Journal of Health Economics 10 (1991) 21–41. North-Holland

QALYs and the equity-efficiency trade-off

Adam Wagstaff*

School of Social Sciences, University of Sussex, Falmer, Brighton BN1 9QN, UK

Received May 1990, final version received November 1990

As the volume of research on quality-adjusted life years (QALYs) has increased, concern has begun to be expressed about the equity aspects of resource allocation decisions based on the results of this research. This paper suggests that a common theme running through the criticisms of the QALY approach is a concern about *inequality*. It also suggests that the method for incorporating distributional concerns which is currently being pursued by advocates of the QALY approach will only ever capture concerns *other than* a concern about inequality. The paper suggests a method for incorporating *both* sets of concerns into resource allocation decisions.

1. Introduction

As the volume of research on quality-adjusted life years (QALYs) has increased, concerns have begun to be expressed about the equity implications of this research. Lockwood (1988), for example, has argued that the QALY approach 'is in principle liable to result in forms of allocation that are *unjust* or *unfair*' [Lockwood (1988, p. 45), emphasis in original]. Similar concerns have been expressed by inter alia Broome (1988), Harris (1988) and Smith (1987).

That these concerns continue to be expressed despite attempts by advocates of the QALY approach to reassure critics suggests that there has been a misunderstanding either on the part of the critics of the QALY approach or on the part of its advocates. One of the aims of this paper is to find out the nature of any misunderstanding. More generally the aim of the paper is to explore the nature of the trade-off between equity and efficiency in the context of resource allocation via QALYs. To do so clearly requires

*I am grateful to John Broome, Tony Culyer, Eddy van Doorslaer, Alan Maynard, Alan Williams and an anonymous referee for helpful comments on earlier versions of this paper. The usual disclaimer applies.

consideration of the concepts of equity and efficiency. What do advocates of the QALY approach mean by efficiency? What is the concept of equity underlying the QALY approach? Is it at variance with that proposed by the critics of the QALY approach? What is the nature of the trade-off between efficiency and equity (if any)? How could such a trade-off be taken into account in resource allocation decisions?

2. Efficiency as health maximization

The policy objective underlying the QALY literature is the maximization of the community's health. An individual's 'health' is measured in terms of QALYs and the community's health is measured as the *sum* of QALYs. An individual's QALY score is calculated by weighting each remaining year of his life by the expected quality of life in the year in question. The latter is calculated as a probability-weighted average of the quality-of-life scores associated with each of the possible health states the individual might find himself in. These scores, which are invariant across individuals in the same health state, are measured on a scale ranging from (typically) zero (dead) to one (perfect health). Frequently, they are derived using utility theory in an experimental setting. Thus, writers frequently refer to the quality-of-life score as a *utility score* and to the QALY approach as *cost-utility analysis* [cf., e.g., Torrance (1986)].

Maximizing health is argued to be a natural objective to want to pursue, given a desire to see resources deployed efficiently. As Drummond (1989) puts it, here we are 'concerned with economic efficiency rather than with notions of equity or social justice' [Drummond (1989, p. 71)]. The goal of health maximization leads, as Williams (1985) notes, to an inescapable conclusion:

> Procedures should be ranked so that activities that generate more gains to health for every £ of resources take priority over those that generate less; thus the general standard of health in the community would be correspondingly higher. (p. 326)

In concrete terms, the calculations of Williams (1985) imply that resources in the British National Health Service (NHS) ought to be redeployed towards procedures such as the insertion of pacemakers for heart block, hip replacement and replacement of valves for aortic stenosis, and away from procedures such as kidney transplants and coronary artery bypass grafting for severe angina with one vessel disease.

By identifying efficiency with health maximization, the QALY approach constitutes a departure from traditional Paretian welfare economics. This is no accident. Advocates of health maximization explicitly reject the traditional

Paretian value judgements [cf. Williams (1977, p. 282)]. Williams (1976) put the case for rejection of the Pareto doctrine in the health care context thus:

> If society has explicitly rejected willingness and ability to pay as an appropriate basis for rationing out health care, the problem of vindication surely shifts onto those who persist in using that criterion... I would hope to persuade thoughtful scholars that the pure Pareto doctrine, à la Mishan, has outlived its usefulness, and is now merely a *starting point* for work in this field, and which we can safely move beyond in the field of health economics. (p. 13)[1]

Culyer (1989, 1990) suggests that the QALY approach amounts to more than a mere rejection of the Pareto doctrine. Also implicit, he argues, is a rejection of *welfarism*: the view that social welfare depends solely on the welfare (i.e. utility) of the members of the community [cf. Sen (1979)]. Building on other work by Sen [Sen (1980, 1982)], Culyer suggests a schema in which goods (e.g. health care) are viewed as possessing characteristics (e.g., clinical efficacy), which in turn influence the characteristics of people (e.g., being able to feed oneself). It is from these latter characteristics that utility is ultimately derived and it is this that is the focus of welfarism. The QALY approach, by contrast, focuses on the penultimate stage: the characteristics of people. Though utility theory *is* frequently used in the derivation of quality-of-life scores, it is used simply to measure people's health rather than the utility they derive from it. That a QALY is not a measure of utility in the welfarist sense becomes obvious when one reflects that the quality-of-life scores used are common to everybody: in resource allocation exercises based on the QALY approach (in contrast to the welfarist approach) two people cannot occupy the same health state and yet experience different utilities.[2]

As a definition of efficiency health maximization is a good deal stronger than the extra-welfarist counterpart of the Pareto criterion – the traditional welfarist concept of efficiency.[3] The axes of fig. 1 indicate the health, measured in terms of expected QALYs remaining before death, of two individuals, or *groups* of similar individuals, A and B. In the absence of

[1]It is perhaps worth noting that this rejection of the Pareto value judgement does not imply that quality-of-life scores should not reflect the views of consumers [cf., e.g., Torrance and Feeny (1989)]. What it *does* imply is that individuals should not be left to decide on their *own* scores and that the rich should not be able somehow to 'buy' higher scores than the poor.

[2]Possibly because of the confusing use of the term utility in the QALY literature some have actually criticized the QALY approach for being welfarist. Thus Lockwood (1988) writes: '... what I want to focus on here is the philosophically more fundamental objection that can be levelled against the QALY approach: namely that, precisely *because* it is uncompromisingly welfarist, it is in principle liable to result in forms of allocation that are *unjust* or *unfair*' (p. 45, emphasis in original).

[3]Since it is presumably considered unethical to administer health care so as to *impair* health, the extra-welfarist analogue of the potential Pareto criterion would seem a nonsense.

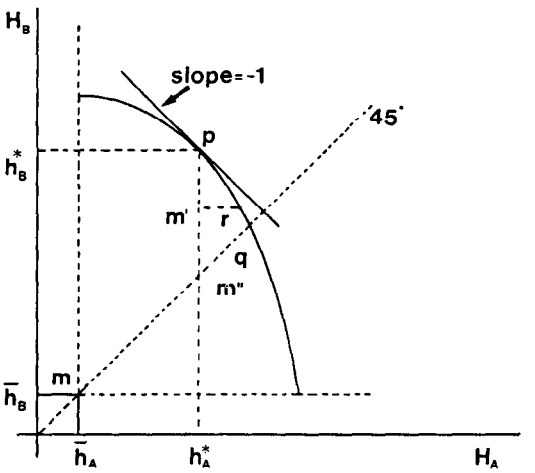

Fig. 1

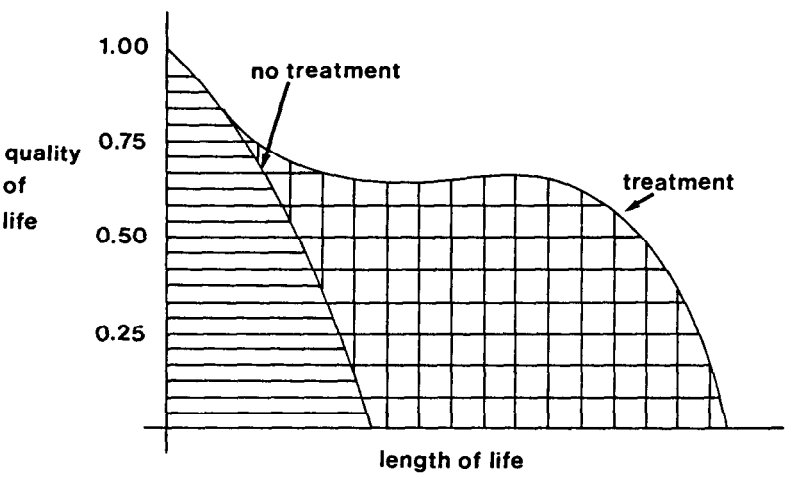

Fig. 2

treatment both A and B would enjoy the same number of QALYs, $\bar{h}_A$ and $\bar{h}_B$. Thus $\bar{h}_A$ corresponds to the shaded area under the no-treatment profile in fig. 2. Point m in fig. 1 can therefore be thought of as an *endowment point*.[4] The location and shape of the bowed-out curve in fig. 1 – the health frontier –

[4]Alternatively $\bar{h}_A$ and $\bar{h}_B$ might be interpreted as the expected QALYs remaining if A and B receive medical management rather than surgery [cf. Williams (1985)].

depends on the resources available, the capacity of A and B to benefit from health care and the costs to society of that health care; the latter include time and money costs incurred by patients, their family and friends, as well as the costs incurred by the health service. In the case illustrated the resource and technological constraints would allow treatment to be administered which would improve the health (i.e., increase the number of QALYs remaining) of *both* A *and* B.

The frontier – the slope of which indicates the marginal cost of a QALY to A in terms of QALYs denied to B – is drawn concave to the origin on the assumption that health care is subject to a diminishing marginal product. This can be interpreted either at an individual level (an individual's capacity to benefit at the margin diminishes as more treatment is administered) or at a group level (if the cases are treated in order of capacity to benefit from treatment, the first cases will enjoy a bigger health improvement than the last). The asymmetrical nature of the frontier in fig. 1 implies that B can be treated at lower cost than A and/or that B has a greater capacity to benefit from treatment than A.[5]

The community's health (defined as the sum of h_A and h_B) is maximized in fig. 1 at point p, where the cost of a QALY is the same at the margin for A and B. The movement from point m to point p involves A receiving an additional number of QALYs equal to $h_A^* - \bar{h}_A$: the latter corresponds to the cross-hatched area lying between the treatment and no-treatment profiles in fig. 2, whilst h_A^* corresponds to the *sum* of the shaded and cross-hatched areas. That health maximization is stronger than the extra-welfarist analogue of the Pareto criterion ought to be obvious from fig. 1.

3. Utilitarianism rediscovered?

The welfarist counterpart of health maximization is, of course, *classical utilitarianism*. The two are *not*, however, the same,[6,7] the difference being that classical utilitarians sought to maximize the sum of *utilities*, whilst advocates of the QALY approach seek to maximize the sum of *QALYs*.[8]

[5]This can be seen by noting that the slope of the frontier at q (where health is equally distributed) is greater than 1 in absolute value. In other words, if we abstract from differences in health status and assume that the two individuals have the same health, the cost of an additional QALY is higher for A than for B.

[6]Some critics of the QALY approach have suggested otherwise. Smith (1987), in his critique of Williams' (1985) paper on QALYs, writes: 'A cost-effectiveness approach to the allocation of health resources presupposes a simple utilitarian or Benthamite concept of justice.' [Smith (1987, p. 135)]. Lockwood (1988) also identifies health maximization with utilitarianism. 'These health-care economists have, it appears, rediscovered utilitarianism. Indeed, the QALY approach has a pleasantly nostalgic air for those familiar with Jeremy Bentham's "felicific calculus".' [Lockwood (1988, p. 39)]. Again, the confusion may be due in part to the confusing use of the term utility in the QALY literature.

[7]Cf. Culyer (1990) on this point.

[8]Recall Culyer's (1990) distinctions noted in section 2.

The fact that the QALY approach, unlike utilitarianism, is not based on individual's own valuations of their health, and instead regards a QALY as of equal value to everybody, allows the QALY approach to avoid what many see as an unattractive implication of utilitarianism, namely that resources ought to be distributed away from people who (for whatever reason) place a relatively low value on their health (e.g., the pecuniary benefits of health improvements may be small for persons who are on low wages or are outside the labour force) [cf. Grossman (1972), Le Grand (1987, p. 14)].

In other respects, however, utilitarianism and health maximization are close bedfellows. Like utilitarianism, health maximization leads to the conclusion that resources ought to be redeployed away from people (or groups of people) who have a low capacity to benefit from treatment. Consider again fig. 1. Suppose that the costs of treating A and B are the same. Then it follows that the asymmetrical shape of the frontier must stem from the fact that B (who ends up receiving more QALYs than A) has a higher capacity to benefit from health care than A, despite the fact that he starts off with the same health as A.

A closely related point is the similarity between utilitarianism and health maximization in their treatment of the elderly. Suppose that A and B both initially enjoy the same quality of life, but that B is younger and therefore starts off with a higher number of expected QALYs. If A and B are identical in other respects, the impact of medical care on their quality of life will be the same, but the productivity of medical care in terms of QALYs will be higher for B than for A, simply because B is younger. The health frontier is therefore asymmetric as before, but the endowment point is a point such as m′ in fig. 1 rather than m. The frontier therefore shrinks (points on the frontier to the north-west and south-east of m′ are now unattainable); B gets all the additional QALYs and A gets none.

Health maximization leads to another notable conclusion, namely that in a fully employed economy resources ought to be redeployed towards people whose output is highly valued and who, as a result of treatment, are able to return to work quickly [cf. Williams (1981)]. Suppose in fig. 1 that the capacity of the A and B to benefit from treatment is the same but that B's output is highly valued and therefore the *net* cost to society of treating him is less than that associated with treating B.[9,10] Health maximization would

[9]It is not entirely clear from Williams (1981) why time appears to be valued at market prices (i.e., wages) in the case of the employed [cf. Williams (1981, p. 272)]. This would seem to run counter to the desire to reject market valuations as a basis for valuations. Nor is it clear why the time of the employed is assigned a positive value (as is the time of housewives), whilst the time of the elderly and mentally handicapped is valued at zero [cf. Williams (1981, p. 279)].

[10]Of course, in this case the notion of a fixed budget becomes a bit lazy. The assumption does, however, seem to underlie the QALY approach.

lead to B receiving more QALYs than A. Unless the highly paid person derives a sufficiently lower utility from health improvements, utilitarianism leads to a similar conclusion.

4. Equity in the QALY literature

Despite the differences between utilitarianism and health maximization one is reminded by the previous section of Sen's remark to the effect that utilitarianism is simply concerned with the sum of utilities and is 'supremely unconcerned with the interpersonal distribution of that sum' [Sen (1973, p. 16)].[11]

It would, of course, be untrue to say that the QALY approach is supremely unconcerned with the distribution of QALYs or even that it is unconcerned with distributional matters. As Drummond (1989, p. 71) puts it, the QALY approach does embody 'a kind of equality' by virtue of the fact that a QALY is considered to be worth the same to everybody. Williams (1974a) put the point more forcibly:

> The acceptance of the fact that an additional year of healthy life is intrinsically worth, say, £1,000, no matter to whom it accrues, would probably lead to a much finer, humanitarian, and egalitarian health service than we have at present, and does not imply the pursuit of profit rules at all, since my supposed figure of £1,000 is merely a precise expression of what society should be prepared to devote to the cause of increasing expectation of healthy life, irrespective of these financial considerations! (p. 255)

The implication seems to be this: since a QALY is regarded as being of equal value to everybody, the outcome of resource allocation via QALYs is *automatically* equitable irrespective of the degree of inequality involved and the type of person who fares badly.

Culyer (1989) offers a more pragmatic defence of health maximization. He suggests that it may actually be equitable to discriminate against those whose capacity to benefit from medical care is limited. If, in fig. 1, B really has a higher capacity to benefit from treatment than A, it is unlikely that his health is initially the same. A more plausible assumption might be that B has worse health (and may therefore be expected to be poorer) than A [cf. Culyer (op. cit., p. 35)]. In this case point m lies below the 45° line, so that the difference between the number of QALYs going to A and B is even

[11]Cf. Pereira (1989).

bigger. Thus in this case the extra resources are going to the person whose health is worse and who happens to be the poorer of the two.

Not all economists, however, feel comfortable with some of the distributional implications of the health maximization approach. Elsewhere Williams has expressed disquiet about the fact that it implies that resources ought, in a fully employed economy, to be redeployed towards those whose output is highly valued [cf. Williams (1981)]. Such a conclusion, he argues, runs counter to egalitarian ethic underlying the principle that a QALY is to be regarded of being of equal value to everybody.[12] Pereira (1989) has challenged Culyer's pragmatic defence of QALY-maximization, arguing that counter-examples are easy to think of: the greater capacity of B to benefit from treatment in fig. 1, for example, might simply stem from the fact that he or she, being rich, is better nourished and more knowledgeable about health matters [cf. Pereira (1989, p. 36), Le Grand (1987, p. 7)].

Advocates of the QALY approach have responded to disquiet about the distributional implications of their work by looking again at the notion that a QALY is to be regarded as of equal value to everyone. One possibility currently being explored is that society may want to attach different values to QALYs going to different people. As Culyer (1990) put it:

> ... do we not feel impelled to cherish the life-years of the very young and the very old?... do we not feel differently about the person whose poor health is the result of their own reckless or feckless behaviour from the way we feel about the person who is more 'responsible'?... (p. 25)

Some preliminary survey results published in Williams (1988) suggest inter alia that there may be some support amongst the population for the idea of discriminating in favour of the young (a popular idea with the elderly in the sample!) and against those who have not taken care of their health (an unpopular idea with smokers in the sample!).

Armed with relative valuations such as these, one could, it is argued, then attach a higher weight to QALYs going to some people (e.g., the young) and a lower weight to QALYs going to others (e.g., those who have failed to take care of their health). This ability to allow QALYs to be weighted differently depending on the person concerned has led to claim that, whatever the prevailing notion of equity, distributional considerations can perfectly easily be integrated into the QALY approach, simply by maximizing a weighted sum of QALYs rather than an unweighted sum [cf., e.g., Culyer (1989, p. 53)]. As Culyer (1990) put it, once this has been done,

[12]Alan Williams has pointed out to me that in his work [cf., e.g., Williams (1985)] he has got round this problem by excluding GNP gains and losses from the calculations. He notes that this can be justified on the grounds that the economy is at less than full employment (a sick person is simply replaced by someone who would have otherwise been unemployed).

> is there any distributional concern left that has not been embodied? If not, the maximizers have the day. If so, then we are perhaps at the heart of what it is the egalitarians fear most from the maximizers. But what it is I cannot discern! (p. 25)

Clearly one might raise objections about the survey approach to eliciting views about equity. It is, after all, difficult to distinguish between what people regard as just (or equitable) and what they regard as desirable. The latter will depend not only on what people think is just, but also on their degree of compassion – or 'caring' as it is termed by Culyer (1980) – and on their own selfish interests. Culyer (1980) put the point rather well:

> ... the source of value for making judgements about equity lies outside, or is extrinsic to, preferences.... The whole point of making a judgement about justice is so to frame it that it is (and can be seen to be) a judgement made independently of the interests of the individual making it. (p. 60)

The response by smokers to the question about people taking care of their own health illustrates the point nicely. Though this issue *is* important, it is not my primary concern here. Instead, the next section considers whether the approach proposed in the recent QALY literature – viz. weighting QALYs according to non-health information, such as age, sex, employment status, etc. – really does embody all (or any) of our equity concerns.

5. QALYs and the equity literature

A number of definitions of equity have been proposed and discussed at length in the health economics literature.[13] Two of these bear close similarities to those proposed by critics of the QALY approach, but interestingly none of them gives rise to the notion that society might want to attach different values to QALYs going to different people.

This section considers these definitions of equity. In each case three questions are asked:

(i) can the definition of equity provide a basis for determining an equitable allocation of resources?;
(ii) if so, is there a conflict between equity and efficiency (defined in terms of health maximization)?; and

[13] See, for example, Le Grand (1982, 1987), Mooney (1983, 1986) and Mooney and McGuire (1987).

(iii) if there *is* a conflict, can the definition of equity be captured simply by weighting QALYs appropriately in the objective function?

5.1. *Equal treatment for equal need*

Under this definition equity requires that persons in equal need of health care receive the same treatment, irrespective of personal characteristics that are irrelevant to 'need', such as ability to pay, gender, place of residence, etc. There is, of course, a vertical equity counterpart to this definition.

The idea that need should determine the distribution of health care resources lies at the heart of Lockwood's (1988) critique of the QALY approach. He writes:

> ... justice has something to do with equality, and something to do with giving appropriate weight to certain sorts of moral *claim*.... The claim that any patient would plausibly be thought to have on the health services ... is a function not so much of the amount of *benefit* that the health services are in a position to confer, as of the person's *needs*, in relation to the services' capacity effectively to meet those needs (p. 45, emphasis in original).

He goes on to suggest that there is a potential conflict between the principle 'To each according to what will generate the most QALYs' and the principle (which Lockwood inclines towards) of 'To each according to his need'. A similar view is expressed by Broome (an economist), who suggests that a fair allocation of resources would be based on the premise that 'equal claims demand equal satisfaction, or more generally that claims demand proportional satisfaction' [Broome (1988, p. 63)]. In a footnote on the next page he suggests that 'need is one of the most plausible sources of claims' [Broome (1988, p. 64)].

Unfortunately, however, neither Lockwood nor Broome offers any definition of the word 'need' both apparently regarding its meaning as self-evident. This is somewhat surprising in view of the controversy that has surrounded the term. As Culyer et al. (1972) put it:

> ... it is difficult to tell, when someone says that 'society needs...', whether he means that *he* needs it, society ought to get it in *his* opinion, whether a *majority* of the members of society want it, or *all* of them want it. Nor is it clear whether what it is 'needed' *regardless* of the cost to society. (p. 114)

Williams (1974b) has suggested that what is meant when a person is said to 'need' medical treatment is that the treatment is required in order to reach

some desired improvement in health. This interpretation of need requires consideration of (i) the scope for improving the person's health through treatment (a *technical* judgement), and (ii) the extent to which the health improvement is desired (a *social valuation* judgement) [cf. Williams (1974b, pp. 71–72)]. This means that one cannot determine how the needs of one person compare with those of another simply by assessing the capacity of each to benefit from treatment. One also has to consider the extent to which society regards the health improvements as desirable. As Culyer (1976) puts it:

> Does the individual dying of thirst in the desert 'need' a glass of water? Technically, yes, if he is to live. Normatively, yes only if others agree that he ought to have it. If they do not, he may *want* it as much as it is possible to want anything, but he does not, in our definition, need it (p. 16, emphasis in original).

In other words, in order to establish needs, society has to decide how the health improvements of one person are to be weighed against those of the other. But of course this cannot be done unless society has first decided on the desired levels of *health* of the two individuals

So, if need is interpreted along the lines suggested by Williams and treatment is therefore regarded simply as an instrument to achieving health improvements, the criterion of equal treatment for equal need cannot *in itself* provide a basis for determining an equitable allocation of resources [cf., Williams (1988, p. 117)]. The answer to question (i) above is therefore 'no'.

5.2. *Equality of access*

Policy statements in Britain and elsewhere make much of the concept of equality of access. The concept of access is, however, often ill-defined in these statements. This confusion is often reflected in the academic literature.[14] However, in Britain at least, there seems to be widespread support amongst economists for Le Grand's (1982) suggestion that access might best be interpreted in terms of the time and money costs that individuals incur in using health care facilities, ideally measured in utility terms.

Mooney (1983), who seems to endorse Le Grand's concept of access, emphasizes that, defined like this, equality of access is all about *equality of opportunity*; whether or not the facility in question is used is immaterial. Frequently it is argued that equality of access should apply only to those in equal need: hence the definition 'equal access for equal need' [cf., e.g., Mooney (1983)]. This justifies persons in unequal need facing different access

[14]On this confusion, see Mooney (1983).

costs, so that, for example, an individual with a minor infection might have to wait longer in a casualty department than a motor accident victim with life-threatening injuries.

It is important to realize that equality of access amongst those in equal need will not guarantee equality of treatment amongst those in equal need [cf. Mooney (1983, p. 183)]. The fact that in countries such as the U.K. government policy tends in practice to involve equalization of expenditures across regions in equal need, even though policy goals tend to be couched in terms of equality of access [cf. Mooney and McGuire (1987, p. 74)], suggests that policy-makers realize this and that they may actually be concerned with treatment rather than access. If, however, 'equal access for equal need' *is* regarded as the appropriate equity objective, one runs again into the problem of defining and establishing needs – something that cannot be done until society has first decided what the desired health levels are of the persons concerned. Thus the answer to question (i) above is again 'no'.

5.3. *Equality of health*

This definition of equity appears to underlie policy documents such as the *Black Report* [Department of Health and Social Security (1980)]. It implies that any reduction (increase) in inequality is to be regarded as a good (bad) thing. Thus any movement away from the 45° line in fig. 1 increases the degree of inequality and is to be regarded as inequitable. The opposite is true of any movement towards the 45° line.[15]

Since the 'equality of health' definition of equity provides a clear statement on the equitable distribution of health, it also provides a clear statement on what constitutes an equitable distribution of health care. One merely has to work back from the desired distribution of health, bearing in mind the current distribution of health, the capacity of individuals to benefit from treatment and the costs of the treatment. Thus in contrast to the definitions of equity considered in the previous two sections, here we are able to answer 'yes' to the first of the questions posed at the beginning of this section.

As can be seen from fig. 1, the answer to the second question is also 'yes'. The efficient distribution of health is at point p and the equitable distribution with the highest attainable per capita level of health is at q. Health maximization requires that B receive more QALYs than A. The 'equality of health' objective, by contrast, requires that A and B receive the same number

[15]Since the objective of equality focuses simply on the distribution of health and not the overall amount of health, this definition of equity has nothing to say about the relative desirability of different points along the 45° line, or indeed along *any* ray through the origin. This means that any move outwards along a ray is neither good nor bad from the standpoint of equity.

of QALYs. A price has to be paid for pursuit of equality: the sum of QALYs is clearly lower than would be the case under health maximization.

In passing it is worth noting that if, as is assumed here, equality of health is interpreted to mean equality of 'expected QALYs remaining', the goal of equalizing health may well not be feasible.[16] This would be the case if point m' were the endowment point. Health maximization would entail giving all the resources to the younger person (B), thereby moving to p. Equity would ideally involve moving to point q, but since this would entail a *reduction* in B's health below its current level, it is not feasible. Point r is the best that can be done from an equity standpoint, so that B receives no treatment at all and A receives all the extra QALYs. These might, of course, take the form simply of an enhanced quality of life rather than any extension of life.

If the equity goal is equality of health, can equity considerations be taken into account simply by weighting QALYs in the health maximizing problem? The answer is 'no'. The reason is that the concern here lies with inequality in the distribution of *health*, not with the question of whether or not A happens to be a compulsive smoker, a parent of a 2-year old child, or whatever. If equality of health is the equity objective, *any* allocation of extra QALYs which increases the degree of inequality of health is a bad thing, *irrespective of the non-health characteristics of the individuals concerned* (cf. section 7 below).

5.4. *Equity and choice*

This is not to say necessarily that non-health characteristics should play no part in distributional considerations. Rather that a concern about them does not stem from a concern about *equality*. This prompts the question: might weighted health maximization be justified by an appeal to some *other* concept of equity that is not based on the notion of equality?

Of particular interest here is Le Grand's (1984, 1987) concept of equity as equal constraints. He argues that inequality is not necessarily inequitable. One cannot, Le Grand (1984) suggests,

> simply observe inequality ... and thereby judge, *on the basis of that inequality alone*, whether or not an allocation is equitable or inequitable (p. 44, emphasis in original).

He suggests that inequalities *are* inequitable if they reflect inequalities in the constraints people face but are *not* if they simply reflect differences in tastes. Le Grand (1987) gives the example of two people who face the same

[16]This is not the only way 'equality of health' might be interpreted. An alternative would be equality of expected QALYs at birth.

constraints but have different preferences: their levels of health are different because one is a smoker and the other is not. He then comments:

> But this does not seem inequitable. Both were fully aware of the dangers involved; both were unconstrained in their choice by other factors; both have made informed decisions based on their own preferences. The results of these decisions are different, and that is reflected in disparities in their health states; but that is the outcome of their own decision, exercised over the same range of choices, and hence is not inequitable. (p. 20)

Viewing equity in these terms would seem on the face of it to provide a promising line of defence for weighted health maximization. Interestingly, however, Le Grand concludes that it would actually be *inequitable* for differences in health-related behaviour to have *any* bearing on the way people are treated by the health care system.

> It is arguable whether equity considerations of this kind should play any role in the actual allocation of treatment at the point of use. Rather, they seem to be more appropriately applied to questions of the *finance* of treatment. That is, the question might be not: is it equitable for these particular patients to receive treatment; but, rather, should they receive treatment at the community's expense, or at their own expense? (p. 23)

He then suggests that smokers, for example, should be charged an annual premium to cover the *expected* costs of treatment but should continue to receive the same treatment (or access to treatment) as non-smokers.

The implication of Le Grand's argument is, then, that it might well be equitable for a person's non-health characteristics to influence his rights vis-à-vis the health care sector. However, any discrimination ought to be confined to the finance of health care; discrimination in the delivery of health care is not warranted. Once again the answer to the first of the three questions posed at the beginning of the section appears to be 'no'.

6. Equality? Or less inequality?

Of the four definitions of equity examined in the previous section, only the 'equality of health' definition would appear to provide a basis for determining an equitable allocation of health care. But it is not a terribly sensible equity objective[17]: blind pursuit of equality can lead to some rather absurd

[17]Indeed many suggest that it is, in any case, rarely stated as a policy objective. On these points see inter alia Mooney (1983, 1986) and Mooney and McGuire (1987).

conclusions. It would, for example, lead to the conclusion that the move from point m in fig. 1 to point r is unjust, even though *both* A and B are in better health at r and the degree of inequality there is small.

This is *not* to say, however, that a concern about the *degree* of inequality in health is to be rejected. Indeed, it is this concern that lies at the heart of Smith's (1987) critique of the QALY literature. Williams (1981) also appeared to be sympathetic to a concern about inequality when, in a discussion about the implications of the 'egalitarian–humanitarian ethic' underlying the QALY approach, he wrote:

> It has frequently been suggested to me in casual conversation ... that distributional equity dictates that it is better for many people to get a little than for a few people to get a lot (i.e., contradicting the notion that one person getting ten years' additional life expenditure is the same as ten people getting one each). (p. 277)

If the concern is about the implications of health maximization for health inequalities is legitimate, but the goal of equality per se is unattractive, how might one proceed?

7. Is a social welfare function the answer?

One approach suggested by modern welfare economics would be to employ a *social welfare function* (SWF). Given the extra-welfarist orientation of the QALY approach, the SWF would be defined not over utility levels but over inter alia the health of the population. This SWF would be constructed so as to reflect an aversion to inequality, but would permit some trade-off between inequality and efficiency. The SWF would then be maximized subject to resource and other constraints.

7.1. The isoelastic social welfare function

To illustrate, let us stick with the case of two individuals, or groups of similar individuals, A and B. Then one attractive SWF is the isoelastic SWF:

$$W=(\tau-1)^{-1}[(\alpha h_A)^{1-\tau}+(\beta h_B)^{1-\tau}], \quad \tau \neq 1. \tag{1}$$

Here, W denotes the level of social welfare associated with the health distribution $[h_A, h_B]$.[18] The parameter α indicates the weight to be attached

[18] A similar welfare function underlies Atkinson's (1970) index of inequality [cf. also Deaton and Muellbauer (1980, p. 227 ff.)].

to A's health and β the weight to be attached to B's health. If, for example, A were a young person (or a group of young people) and B an elderly person (or a group of elderly people), society might take a view that α should be larger than β. Alternatively society might wish to adopt the principle underlying the early QALY literature, namely that the identity of the person receiving the QALY is irrelevant to the social value of that QALY. Such a view, which is in effect an *anonymity principle*, would result in α and β being set at the same value.

The parameter τ indicates the degree aversion to inequality in health outcomes. The case where $\tau=0$ and $\alpha=\beta=1$ is the case considered in the empirical literature on QALYs to date and the case where $\tau=0$ but $\alpha\neq\beta$ underlies the recent proposals of Culyer and Williams. In neither case is any aversion to inequality indicated: the difference lies in the slope of the welfare contour. If $\tau>0$, some aversion to inequality is indicated: with $\tau<\infty$, the contours of the welfare function are convex to the origin and as $\tau\rightarrow\infty$ the welfare contour becomes L-shaped with its corner on the 45° line – the so-called *Rawlsian* SWF, since it implies a concern solely with the health of the least healthy person [cf. Rawls (1972), Atkinson and Stiglitz (1987, p. 339)].[19] Though it is sometimes claimed otherwise [cf., e.g., Barr (1987, p. 152)], eq. (1) cannot encompass egalitarianism: if the goal is absolute equality, then all that matters is the distribution of health, not per capita health status; in this case all points along a given ray through the origin give rise to the same level of social welfare and the closer the ray is to the 45° line, the higher is the level of social welfare. Nor can eq. (1) capture a less extreme version of egalitarianism where some trade-off is permitted between 'distance' and the absolute levels of health: in this case the social welfare contours are upward-sloping lines with their corners on the 45° line [cf., e.g., Atkinson and Stiglitz (1987, p. 341)]. However, unless the health frontier has some upward sloping parts, the Rawlsian maximin criterion – which *can* be captured by eq. (1) – will lead to the same outcome as this family of egalitarian SWFs.

Figure 3 shows the implications of aversion to inequality for the choice of point on the health frontier. Health maximization would pick point p out as the best point, whilst the 'equality of health' objective would select point q. The non-linear SWF, by contrast, chooses point s: here there is less inequality than at p, but more inequality than at q. It is also evident that the sum of health statuses is lower at s than at p. Thus the SWF provides an answer to the question posed by Culyer (1990), namely: '...what *is* the acceptable price that one should pay for greater equality?' [Culyer (1990, p. 24)].

[19]For a defence of this interpretation of Rawls' maximin principle in the present context see Le Grand (1987).

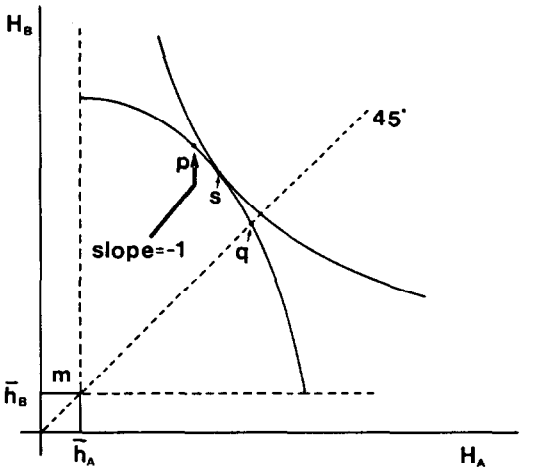

Fig. 3

7.2. Social welfare maximization and definitions of equity

The attraction of the SWF approach, then, is its ability to capture both efficiency *and* equity considerations. It provides a way of examining the extent to which society wants to accept a lower per capita health status in order to achieve greater equity. The SWF approach also provides the non-technical information required in order to determine *needs*, since it indicates which health improvements are socially desirable. Once the desired distribution of extra QALYs has been established, one can address the technical problem of establishing the least-cost treatments to be used to effect these health improvements.

This raises the issue of whether there might be some conflict between welfare maximization (which reflects one view of equity) and the notion of equal treatment for equal need (which ostensibly reflects another). Consider first the case where A and B start off with the same health (the endowment point lies on the 45° line), have the same capacity to benefit from treatment and can be treated at the same cost (the health frontier is symmetrical about the 45° line). Welfare maximization in this case requires that both A and B receive the same number of extra QALYs and the same amount of health care resources. It is clear that giving A and B *different* amounts of resources would not only violate the principle of equal treatment for equal need, but would also reduce social welfare. This is also true in the more general case where capacities to benefit from treatment differ. In such a case it is quite conceivable that welfare maximization may require that one person gets more extra QALYs than another, yet both require the same amounts of

health care resources to achieve their respective health improvements. Differential treatment would again violate the principle of equal treatment for equal need, as well as reducing social welfare. The conclusion is therefore reached that *equal treatment for equal need is actually implied by welfare maximization.*[20]

By contrast, equal access for equal need is not implied by welfare maximization if the SWF is defined over health levels. Welfare maximization identifies persons in equal need, but assigning those in equal need the same access costs will not guarantee that they receive the same treatment [cf. Mooney (1983, p. 183)]. Welfare maximization – like 'equal treatment for equal need' – may require differentiation of access costs amongst those in equal need.

8. Social welfare maximization and resource allocation

What implications does acceptance of an aversion towards inequality have for resource allocation via the QALY approach?

8.1. Resource allocation: Theory

Health maximization leads, as has been seen, to the rule that resources ought to be deployed so that the marginal cost per QALY is the same across all types of health care activity. As might be expected, social welfare maximization leads to a modification of this rule, namely that the slope of the frontier (the ratio of marginal costs) is to be equated to the slope of the SWF contour. The latter is equal to $(h_A/h_B)^{-\tau}$, where h_A and h_B are the *final* health levels of A and B (i.e., their health *levels after* receipt of any extra QALYs). The latter is the sum of (i) the expected number of QALYs remaining in the absence of treatment (the shaded area in fig. 2) and (ii) the QALYs that would be gained as a result of treatment (the cross-hatched area in fig 2). Since (i) is given, the decision variable in each case is (ii). This affects both the slope of the SWF contour (via its effect on final health) and the slope of the frontier (via its effect on marginal costs).

Note that when $\tau=0$, i.e., when there is no concern about inequality, the rule reverts to the original rule that the marginal cost per QALY is to be equalized across programmes. In this case the pre- *and* post-treatment levels of health are irrelevant to resource allocation decisions. In the case where there is some concern about inequality, $\tau>0$ and $(h_A/h_B)^{-\tau}$ is different from one. Here resource allocation decisions cannot be made simply by examining the marginal cost of additional QALYs. Consideration must also be given to

[20]One is reminded here of Feldstein's (1976) conclusion that equal taxation of equals is implied by social welfare maximization.

(i) the expected health of A and B in the absence of treatment and (ii) the degree to which society is averse to inequality.

8.2. Resource allocation: Practice

The implication of the above is that, when allocating resources between health care programmes and between individuals within a given programme, attention needs to be paid both to the degree of inequality implied by each decision and to the importance society attaches to that inequality.

How might the approach suggested be operationalized? To illustrate take the problem of deciding on resource allocation between two health care programmes. Suppose the programmes in question are hip replacements and kidney transplantation. Thus A might represent all persons with hip disease of a given type and severity, and B all persons with kidney disease of a given type and severity. Unless society is indifferent towards inequalities in health, information is required on:

(i) the prognosis (in terms of QALYs) for a person with a diseased hip in the absence of a hip replacement (the shaded area in fig. 2);
(ii) the number of people who could benefit from a hip replacement;
(iii) the prognosis (in terms of QALYs) for a person with a diseased kidney in the absence of a kidney transplant (the shaded area in fig. 2);
(iv) the number of people who could benefit from kidney transplants;
(v) society's aversion to inequalities in health (i.e. a value for τ); and
(vi) indications of how the marginal cost of a QALY in each programme changes as the number of cases treated changes.

The product of (i) and (ii) gives the *total* number of expected QALYs that would be enjoyed if hip replacements were not undertaken.

Some of this information is already available. Items (i) and (iii) in the list above have already to be calculated in QALY calculations [cf., e.g., Williams (1985)]. Two possibilities suggest themselves for items (ii) and (iv): waiting lists (a rather crude and unattractive solution) and survey data (a more promising approach, but one that may require special surveys).[21] In the long term experimental methods might be used to try to elicit values of the inequality aversion parameter τ [item (v)]. In the short term, in the absence of firm guidance from policy-makers (or the general public at large), one sensible strategy might be to calculate the optimal allocation of resources for

[21]In Britain, for example, the annual General Household Survey – the closest survey Britain has to the health interview surveys carried out in many other countries – is very limited in its information on specific health problems.

each of several different values of τ. Information on the final item (vi) is currently rather patchy (current figures correspond to a point on a marginal cost curve), but could presumably be generated, perhaps by drawing on the experience of the past and/or that of other countries.

9. Conclusions

The main negative message to emerge from the foregoing is that health maximization – even in its weighted form – fails to reflect the aversion society apparently feels towards inequalities in health outcomes. The main positive message of the paper is that this aversion *could* be incorporated into resource allocation decisions by using an appropriately specified SWF. Such an approach would allow researchers to capture *both* efficiency *and* equity considerations, by allowing them to take into account how far society is prepared to accept a lower per capita health status in order to achieve greater equality in health outcomes.

Of course, obtaining optimal resource allocations via a SWF would be more burdensome computationally than the health maximization approach. In addition to requiring more information, it would also probably necessitate the use of non-linear programming techniques. But if society *does* take the view that building aversion to inequality into resource allocation decisions is desirable, does it really make sense to continue to force QALYs into the health maximization straightjacket? Might not the extra work entailed in the SWF approach be the price that has to be paid for being in tune with society's views about equity?

References

Atkinson, A.B., 1970, On the measurement of inequality, Journal of Economic Theory 2, 244–263.

Atkinson, A.B. and J.E. Stiglitz, 1987, Lectures on public economics (McGraw-Hill, London).

Barr, N., 1987, The economics of the Welfare State (Weidenfeld & Nicolson, London).

Broome, J., 1988, Goodness, fairness and QALYs, in: M. Bell and S. Mendus, eds., Philosophy and medical welfare (Cambridge Univ. Press, Cambridge), 57–73.

Culyer, A.J., 1976, Need and the National Health Service (Martin Robertson, Oxford).

Culyer, A.J., 1980, The political economy of social policy (Martin Robertson, Oxford).

Culyer, A.J., 1989, The normative economics of health care finance and provision, Oxford Review of Economic Policy 5, 34–58.

Culyer, A.J., 1990, Commodities, characteristics of commodities, characteristics of people, utilities and the quality of life, in: S. Baldwin et al., eds., The quality of life: Perspectives and policies (Routledge, London), 9–27.

Culyer, A.J., A. Williams and R. Lavers, 1972, Health indicators in: A. Schonfield and S. Shaw, eds., Social indicators and social policy (Heinemann, London), 94–118.

Deaton, A. and J. Muellbauer, 1980, Economics and consumer behavior (Cambridge Univ. Press, Cambridge).

Department of Health and Social Security, 1980, Inequalities in health (DHSS, London).

Drummond, M.F., 1989, Output measurement for resource allocation decisions in health care, Oxford Review of Economic Policy 5, 59–74.

Feldstein, M.S., 1976, On the theory of tax reform, Journal of Public Economics 6, 77–104.
Grossman, M., 1972, The demand for health: A theoretical and empirical investigation (Columbia Univ. Press, New York).
Harris, J., 1988, More and better justice, in: M. Bell and S. Mendus, eds., Philosophy and medical welfare (Cambridge Univ. Press, Cambridge), 75–96.
Le Grand, J., 1982, The strategy of equality (Allen & Unwin, London).
Le Grand, J., 1984, Equity as an economic objective, Journal of Applied Philosophy 1, 39–51.
Le Grand, J., 1987, Equity, health and health care, in: Three essays on equity (Discussion paper WSP/23, STICERD, London School of Economics).
Lockwood, M., 1988, Quality of life and resource allocation, in: M. Bell and S. Mendus, eds., Philosophy and medical welfare (Cambridge Univ. Press, Cambridge), 33–55.
Mooney, G., 1983, Equity in health care: Confronting the confusion, Effective Health Care 1, 179–185.
Mooney, G. and A. McGuire, 1987, Distributive justice with special reference to geographical inequality in health care, in: A. Williams, ed., Health and economics (Macmillan, London), 68–81.
Pereira, J., 1989, What does equity in health mean?, Discussion paper 61, Centre for Health Economics, University of York.
Rawls, J., 1972, A theory of justice (Oxford Univ. Press, Oxford).
Sen, A.K., 1973, On economic inequality (Clarendon Press, Oxford).
Sen, A.K., 1979, Personal utilities and public judgements: Or what's wrong with welfare economics?, Economic Journal 89, 537–558.
Sen, A.K., 1980, Equality of what?, in: S. McMurrin, eds., The Tanner lectures on human values (Cambridge Univ. Press, Cambridge), 185–220; reprinted in Sen, 1982.
Sen A.K., 1982, Choice welfare and measurement (Blackwell, Oxford).
Smith, A., 1987, Qualms about QALYs, The Lancet i, 1134.
Torrance, G.W., 1986, Measurement of health state utilities for economic appraisal: A review, Journal of Health Economics 5, 1–30.
Torrance, G.W. and D. Feeny, 1989, Utilities and quality-adjusted life years, International Journal of Technology Assessment in Health Care 5, 559–575.
Williams, A., 1974a, The cost–benefit approach, British Medical Bulletin 30, 252–256.
Williams, A., 1974b, Need as a demand concept (with special reference to health), in: A.J. Culyer, ed., Economic policies and social goals (Martin Robertson, London), 60–76.
Williams, A., 1976, Cost–benefit analysis in public health and medical care: Comments on a thesis written by Bengt Jönsson, Report 1976:28, Department of Economics, University of Lund.
Williams, A., 1977, Measuring the quality of life in the elderly, in: L. Wingo and A. Evans, eds., Public economics of the quality of life (Johns Hopkins Press, Baltimore).
Williams, A., 1981, Welfare economics and health status measurement, in: J. van der Gaag and M. Perlman, eds., Health, economics and health economics (North-Holland, Amsterdam), 271–281.
Williams, A., 1985, Economics of coronary artery bypass grafting, British Medical Journal 291, 326–329.
Williams, A., 1988, Ethics and efficiency in the provision of health care, in: M. Bell and S. Mendus, eds., Philosophy and medical welfare (Cambridge Univ. Press, Cambridge), 111–126.