concerning what things matter to individuals most deeply and widely. There is less agreement across the boundaries of societies, but there is still considerable overlap between the lists of important goods that members of different societies will draw up. Consider, for example, Griffin’s list of the five most important constituents of a good life (1986, p. 67): (a) achievement; (b) the components of human existence: agency, autonomy, basic capacities, and liberty; (c) understanding; (d) enjoyment; and (e) deep personal relations. People will weight the entries on this list differently, and there is bound to be controversy concerning its contents, especially across cultures. However, there will be considerable agreement too. For example, intimate personal relations matter enormously to almost everyone.

The constituents of a meaningful life, such as close companionship, are not in general things that government should provide. Individuals must provide them for themselves. The role of government is to provide individuals with the means and opportunities to pursue for themselves as wide a range as possible of these goods.

Returning now specifically to health, while from a personal perspective what matters most about health states is their bearing on how successfully an individual functions in the way of life the individual has chosen, the criterion of evaluation from a public perspective should be how severely health states limit the range of alternative lives and pursuits that are open to individuals. From a personal perspective, the evaluation of a health state depends on its intrinsically rewarding or repellent features and on its bearing on the specific goals of the individual. Bad health diminishes subjective well-being or prevents individuals from achieving their ends. From a public perspective, on the other hand, the evaluation of a health state should depend on the extent to which its characteristics and consequences diminish the range of good lives and valuable projects that are available to people. From a public perspective, the significance of bad health lies not in ultimate outcomes, but in the extent to which it diminishes capabilities.

In other words, health states should be evaluated in terms of how severely they limit the range of valuable lives individuals can live. Rather than asking respondents ‘Do you prefer H₁ to H₂?’ or ‘Are people better off in H₁ or in H₂? or simply ‘Is H₁ better than H₂?’ the question to ask is ‘Does H₁ constrain the possibilities of living well and pursuing valuable objectives more than H₂ does?’ The evaluative relation ‘G’ in terms of which health states should be compared and ultimately quantified is not preference, consequences for well-being, nor some unspecified ‘better than’ relation. It is rather something like ‘capability enhancement’ or, put negatively, ‘capability constriction’.

Ranking and then quantifying health states in terms of capability enhancement rather than preference or well-being does not necessitate abandoning existing health-classification systems. In principle it is possible to compare the extent to which health states, as defined in some familiar health-classification scheme, limit the range of valuable lives individuals can live. However, such comparisons are very difficult. One reason why is that they require copious information. Evaluators need to know how health states influence what people can do and they feel. For example, before evaluators can judge how severely blindness limits what kinds of lives people can live, they need to know what activities

---

10 The discussion here shows the influence of Daniel’s views (1985, 2007).
11 Though it seems much easier just to express a preference, this appearance is deceptive. To construct preferences that are worth taking seriously, one needs to consider everything relevant to the evaluation of health states, including their consequences for the range of lives people can live. Hence, responsibly formulating a preference is in fact more difficult, not less.
blindness interferes with. Some of this information is obvious, but much of it is not. Filling in the relevant information will of course neither solve the problems of evaluation nor make them easy to solve. However, it will help.

Hence, in addition to proposing a new criterion to evaluate health states, I also suggest a new health-state classification in terms of just two dimensions: activity limitations and health-related feelings – the two aspects of health that most directly bear on the question of how health limits capabilities. In making specific activities impossible or difficult, health states can rule out certain kinds of lives and projects. In addition, suffering and distress, both physical and mental, interfere directly with achieving many good lives. The features of health states that are relevant from a public perspective can be summarized by an activity-limitation/feeling (a/f) pair. For policy purposes, health-state evaluation can focus on just two dimensions along which health states can be located: activity limits and feelings.

Although it would be possible to construct questionnaires to determine directly activity limitations and feelings, what I have in mind instead is mapping health states, as defined in a more complicated health-stated classification system such as the EQ5D or the HUI or the SF36, into the equivalence classes defined by the simpler classification in terms of feelings and activity limitations. This task is largely non-evaluative. The a/f pair that typically corresponds to which array of functional limitations is a complicated matter of fact. In this way, this proposal reduces the informational demands on those evaluating health states in terms of the extent to which they limit how people can live.

Assigning an overall scalar capability enhancement or constriction value to health states remains an enormously difficult task. The fact that there are only two dimensions lessens the complexity and diminishes the cognitive demands on those attempting to assign weights to a/f pairs. Evaluation remains challenging. Apart from the special context of purely instrumental evaluation with respect to some well-specified goal, neither philosophers nor social scientists have any satisfactory theory of either how people in fact manage to evaluate alternatives or how they (rationally) ought to evaluate alternatives. In this context, the only solution to the problem of evaluation is to rely on the abilities people have to rank alternatives, however poorly understood those abilities may be, and to hope that a well-designed deliberative process will improve the results. This is not a return to a preference-based ranking, because the problem evaluators face is not to figure out their preferences – their all-things-considered ranking of health states. Their problem is different and much more narrowly defined. Their task is to rank and ultimately to assign numbers to a/f pairs, where these numbers represent how severely a/f pairs limit how people can live and what significant projects they can pursue.

Having separated off the factual questions and having ascertained what experts have to say about their answers, evaluative disagreements are bound to remain. People can invoke well-known ideals governing what kinds of lives and activities are most valuable to argue that some a/f pairs make available a wider range of better lives for more people. It would however be overly optimistic to suppose that consensus will result. What then? I have already rejected aggregation. If one rejects the dubious possibility of relying on some sort of evaluative expertise, the only way left to cope with disagreement is procedural, which in the case of roughly democratic societies involves deliberation coupled with some mechanism for decision making while deliberation proceeds. Health-state evaluation should, I suggest,

12It is important to distinguish the classification of the activity limitations health states cause from the evaluation of a/f pairs in terms of the extent to which they limit the range of lives and projects open to individuals. The classification, unlike the evaluation, is primarily a factual matter.

13To reiterate the conclusions of Sections 2 and 3: having no idea how to value health states, health economists have supposed that members of the target population have somehow managed the task and that their preferences indicate the personal values they assign to health states. Health economists then regard public values as an aggregation of personal values. Since I am skeptical of the ability of individuals to adjudicate among all the considerations that bear on the value of health states and doubtful that preferences indicate personal values, I question whether the results of preference elicitation techniques inform health economists concerning the personal values of health states. Since, in addition, I deny that the public value of health states should be an aggregate of personal values (let alone preferences), I need to describe some alternative way of determining the public value of health states, which does not rely on the ‘black box’ of individual preferences.
depend on deliberative groups coupled with public debate, rather than on surveys of individual judgments. I will return to these issues in Section 6.

To summarize this section, reflection on the goals of public policy in a liberal state suggests that the significance of health with respect to public policy lies in its consequences for what kinds of lives people can have and what kinds of projects they can pursue. These goals suggest a two-part proposal for health-state evaluation. First health professionals should map health states as individuated in a more detailed taxonomy, such as the SF36, the HUI, or the EQ5D, to a small set of activity/feeling (a/f) pairs. Such a classification conveys the information that is most relevant to judging how poor health diminishes capabilities. The second step is to assign numbers to a/f pairs on a 0–1 scale by (a) asking deliberative groups to consider how seriously a/f pairs limit how people can live and what goals they can pursue and then (b) subjecting the findings of such groups to public debate. The new capability-diminishing criterion of evaluation can be applied to health states as individuated by existing health-state classification systems and so does not require the first step, but adopting the first step reduces the cognitive demands on evaluators.

5. IMPLEMENTING AN A/F CLASSIFICATION

Although this programmatic paper is not the place to lay out the details of a still-to-be-constructed method of health-state evaluation – even if I were capable of doing so – it is important to put some flesh on the philosophical skeleton described in Section 4.

To clarify this proposal, it helps to compare it to two existing systems of health-state classification and evaluation. The first is the HALex – the Health and Activity Limitation Scale, which was constructed in the 1990s by Pfeniffer Erickson, Ronald Wilson, and Ilda Shannon at National Center for Health Statistics (Erickson et al., 1995; Erickson, 1998). The HALex is an attempt to make use of existing data from the National Health Interview Survey to provide information concerning the level of generic health in the United States and in some subpopulations. The National Health Interview Survey was not designed to locate people’s health within any health-state classification system. Hence, the task was to work backwards from the information in the survey to some classification of health states, to which quality weights could then be assigned.

As it happens, the National Health Interview Survey includes questions concerning activity limits and it also asks individuals for a subjective appraisal of their health (‘Would you say your health in general is excellent, very good, good, fair, or poor?’). In constructing a health classification, Erickson and her co-workers mapped answers to questions about specific activity limitations to six general levels of activity limitation whose abbreviated descriptions are as follows (Erickson et al., 1995, Table II, p. 11): 15

1. Not limited.
2. Limited in other activities.
3. Limited in major activity.
4. Unable to perform major activity.
5. Unable to perform instrumental activities of daily living (IADL).
6. Unable to perform activities of daily living (ADL).

14The mapping depends on how one treats remediation. For example, myopia limits activities more seriously if people do not have corrective eyeglasses than if they have them. If the remediation will be present regardless of policy choices, then the relevant health states should include the existing remediation, but if the remediation depends on the policy choices, then the relevant health states are unremediated.

15The full descriptions are more complicated and classify the health of children and the aged. In this paper I do not address the problems of classifying the health states of the young and very old.
ADL include eating, getting in/out of bed, inside mobility, dressing, bathing, and toileting. IADL (in societies such as the United States) consist of light housework, laundry, meal preparation, grocery shopping, outside mobility, travel, money management, and telephoning (Kindig, 1997, p. 56). The list of IADL may vary across societies and over time. Combining these six levels of activity limitations with the five possible subjective reports of health results in thirty possible health classifications (or 31 including death), as shown in Figure 2.

Full health, which is here operationalized as no activity limitations and a self-report of excellent, is assigned the value of 1, and death is assigned the value of 0. The details concerning the assignment of values to the cells are not germane here. The point of this discussion is neither to defend nor to criticize the HALex as a health measure. Given the purposes for which it was constructed and the constraints its authors faced, it is very sensible.

What is of interest here are three conceptual features of the HALex, which help to clarify and to articulate my proposal to classify health states in terms of activity limitations and feelings. Notice first that the columns in Figure 2 consist of five levels of subjectively perceived health. Which column an individual's health is located in accordingly depends on the individual's assessment of his or her health. The columns differ with respect to judgment, not health-related feeling. Someone who leaves the doctor's office with a diagnosis of cancer might feel fine physically but judge that he or she is in poor health. Conversely, someone might be in serious pain, having just passed a kidney stone, yet judge that he or she is in very good health. Because people's subjective assessment of their health depends on their expectations concerning future health states, subjective health assessments are not a dimension along which (momentary) health states should be classified. The columns in the HALex should be distinguished from the levels of health-related feeling that, I suggest, should influence the public value of health states.

Before commenting on two other important features of the HALex, let me turn to a second health evaluation scheme, whose classification resembles the proposal in this paper more closely. This is the Rosser and Kind Disability and Distress Index, shown in Figure 3.

What is relevant here is the classification of health states, not the weights, which were elicited using psychometric techniques explained in Rosser and Kind (1978). The disability levels in the rows of Rosser–Kind Index resemble the activity levels in the HALex, but the columns distinguish levels of distress (feeling) rather than subjective health judgments.

There is an ambiguity in the characterization of the rows (which is the second point I wanted to make about the HALex), which may explain in part why the Disability and Distress Index was superseded by health-classification systems such as the EQ5D. Do the rows distinguish among the consequences of functional limitations that are relevant from a particularly evaluative perspective – which is what I propose – or do they distinguish among levels of functional limitation? Most of the levels of disability (like the activity levels in the HALex) seem to differentiate consequences of functional limitations rather than the limitations themselves. On the other hand, the difference between being unconscious and being conscious and confined to bed is a functional difference. It seems that Rosser conceived of the Disability and Distress

---

16This remark is not intended as a criticism of Erickson and her co-workers. They had to work with the questions on the National Health Interview Survey.
Index as classifying functional limitations rather than their consequences (1988, p. 134). In a co-authored paper, she describes disability and distress as ‘two primary dimensions of morbidity’, and she writes that ‘further work is required before the set of constructs necessary for a comprehensive, reliable and valid classification of morbidity applicable to all age groups is complete’ (Rosser and Watts, 1978, p. 534).

Classifying functional limitations by their consequences makes it easier for respondents to understand the impact of functional limitations and to evaluate them. However, the consequences of functional limitations are selective and less descriptive of health itself and for that reason less useful with respect to other (non-allocative) purposes for which generic health measures are sought. The ambiguities concerning what is intended by a health-classification system reflect the multiplicity of purposes that researchers have hoped that health classifications could serve. A health-classification system is a compact description of people’s momentary health states. What information it should provide depends on what purposes the information will serve. It would be nice to have a classification and an evaluation of health states that provide the information sought by those who wish to make clinical judgments, by those who seek to determine the causes of ill health, and by those who need to make decisions concerning the allocation of health-related resources. Nevertheless, I am arguing that the third purpose is better served by a classification and evaluation of health states designed for that purpose alone.

The final noteworthy feature of both the HALex and the Disability and Distress Index is the way that the ‘levels’ of ‘activity’ or ‘disability’ are defined. The specification of these levels should not be confused with the evaluation of health states, however that evaluation is to be done. Though lower activity levels are supposed to be worse, the classification of health states in terms of their consequences for activities is largely factual rather than evaluative. Whether a health state constitutes a limitation in a ‘major activity’ (in the HALex) is a factual matter, such as whether someone is able to work, rather than an evaluative question concerning how bad the limitation is.

In addition, in both the HALex and the Disability and Distress Index, the classification of activity levels depends on the individual’s past choices concerning how to live and what goals and occupation to choose. Hence, for example, the loss of a finger would be a limitation in the major activity of a violinist. Yet the loss of a finger does not of course greatly constrict the range of activities in which people can participate, and it does not rule out or make difficult many common or important activities. On a choice-insensitive interpretation of the scale of activity limitation, the loss of a finger would not count as causing a major activity limitation.

Although a classification of activity limitations that presupposes an individual’s prior occupational choices is sensible in the context of measuring the personal value of health states, it is inappropriate if the goal is to provide a classification that facilitates evaluation of health states in terms of capability. In the relevant choice-insensitive interpretation of levels of activity limitation, the loss of a finger does not cause a significant activity limitation, even though it prevents violinists from pursuing their profession. Conversely, paraplegia does significantly limit activities, even if all paraplegics have

<table>
<thead>
<tr>
<th>Disability Levels</th>
<th>Distress Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Slight social</td>
<td>Mild</td>
</tr>
<tr>
<td>Severe social, slight work</td>
<td>Moderate</td>
</tr>
<tr>
<td>Severe work</td>
<td>Severe</td>
</tr>
<tr>
<td>Unable to work</td>
<td>Unable</td>
</tr>
<tr>
<td>Chair-bound</td>
<td>Chair-bound</td>
</tr>
<tr>
<td>Bed-bound</td>
<td>Bed-bound</td>
</tr>
<tr>
<td>Unconscious</td>
<td>Unconscious</td>
</tr>
</tbody>
</table>

Figure 3. Rosser–Kind Disability–Distress Index.

These are Alan Williams’ abbreviations (1988, p. 285) for the more detailed descriptions Rosser provides (1988, p. 137).
chosen occupations and ways of living with which paraplegia does not interfere. For the purposes of assisting the public evaluation of health states, activity limitations must be classified without reference to the previous choices of individuals concerning how to live and what to try to accomplish.

I thus suggest that the classification of health states for the purpose of quantifying their public value should be something like that shown in Figure 4.

Figure 4 is intended only to illustrate how the proposal defended here might be articulated, and I make no defense of its details. As in Figure 2, ‘ADL’ stands for ‘activities of daily living’, and IADL stands for ‘instrumental activities of daily living’. One way to classify activity limitations is to distinguish between ‘core’ and ‘peripheral’ IADL and to distinguish whether health states limit people in most or only a few of these. ‘Core IADL’ might consist of a subset of IADLs, such as light housework, meal preparation, outside mobility, money management, and telephoning (Kindig, 1997, p. 56). Peripheral IADL might consist of the remaining items on the list of IADLs (travel, laundry, grocery shopping) supplemented with instrumental activities such as reading, using a computer, driving, and other shopping. Each row is defined to be a more severe limitation than the row above. Hence, for example, the third row should be defined more precisely as ‘limitations in peripheral IADL without limitations in core IADL or in ADL’. Designing a useful categorization of activity limitations and feelings requires research, and there are many other ways to proceed than what is suggested in Figure 4. There is nothing sacrosanct about the number of rows or columns. Although a classification of activity limitations typically takes for granted the array of activities present within a society, it must be sensitive to the possibility that there are normatively unacceptable prior constraints on activities. For example, in a castes society significant cognitive deficiencies in individuals belonging to lower caste (who are the majority) might limit few of the activities available to them, because that range of activities is already severely limited even among those in full health. In such unjust conditions the relevant question concerns how health states would limit activities if they were not already unjustly limited.

The columns in Figure 4 are supposed to distinguish different kinds of health-related feelings. Two people who are in the same health state as measured by the SF36 or the HUI should have the same health-related feeling, even if one has just won the lottery while the other’s house has just burned down. There seems to be no good way to measure health-related feelings directly, because (apart from specific aches and pains) people cannot easily distinguish what part of their feelings is health related. However, it should nevertheless be possible to assign levels of health-related feelings to health states by averaging overall mood or happiness levels reported by individuals in those health states.

I do not know how people’s health will be distributed among the 20 a/f pairs defined in Figure 4 and consequently how well such a system discriminates among health states. Presumably there will be fewer actual health states in the cells at the lower left-hand and upper right-hand corners than in cells near the upper left-hand corner. As a preliminary inquiry, it would not be difficult to sketch a rough mapping from the 243 states in the EQ5D to the 20 pairs in Figure 4 and then to use data concerning how populations are distributed among the EQ5D’s states to get some idea concerning the distribution of a/f pairs in the population.

The classification illustrated in Figure 4 (like Nord’s ‘scale of severity of illness (1999, p. 119) is of use only to help guide the allocation of health-related resources, and even for this purpose, the
information it provides is very thin. However, if (as suggested in Section 2), the evaluation of health states is inevitably imprecise, such a barebones classification may suffice. What one loses in detail one may more than gain in simplicity. Classifying health states this way for this purpose does not preclude classifying them in other ways for other purposes, and indeed it seems sensible to piggyback this classification of health states on top of a more refined classification.

6. VALUING ACTIVITY-FEELING (A/F) PAIRS

I argued in Section 4 that the goal of public policy (in a liberal state) is to create a space in which members of the population can enjoy a variety of good lives and pursue a wide range of significant personal objectives. This overall goal suggests that, within the constraints of fairness, society should use health-related resources as efficiently as possible to enhance capabilities – that is to expand the range of valuable lives available to people. To compare the benefits of alternative policies requires a quantitative (intervally significant) scalar measure of the public value of a/f pairs – that is, of the extent to which they enhance or diminish capabilities.

As argued in Section 4, the best way to measure these values is to use deliberative groups and public debate. In order to get numbers that mean what they are supposed to mean, that are reasonably stable and consistent over time, and that are sensitive to relevant considerations, the deliberative groups and the task they will tackle need to be carefully delineated. What follows is an illustrative sketch, not a concrete proposal, which would depend on sociological and psychometric research and experimentation.

The task is to assign numbers to a/f pairs – that is, to the cells in Figure 4 (although, as mentioned in Section 4, nothing in the criterion of evaluation requires this classification, as opposed to existing health-state classifications). Full health (the top left cell) has the value 1 and death has the value 0. The numbers assigned to health states are supposed to quantify how health affects capabilities. These numbers are not supposed to represent personal or group preferences, which could be quite different. The values should be elicited from the workings of deliberative groups and subjected to public debate rather than from surveying individuals, because assigning these weights demands a more varied experience than any one individual is likely to have had as well as arduous reflection, which individuals cannot be expected to carry out on their own. In addition, the give and take of discussion will help correct obvious errors. Since the numbers that deliberative groups come up with are designed to guide the allocation of health-related resources in some specific population, the groups assigning the weights should be representative of that population. However, the fact that a deliberate group is representative of a population does not imply that the weights it assigns are correct or should not be contested. The conclusions drawn by deliberative groups should be inputs into a public debate.

To assign weights to a/f pairs, members of the deliberative group will need to understand the classification thoroughly. Employing this understanding and thinking carefully about the goal of the ranking – to determine how a liberal state ought to prioritize health deficits (subject to the constraints of fairness) – individuals will have a basis upon which rank the a/f pairs. Constraints are also built into the classification system: each cell should be ranked at least as highly as any cells below it or to its right.

The problem of assuring the interval significance of the weights remains, but there is no reason why familiar methods for eliciting these weights could not be adapted for use within deliberative groups, despite their known drawbacks. In addition, there is room for direct argument concerning the relative magnitudes of differences between the cells. For example, people can deliberate about whether health policy should be more concerned to move individuals who feel okay from the fifth row to the fourth

---

18There are other ways of posing the questions that might make them easier to tackle. For example, members of the deliberative group could be asked which pair they would prefer from behind a ‘veil of ignorance’ (Rawls, 1971) that prevents them from knowing their central objectives. Such a veil of ignorance would serve to disconnect this ranking from the personal considerations that influence ordinary preference rankings.
than from the second row to the first or whether, among those in the third row, reducing suffering to discomfort expands people’s possibilities more than helping those who feel merely okay to feel vigorous.

Those experienced with standard survey methods may be scoffing loudly (unless they quit reading long ago). There are three reasons why they may scoff. First, they might question whether deliberative groups could ever succeed in assigning weights. However, this skepticism is misdirected. If health economists can assign numbers to health states by means of individual surveys, they can do so by means of deliberative groups. One flawed way to do so would be to provide members of deliberative groups with redesigned surveys resembling those currently used to assign weights to health states, to let group members discuss the questions, to ask the members to respond separately, and finally to average the results.

Instead of doubting whether it is possible to get quantitative weights from deliberative groups, skeptics might instead question whether one can get sensible or accurate numbers. It might be argued that the task is too poorly defined and too demanding to expect a representative group of individuals to provide anything sensible. Though such qualms are justified, note that current methods of assigning weights to health states are even more demanding, since they do not help individuals to identify the aspects of health states that are most relevant and do not permit the information sharing and time for reflection that a deliberative group would provide.

Finally, skeptics might question whether the public could accept as legitimate the quantitative weights that deliberative groups assign to a/f pairs. Even if more thoughtful and better informed than survey results, the weights might appear arbitrary, and the method of determining them might appear undemocratic. This is potentially a serious difficulty, but it is a problem of misapprehension, rather than an objection in principle. The best way to address it is to encourage public debate over the values deliberative groups assign to a/f pairs.

Having tentatively determined some set of weights, the policy consequences of those weights would then provide an additional line of argument that would be available to the public. Because there are only 20 cells in the table, these numbers and their significance can be the subject of public debate.19 The bearing of consequences is complicated, however, because health policy is influenced by concerns about fairness in addition to improving overall health. One cannot expect that public debate will always productive here (or elsewhere), but since the evaluation of health states is a public normative task, it is hard to see how to do better.

Once weights are assigned to the a/f pairs, it will be possible to employ the mapping between a/f pairs and health states such as those defined by the EQ5D to assign weights to the latter. If it turns out that the capability enhancement weights are similar to the preference-based weights currently in use, then this excursion will not change any cost-effectiveness conclusions. Nevertheless, the excursion will still be valuable. It would legitimize the current procedures and show that their apparently unjustifiable features do not skew the results. In addition, it would take the mystery out of the numbers and make possible systematic public scrutiny. Because ranking health states by aggregating preferences is so different from ranking health states by their consequences for capabilities, one would expect that the weights derived in these ways will not coincide and that the proposal sketched in this paper will imply revisions in current views of the cost effectiveness of treatments and policies.

Let me reiterate that assigning weights to a/f pairs is not a matter of eliciting preferences. If one sought to elicit preferences among a/f pairs, one would be asking respondents the wrong question, and one would have abandoned the ideal of relying on a public prioritization of what matters about health. Though it is easier to form individual preferences among a mere 20 a/f pairs than among hundreds or thousands of health states, the task of preference elicitation is still not well motivated. Health economists would still be passing the buck if they refused to specify the criteria in terms of which a/f pairs should be ranked. There is no reason why an aggregation of personal preferences among a/f pairs should match their public value.

---

19 In their interviews Rosser and Kind informed their respondents of the implications of the weights for the proportion of resources allocated for the relief of people in different health states and for the point of indifference between curing individuals in different health states.
By specifying the evaluative criterion, simplifying the classification of health states to be evaluated, and relying on a deliberative process and public debate to elicit public values, this proposal attempts to measure accurately those aspects of health that should influence the allocation of health-related resources. Because a/f pairs specify the relevant consequences of health states, they do not require respondents to work out for themselves what the consequences are. Hence, they are easier to evaluate. Defining the evaluative criterion in terms of the a/f pairs that should be ranked screens out irrelevant considerations. In addition, rankings of the extent to which a/f pairs enhance or diminish capabilities are less likely to be confounded with judgments concerning the overall value of people’s lives and so less likely to suggest offensive views concerning the lives of those who are disabled. Instead of attempting to value a/f pairs in terms of personal choice-worthiness, the deliberative groups aim to determine their public value and the deliberative process helps to compensate for the restricted experience, knowledge, and wisdom of individual group members. Moreover, there is no aggregation or averaging to be done. What is at issue are public not personal values, to be determined by reasons, rather than by polling or voting.

7. CONCLUSION

The proposal defended here to classify health states by the feelings they involve and the activity limitations they cause and then to evaluate them by their consequences for the range of lives people can lead has three main advantages over the current practice of assessing health states by aggregating personal preferences among health states, which are individuated in terms of functional deficiencies and their consequences. First, by asking individuals in deliberative groups to evaluate a/f pairs rather than functionally individuated health states, this proposal demands less knowledge of respondents. Rather than relying on survey respondents to figure out how what aspects of health states are most relevant to their public value, this task is accomplished by the mapping from health states to a/f pairs. Hence, the cognitive load on individuals is diminished. Second, this proposal specifies clearly the criterion in terms of the health states that should be evaluated. Evaluators should consider how health states affect the range of valuable lives individuals can live and on the range of important objectives they can pursue rather than relying on the potpourri of considerations influencing preferences. Third, employing a deliberative method further lessens the burdens on individuals and helps to make assessment turn on what matters to the public value of health.

This is a programmatic paper whose main purpose has been to sketch a new way to evaluate health. To be worthy of serious consideration, its proposals need to be articulated, and a great deal of empirical work needs to be done. However, nobody is going to do that work unless he or she is convinced that the proposal is coherent and feasible and that it promises significant advantages over current practice. I hope to have begun doing that convincing. As a reasonable first approximation, the relevant information about health states from the perspective of health policy concerns how people feel and what people activities people can carry out. Using that information, health states should then be evaluated by means of a deliberative process in terms of their consequences for the range of valuable lives individuals can live. In this way, health economists can quantify the importance of feelings and activity limitations and thereby construct a sensible measure of health that can be used to guide health resource allocation decisions.

ACKNOWLEDGEMENTS

This paper has been written over a long period and has benefited from advice and criticism from more individuals than I can easily remember. In addition to the thoughtful referees for this journal, I would like to thank Yukiko Asada, Dan Brock, Paul Dolan, Dennis Fryback, Paul Kind, John Mullahy, Sarah Marchand, Erik Nord, Alberto Polloni, Dan Wikler, Benjamin Wilfond, and both the members of a seminar I taught during the Spring of 2006 on Health, Welfare, and Preferences and a sympathetic audience at University College, London.
REFERENCES


