

THE UNIVERSITY *of York*

CENTRE FOR HEALTH ECONOMICS
YORK HEALTH ECONOMICS CONSORTIUM
NHS CENTRE FOR REVIEWS & DISSEMINATION

Interpersonal Comparisons of Welfare

Alan Williams

DISCUSSION PAPER 151

INTERPERSONAL COMPARISONS OF WELFARE

Alan Williams

October 1996

CENTRE FOR HEALTH ECONOMICS DISCUSSION PAPER SERIES

The Discussion Paper (DP) Series was established in 1984 to make research available quickly to a broad readership drawn from the academic community, professional health care staff and policy makers, both within the UK and internationally. Material published in DP form ranges from work in progress, to edited versions of completed project reports, from papers intended ultimately for journal publication, to those which seek to stimulate immediate debate. Hence the variation in length and complexity of DPs.

In the past, DPs have been subject to a minimum of editorial revision, it being seen principally as the responsibility of individual authors to ensure that appropriate standards were met. Since June 1996, a more formal system of internal peer-review has been introduced to consolidate the high levels of quality that have been achieved in the past. That system involves the refereeing of DPs in draft form, with any recommended revisions being verified independently by a senior member of staff.

DPs are not a substitute for publication in peer-reviewed journals. It is to be hoped that journal editors will recognise the nature of the DP series, and accept that this does not compromise the status of those papers published by the Centre which may later be submitted for consideration in more formal publications. Comments and suggestions on DPs are encouraged and should be sent directly to the authors.

THE AUTHOR

Alan Williams is Professor of Economics at the University of York, undertaking research in the Centre for Health Economics on the ethics and economics of priority setting in health care.

FURTHER COPIES

Further copies of this document are available (at price £5.00 to cover the cost of publication, postage and packing) from:

The Publications Office
Centre for Health Economics
University of York
York
YO1 5DD

Please make cheques payable to the University of York. Details of other papers can be obtained from the same address or telephone York (01904) 433648 or 433718.

ABSTRACT

Amongst the many methodological and philosophical problems that have to be confronted sooner or later by those involved in the measurement of health status is how best to compare the health status of one person with that of others. It is an activity which some people do routinely many times a day, yet which others say is totally illegitimate (on moral and/or on scientific grounds). My purpose in this short essay is to review this territory from the viewpoint of a health economist, and to offer a way forward which seems practical and yet morally and scientifically defensible (which is not to claim that it should therefore please everybody!).

INTRA-PERSONAL COMPARISONS

To get into this difficult territory gently, let us first of all start with the problems encountered when making INTRA-personal comparisons of welfare related to health. Typical questions are:

- A.1 How ill is A?
- A.2 Is A getting better or worse?
- A.3 How much better is A today than yesterday?
- A.4 Would treatment X or treatment Y be better for A?
- A.5 How much better for A would the better treatment be?

These are the judgements that are made daily by health care professionals, other carers, and patients themselves. How do they do it?

There are typically four classes of evidence that are brought to bear by all three of the parties mentioned, though they may each select and weight particular items differently. The first distinction to be made in classifying this evidence is between that relating to the person's current health status and that relating to the person's expected future health status. The second distinction is between survival or life expectancy considerations on the one hand, and health-related quality-of-life (HRQOL) considerations (such as pain and disability)¹ on the other. These distinctions generate the fourfold classification set out in the accompanying Table, and some typical examples of each class of information are given there. In addition to this evidence, the health care professionals will also use a wider set of phenomena, of a more technical kind, to do with the underlying pathology of the medical conditions in question. Some of the symptoms relating to the medical data will also be directly accessible to lay people (eg fever, nausea, unusual appearance), though they may well be less good at judging their significance. It is also likely that the patients and their informal carers will

¹ Harri Sintonen's major research contributions in this field are epitomised in the development of the EuroQol Instrument and in his own 15D instrument. The former covers the HRQOL dimensions of mobility, selfcare, usual activities, pain/discomfort, and anxiety/depression. The latter covers moving, seeing, hearing, breathing, sleeping, eating, communicating, eliminating, working, social perception, mental functioning, pain/ache, depression, distress, and perceived health (see Sintonen H and Pekurinen M "A fifteen-dimensional measure of health-related quality of life (15D) and its applications", in Walker SR and Rosser RM (eds) Quality of Life Assessment: Key Issues in the 1990s, Kluwer 1993 pp 185-195.

use comparative data from the patient's own past health, which may not be accessible to the health care professionals.

Table 1 Evidence typically used in making intra-personal comparisons of health status

Phenomenon	Time Horizon	
	Current	Future
SURVIVAL	Alive or dead?	Likelihood that still alive at some point Life expectancy
HEALTH-RELATED QUALITY OF LIFE	How.... immobile? conscious? much pain? able to care for self? cheerful or depressed?	How much improvement can be expected how soon in each of the phenomena in the previous column?

But actually processing this evidence may not be a straightforward task, and differences of opinion may well emerge between the various parties even when they are agreed about the evidence itself. The most straightforward matter, on which there is rarely disagreement, is whether the person is alive or dead. But the person's survival prospects, which should in principle be a factual matter determined by evidence, will usually turn out to be a matter of judgement, either because the relevant predictive model does not exist, or because even the best available predictive model is subject to wide margins of error. But predictive models are based on data from many individuals, so in the process of estimating someone's survival prospects we are implicitly asserting that that person is similar in all significant respects to the people whose data entered that predictive model. The assumption of equivalence which underlies this interpersonal comparison may be purely factual, eg based only on information such as age and sex, but it is more likely that it will also be based on current severity of illness in HRQOL terms, in which case it will raise all the issues to be discussed below. It therefore turns out that we cannot judge a person's survival prospects without making interpersonal comparisons which commit us to assuming that certain things are equivalent between this person and other people, and some of these "equivalences" may go beyond objective facts.

Amongst the things that "go beyond objective facts" is the assessment of a person's HRQOL. Here too there may be some "objective facts" about which there will be little disagreement (eg that the person is bedridden). But some of the supposedly factual (descriptive) evidence may be available directly only to the patient (eg as regards his/her mood or level of pain). Observers assessments must necessarily be based on the indirect evidence of what they see or are told, with the result that they may have different opinions about these "facts". But more important than these potential differences about how a person's HRQOL is to be described, is the problem that what is "serious" and what is "slight" is a matter on which different people may (legitimately) make different judgements depending not only on the "facts" of the situation but also on their personal value system. One person may think that not being able to walk is far more serious than being in continuous moderate pain, whilst someone else thinks that the reverse is true. So judging the seriousness of a health state can be a problem, even though it is one and the same person (at a given point in time) that we are thinking about, and even though we agree on how the health state is to be described.

This disagreement about values could be resolved by deciding that only one party has the right (or expertise) to make such a judgement. One obvious candidate for the dominant role is the patient him- or her-self (provided he or she has all the available relevant factual evidence). Another candidate is one of the health care professionals, on the grounds that they have a lot of technical knowledge not available to the other parties, and a lot of accumulated knowledge from dealing with other patients in similar situations. But this last observation is capable of two rather different interpretations. It could be taken to refer to the accumulation of factual knowledge that enables better predictions to be made. But it could also be a claim that, against the background of all the other cases known to them of people with this condition, this is (say) a mild case. In both cases interpersonal comparisons again enter the picture, some of which will be factual and some valuational.

It is for reasons such as these that preference-based HRQOL measures have been developed which separate description from valuation (to the extent that it is possible to do so).² Ideally such measures should be capable of distinguishing differences in judgements about the severity of health

states that stem from differences in the way different states are described, from differences in the value attached to a given state once it has been described.

So where does this leave us with the five typical questions posed earlier? First of all, "How ill is A?". It seems that this can only be answered without making interpersonal comparisons if we restrict ourselves to A's current situation, since prediction inevitably requires us to assign A to a class of other people whose experiences we know and whose experiences we expect A to share. But even if we restrict ourselves to using evidence about the person's current health status, apart from the simplest case of deciding whether A is alive or dead, pure factual description does not get us very far, and someone's value judgements have to enter in. Different people may have different value systems, so we may not find a consensus. An unambiguous decision may only emerge if we assign to one person the right to determine "how ill is A?". It is open to argument who that person should be, and that choice may depend on what decision and consequential actions follows from the answer.

The second question was "is A getting better or worse?". This requires us to answer the first question on two successive occasions, which obviously leaves us with all the problems already outlined and some more besides. The most important additional requirement is that whatever system of judgement was used on the first occasion must also be used on the second, otherwise the two judgements are non-comparable. The notion of a "system of judgement" includes the set of values that is used, which may imply that the same person must make the two judgements. Again, we may well find that different people come to different conclusions as to the correct answer to this question in a given situation. This second question also brings to the fore something that was implicit in the process by which the first question was answered, namely that there has to be an unambiguous ordering of the health states. This requirement is now more crucial, especially if it is acceptable for the answer to this second question to be "no change", for this will expose whether the ordering is a "weak" or a "strong" one (ie whether or not it contains states that are judged to be of equal seriousness).

² The caveat "to the extent that it is possible to do so" is entered because all HRQOL measures have to select some dimensions and reject others, which is to imply that the rejected items are of relatively little value. Moreover, within the included items terms will typically be used which imply some ordinal ranking at a descriptive level (eg mild, moderate, severe) which has limited valuational content.

The third question was "How much better is A today than yesterday?". This moves us into much tougher territory, because mere ordinal ranking of health states will no longer suffice. Strictly speaking, we now need a measure of health seriousness with interval scale properties, and it is at this stage that informal methods of judgement are likely to fail us. It is true that we commonly use terms such as "a lot better" or "slightly better" to express what is, in effect, an ordinal ranking of differences, and for most intra-personal purposes this may suffice. Going further is a major scientific task and should perhaps not be attempted if intra-personal judgements are all that we need.

The fourth question was "Would treatment X or treatment Y be better for A?". Here we are in somewhat deeper water because we again have to begin by assigning A to a class of people who share A's current situation in all significant respects, and then assert that A is likely to share their experiences. Since their experiences are typically quite varied, this range of collective variation is typically translated into a statement of probabilities facing A. If these prospective experiences are described in a suitable way, A's values (or someone else's values) could be applied to them to determine which of the two prospects are "better for A".³ Otherwise, we are back with the problems generated by the second question.

The final question was "How much better for A would the better treatment be?". This question contains an ambiguity which has first to be resolved, as to whether the question is "how much better for A is the better treatment over the worse treatment?" or "how much better for A is the better treatment over no treatment?" This distinction is important when interpreting the relevance to A of data from clinical trials, some of which compare different treatments, whilst others compare some treatment with no treatment (or with placebo treatment). But to answer either version of this question we have to address simultaneously all the problems mentioned in relation to each of the previous questions. If all that is needed is an ordinal ranking of differences, answers such as "a lot better" or "not much better" may do. But if A or A's family are going to have to pay (in some

³ A whole other set of ambiguities are buried away here which concern the policy relevance of possible differences between A's own "preferences" against professional judgements of A's "needs". For further exploration of this treacherous territory see Alan Williams "Priorities - not Needs!" in Corden A, Robertson E, Tolley K (eds) Meeting Needs in an Affluent Society Avebury 1992 pp 57-63.

form or other) for A's treatment this may not be sufficient, and a much more carefully specified answer may be required (eg "an extension of life expectancy by five years, with moderate pain and some slight difficulty walking, but able to get on with usual activities and able to care for yourself", as opposed to "....."). This more careful specification may well bring into prominence the need to weigh changes in quality of life against changes in life expectancy. Moreover, it is not only patients and their families who may be concerned about the careful balancing of benefits against costs. A more precise measure of benefits may also be required if some third-party payer is going to foot the bill for the treatment, for they may have some cut-off point below which they consider the treatment insufficiently beneficial to justify its costs. In making this judgement they may well insist that benefits and costs are measured in a pre-specified manner, and this may well require a full interval scale measurement of benefits based on an explicit set of values. All this may be required even though we are still only dealing with intra-personal comparisons of the value to be attached to one person being in different health states!

In summary, the broad notion that it is OK to make intra-personal comparisons of welfare with respect to health states, but once we get into interpersonal comparisons we are in deep trouble, has been shown to be far too simplistic. Many intra-personal comparisons necessarily entail the making of interpersonal judgements, and if the intra-personal comparisons need to go beyond ordinal rankings (of states or of differences between states) the measurement requirements can only be met by properly developed preference-based HRQOL measures, and these may need to be capable of integrating the value of changes in length of life into the final measure of benefit.

INTERPERSONAL COMPARISONS: PERSON TO PERSON

Interpersonal comparisons can be divided into two categories: direct comparisons between one person and another, and statements about whole populations. In this section I will concentrate on direct comparisons between one person and another, and in the following section I will deal with whole populations.

Direct comparisons between one person and another are frequently made when establishing clinical priorities, ie when deciding who needs to be treated now and who could wait, who should be

referred for more specialised assessment and/or treatment, etc. Again it will be useful to start with some typical questions on which to focus the argument, which, following the earlier pattern, generates questions which in their transmuted form now are:

- B.1 How ill is A compared with B?
- B.2 Has A improved more than B?
- B.3 How much more has A improved than B?

The fourth and fifth questions posed earlier (about alternative treatments) are answerable in this new context once these three basic questions are answerable.

The new problem which arises here is that A and B may not be suffering from the same medical condition, and they may not have the same background characteristics (age, sex, previous history of illness, etc). This means that we are having to bring within a common frame of reference very disparate information. This implies that the “evidence” contained in Table 1 has to be generic (ie not specific to a particular condition, treatment, class of patient, and not specific to the time and place at which such an assessment is being made). It also requires a common set of values to be applied. But whose values? In general we cannot assume that A’s and B’s values will be the same, so we may get different answers to each of the questions depending on whether we adopt A’s perspective or B’s perspective. There are several ways out of this, namely:

1. to assign one or other of them the right to make the judgement
2. to devise some “average” set of values from their individual values (weighted equally?) and use this “compromise” set of values
3. to hand the problem over to a third party who will use some independently determined set of values (not necessarily his or her own).

In practical terms the third solution seems to be the one most commonly adopted, with the health care professionals playing the arbiter’s role.

If the issues raised above were resolved satisfactorily, the first two questions would be answerable. The third question requires a further step, namely a level of quantification that has interval scale properties, as mentioned in the previous section. This sets rather stringent requirements upon HRQOL measures, and makes one wonder whether the health care professionals, who typically make these judgements rather informally, realise the implications of what they are doing.

INTERPERSONAL COMPARISONS: POPULATIONS

The second category of interpersonal comparisons concern whole populations, since any aggregation requires a common basis of comparison, and hence, in the health care field, some assumption about what is equivalent between one person and another. The standard questions then become:

- C.1 How ill is a particular population?
- C.2 Is their health getting better or worse?
- C.3 How much better is their health this year compared with last year?
- C.4 Would treatment X or treatment Y be better for a particular population?
- C.5 How much better for that population would the better treatment be?

It will be obvious that all that has happened to the original version of these questions is that I have replaced "A" with "a particular population".

First of all, "a particular population" can mean a whole community, as when one of the questions is posed with respect to the general population living in a particular area served by a health authority or a hospital. Or it could refer to an insured population, or those on a practice list, or old people or children or people at risk of having some specific medical condition. Yet again, "a particular population" could be those in different arms of a clinical trial. But they all share the important feature that the relevant health status data, although collected from individuals, has to be aggregated. In the process of aggregation the different items have to be defined clearly enough for each to be counted in the same way for each individual, and each item of information from each

individual has to be given some explicit weight in the aggregation process. The usual weight is 1, and it is usually the same for everybody.

For instance, using survival rates as a measure of health requires death to be defined, and the date and time at which the observations are to be made has to be specified. Thus, in clinical trials, it is common to use measures such as the 1-year survival rate to compare outcomes between treatments. It implies that all survivals are equal (ie it does not matter whether a survivor is at death's door or 100% fit, and everyone is anonymous, so who they are does not matter). In more sophisticated analyses survival rates may be differentiated by age or sex or by some other personal characteristic, so that only within-group survivals are then considered equivalent (though the use of explicitly differentiated weights between groups is rare). If "survival rates" are replaced by "life years gained" or "changes in life expectancy", this will, in effect, weight a "survival" by the number of years of life which are expected to follow it. The implicit assumption then is that an additional year of life is held to be equivalent no matter who gets it (ie it does not matter what the quality of those additional years is). The point I am emphasising here is that these conventional and widely used (and apparently unproblematical) measures of population health all involve a commitment to something being of equal value no matter who gets it.⁴

As has been remarked several times already, we should not expect consensus about what is to be held to be equivalent between different members of any "particular population". Even if it were agreed that this equivalence choice is something they should decide for themselves, it is most unlikely that they would be able to reach unanimous agreement. Moreover, there is no scientific way of resolving this matter, though the collection of evidence about what different people think might help make the dialogue better informed (though possibly at the cost of making a decision more difficult). At this stage we are facing an essentially political question concerning whose values should count, and for how much. The aggregation process "solves" that political problem, often unwittingly, by defining what is to be counted, and how the counted items are to be combined into an aggregate. It also goes a step further and chooses an appropriate measure of central tendency (usually the arithmetic mean) with which to summarise the data, which again "solves" an

⁴ Obviously the same is true of morbidity measures, which assume that the condition in question is equally serious no matter who gets it.

unstated political question, namely what weight should we give to the "extremes" in a skewed distribution? In a political system where decisions are usually made by majority voting, the position represented by the median person might be more significant for policy purposes.

All of these problems are brought to the fore in the construction of preference based HRQOL indexes, and the more unthinking commentators have come to regard them as problems that you get into with HRQOL measures but which can somehow or other be avoided with more conventional measures. My purpose here has been to demonstrate that this is not so. Any measure of health that applies to a whole population will (knowingly or unknowingly) have taken a position on each of these issues. And, if used to recommend one action or situation compared with another, all such measures will necessarily involve the making of interpersonal comparisons of welfare.

ECONOMISTS' HESITATIONS

Many practising economists believe that by working within the framework of Paretian Welfare Economics they can avoid these problems. They believe (rightly) that there is no scientific way to make interpersonal comparisons of welfare, hence any procedure which purports to do so is suspect and not part of the proper corpus of economics. I now want to examine the basis of that argument, in the course of which I shall seek to demonstrate that it is both untenable and unhelpful, and should be abandoned. I will offer something in its place.

In the old days (ie about a century ago) economists held that social welfare was maximised when the sum of individual utilities was maximised. They regarded individual utilities as cardinally measurable (in principle at least), and interpersonal comparisons of utility were permitted (and used, for example, to support arguments for progressive taxation). It was acknowledged that this required a common rubric for utility measurement, but with it the maximisation of social welfare could proceed unconstrained.

The "new" welfare economics of the 1930s and 1940s changed all that. Concern about the stringency of the assumptions required by the old ways of thinking led to a more restrained

formulation becoming the dominant mode of thinking. In this "Paretian" framework of thought, social welfare is maximised when no individual's utility can be increased without some other individual's utility being decreased. This requires individual utilities only to be ordinally measurable, it does not require interpersonal comparisons of utility, and it does not require a common rubric for utility measurement (each individual can do it in his or her own way). The problem is that it is constrained to making statements only about movements from a specified situation, and so the judgements which emerge are necessarily context-specific. And any situation in which someone is made worse off is "out-of-bounds" to the assessor.

This last point proved fatal to the pure Paretian position, because social choice situations in which nobody is made worse off are so rare as to be negligible. So a modified rule was recommended, which is that social welfare is maximised when it is no longer possible to find a change in which, after the gainers have fully compensated the losers in money terms, at least one individual's utility has increased and nobody's utility has decreased. So we now have individual utility translated into money terms, but with no common rubric for utility measurement itself. Interpersonal comparisons of utility are still avoided, but we are constrained to the initial situation as the base line from which judgements are made.

Unfortunately this did not enable much progress to be made in practice, because it is not feasible to identify all the gainers and losers and to organise the compensating transfers except at inordinate cost. So the final transmutation was to say that social welfare is maximised when it is no longer possible to find a change in which the money value of the gains to the gainers exceeds the money value of the losses to the losers, even though no compensation is actually made.⁵ This leaves us in a situation in which individual utility is measured in money terms as a common rubric, interpersonal comparisons are no longer avoided, and the initial situation imposes limits on what can be said.

⁵ This might not be too bad if it could be shown that the net redistributive effects of thousands of individual public projects and programmes were scattered randomly throughout the society, so that on balance everyone gained something. I have explained elsewhere why I think this is inherently implausible, and what might be done about it. See Williams A - "Income Redistribution and Public Expenditure Decisions", in Posner MV (editor) Public Expenditure: Allocation between Competing Ends Cambridge Univ Press 1977 pp 65-80.

This is where applied modern welfare economics currently rests⁶, leaving us with some rather severe problems, of which I wish to highlight the following:

1. it takes the initial distribution of utilities as the starting point for all welfare judgements;
2. it ignores the general unacceptability of the principle that so long as nobody is made worse off it does not matter who is made better off;
3. cardinal measurement of utility is supposedly out (von Neumann-Morgenstern utilities being a possible exception!) so only ordinal comparisons are possible;
4. it allows interpersonal comparisons of utility to sneak back in unrecognised.

Concentrating on this last point, if interpersonal comparisons of welfare are, in practical terms, inescapable, would it not be better to come out into the open and address the issue directly and consider the best way to make them. Coming out into the open about it would also allow us to escape from Arrow's Impossibility Theorem, which arises fundamentally from the fact that we are not allowed to find solutions to social choice problems which make such comparisons. He showed that it was a vain search if the rejection of interpersonal comparisons were combined with other plausible desiderata for social choice mechanisms.

THE VIEW FROM HEALTH ECONOMICS

As was argued earlier, the problem of outcome measurement in health care cannot be addressed in an acceptable manner without making interpersonal comparisons of welfare. Even apparently simple intra-personal judgements about the seriousness of that person's illness, or about the relative efficacy of two treatments for that individual, require interpersonal comparisons. Health care professionals, patients and their relatives make such judgements daily. When we economists are patients and relatives we do so too. So why do we hold back when acting in our professional capacity? After all, our supposed scruples about never doing so are clearly self-delusion. We use

⁶ For those who think that this chatty account of the history of this bit of economic thought is a travesty of what actually happened, I recommend a much more scholarly account which might convince them that I am more or less right, namely: Backhouse R A History of Modern Economic Analysis, Basil Blackwell, Oxford, 1985, esp Chapters 1, 15, 24 and 30.

unemployment rates to judge rival policies just like doctors use survival rates to judge rival treatments. Quite apart from the more specific calculations of the money value of the loss of consumer surplus in Paretian Cost-Benefit Analysis (in the transport field for instance), we use sums of money as proxies for welfare (eg GNP per head) in numerous contexts, and compare the welfare of whole populations on that basis. Who are we kidding?

This is not to argue that we should not continue to remind everybody (if they need reminding) that there is no scientific basis for making interpersonal comparisons of welfare. But some of the attributes of scientific ways of thinking can nevertheless be adduced to help to discipline the process by which such judgements are made. The first of these is that they must be made explicit, and their implications honestly presented. Others may then criticise any particular judgement, on the grounds that it has implications that have not been properly explored, and/or that the implications are (in their view) undesirable. In the latter case they ought to offer an alternative which overcomes that undesirability, so that others, in turn, can explore its implications. Critics should not be permitted to claim or imply that no interpersonal comparisons are necessary, because leaving a void at the centre of an argument is scientifically unacceptable.

The actual choice of the basis for making interpersonal comparisons of welfare in any context must depend on its appropriateness to the policy issues involved. If health care policy is guided by the principle that a person's income and wealth should not determine priorities, then valuations of health that are influenced by those elements should not be used when coming to some aggregate measure for community use. Similarly, if we want equals to be treated equally, we must be prepared to specify what it is that is to be held to be equal between individuals.⁷ Do we want it to be a "survival", a year of life, a quality-adjusted year of life, an age-adjusted year of life, or what? And should priority be given to some (eg parents of young children) over others (eg their childless contemporaries) when health benefits are being aggregated? Should people who have cared for their own health count for more than those who haven't? All these are active issues in the health care field, and economists can do a lot to illuminate them by exploring alternatives. It seems both

foolish and irresponsible not to do so, by pretending that when interpersonal comparisons of welfare enter the picture, economics is silent.

CONCLUSIONS

Instead of the economics profession adopting a holier-than-thou attitude concerning the unscientific basis of interpersonal comparisons of welfare, it would surely be more honest and more productive to make such comparisons openly, and to make their implications as clear as possible. It would also be helpful to offer some justification for any such judgements, especially where there are obviously others that could have been made (and even more especially if the chosen ones differ from those that others conventionally make).

This is what the health economists have done in the HRQOL field. For instance, when advocating the use, in a social choice context, of preference-based measures based on the views of the general public, one is committed to the following propositions:

1. the views of all the citizens affected by a decision should be taken into account when judging its desirability;
2. in the absence of consensus, the mean or median value should be taken to represent the views of the citizenry as a whole;
3. being healthy is the same for everybody, and being dead is the same for everybody.

The first two assumptions fall within the realm of political philosophy, and other analysts may hold other views, eg that some people's views should count for more than others, and perhaps even that some people's views should not be counted at all. But these "others" know precisely what they are shooting at, and intellectual integrity requires them to offer some other proposition in place of what they object to, and follow through its implications thoroughly and honestly so that they can be

⁷ I have explored this matter further in "Economics, Society and Health Care Ethics", Gillon R (ed) Principles of Health Care Ethics, Wiley 1994, pp 829-842.

subjected to the same critical scrutiny. The third assumption is implied by the common practice of standardising individual valuations so that dead equals zero and healthy equals one. Its justification has to be that it is the appropriate ethical position to adopt when making social choices in health care. This may not be so in all health care systems, and, as indicated earlier, it raises some interesting issues about possible alternative formulations of what is to be regarded as equal between people.

There are doubtless other tricky assumptions of an interpersonal kind implied by the construction of health status indices, and they need to be teased out and explored. But as indicated earlier, many of them are also implied by conventional measures, which do not explicitly contain valuations or preferences and yet which are used to recommend one course of action over another. In my view, these measures are more dangerous than the explicitly preference-based ones, because it is extremely difficult to detect where the implied valuations enter and on what evidence they are based.⁸ We seem to be in the ludicrous situation that the more arbitrary the scoring system the less criticism it attracts, a situation which ought not to be permitted to occur in a scientific community.

Economists should continue to lay bare for independent inspection the assumptions upon which their work is based, and I think that they should expand the range of these assumptions by abandoning the self-denying (and self-deceiving) ordinance that interpersonal comparisons of welfare are out of bounds. But I also think that those who criticise our assumptions should apply their critical stance in an even-handed manner, so that the same criteria are applied to what the others are doing in the same field, so as to avoid the impression that what has not been criticised (because it is part of the conventional wisdom?) must be OK.

Reinhardt⁹ has trenchantly criticised the role of Paretian Welfare Economics in the analysis of health care systems, and Culyer¹⁰ has advocated that health economics should break out of the

⁸ See, for instance, my critique of SF-36 in "Review Article", Health Economics 1(4) 1992 pp 255-258.

⁹ Reinhardt UE "Reflections on the Meaning of Efficiency: Can Efficiency be Separated from Equity?" Yale Law and Policy Review 10 1992 pp 302-315.

¹⁰ Culyer AJ - "The Normative Economics of Health Care Finance and Provision" Oxford Review of Economic Policy, 5 (1989) pp 34-58.

conventional Paretian framework altogether and adopt an “extra-welfarist” perspective in which the extreme individualistic stance is modified. Sugden and I pursued a similar line many years earlier (though in a more pragmatic vein), distinguishing the Paretian approach to cost-benefit analysis from “the decision-making approach”¹¹ the latter being characterised by greater freedom in the choice of policy objectives and in the choice of value source. The scientific hallmark remains consistency and explicitness, but interpersonal comparisons of welfare should be contained within it.

It may be that we are in the early stages of a Kuhnian¹² “scientific revolution” in the field of welfare economics, in which the Paretian paradigm is slowly being replaced by a new one. If so, health economists are going to have a hard time as they attempt, at one and the same time, to persuade health care professionals to adopt modes of thinking that flow from conventional economics but have been modified to make them more obviously applicable to health care, while at the same time trying to persuade mainstream economists that these modifications are not mere tactical adjustments to win over the non-economists, but offer a means by which economists can apply their analytical skills to a much broader range of problems than they have hitherto felt able to tackle. But the defenders of the old order occupy well explored and familiar territory with which the cognoscenti are comfortable and into which all novices are initiated. Those groping for a resolution of otherwise intractable problems have no such security their ideas are exploratory, they will often make mistakes or turn out to have gone up blind alleys, and any result which challenges the conventional wisdom will be subject to scrutiny with a thoroughness (and even hostility) that more welcome findings escape. Thus reason may appear to be on the side of the methodological conservatives, despite their inability to grapple satisfactorily with quite pressing social policy problems. That is why the challenge to critics must always be “so, working within your paradigm, how, in a practical context, would you solve this problem?”. If they come up with a good answer, the revolution is premature!

¹¹ Sugden R and Williams A The Principles of Practical Cost-Benefit Analysis, Oxford University Press 1978, see especially Chapters 7 and 16. It is a theme I have returned to more recently in “Cost-Benefit Analysis: Applied Welfare Economics or General Decision Aid?” in Williams A and Giardina E (eds) Efficiency in the Public Sector, Edward Elgar, 1993, pp 65-79.

¹² Kuhn TS The Structure of Scientific Revolutions Univ of Chicago Press, Second edition, 1970.