Qalys

John Broome*

Department of Economics, University of Bristol, Bristol BS8 1TN, UK

Qalys (quality-adjusted life years) are used to make judgements about the allocation of resources in medicine, and also about the medical care of particular individuals. This paper assesses that practice. It raises an objection to one common method of making the adjustment for quality, and it describes one serious unsolved difficulty that afflicts the use of qalys. But, subject to those qualifications, the paper tries as far as possible to provide sound theoretical grounds for the practice.

1. Introduction

In medicine, decisions have to be made between alternative courses of action: how to treat a particular patient, say, or how to allocate resources nationally between different specialities. Qalys (quality-adjusted life years) are intended to help in this sort of decision-making.

Specifically, they are intended to measure the *benefit* – or the *good*, as I shall say – that will result from each of the alternatives. The idea is that the benefit of a course of action is the extra years of life it gives people, adjusted for quality; better years count more than worse ones. In medical decisions, benefit is obviously an important consideration, but it is often not the only one. Another is fairness: when treatment is to be given some patients and denied others, to treat those whose treatment would do the most good is not necessarily the fairest thing to do. Other things being equal, for instance, treating a younger person is likely to do more good in total than treating an older one, because the younger has longer to enjoy the benefits. But if resources are concentrated on the young for this reason, that may be unfair to the old. So benefit and fairness may conflict.

Qalys are only concerned with benefit. Consequently, they cannot entirely determine which decision is the right one. The friends of qalys have not

always acknowledged this limitation [e.g. Williams (1985)], and this has exposed qalys unnecessarily to attacks from their enemies. The main objection raised against them is that their use is unfair [e.g. Harris (1987)]. Qalys certainly do not take account of fairness; they cannot be expected to. Fairness must be considered separately [see Broome (1988) and Lockwood (1988)]. Nevertheless, benefit is plainly important, so qalys have an important role open to them. This paper examines how well they can fill it. How well does the total of qalys produced by an action measure its benefit?

I shall not consider how to accommodate uncertainty in a valuation. I think an action whose results are uncertain should be valued by first fixing a value on each of its possible results, and then following the recommendations of expected utility theory. But when it comes to 'social' valuations, involving the good of more than one person, expected utility theory is controversial. In particular, it prevents one from giving value to equality in the distribution of risk between people, and on the face of it that seems unreasonable. [See, for instance, the useful example in Loomes and McKenzie (1990, p. 97).] But the issues about risk are deep and complex, and I have tried to deal with them elsewhere [Broome (1991, chs. 5, 7, and 10)]. So here I shall concentrate on valuing outcomes without uncertainty. I take this to be a necessary first step towards valuing risky actions. Expected utility theory will come into my analysis, nevertheless, because gambles are commonly thought to be a useful device for *measuring* qalys.

My conclusion will be cautiously favourable towards qalys. I shall set qalys within a well-accepted theoretical context. I shall show that their use relies on several major assumptions, some of which are at best very broad approximations, but perhaps these assumptions are acceptable as a first approach to the problem. However, qalys do run up against one large difficulty, described in section 8, which at present I can see no way around.

2. Individual preferences

Start with the simplest case. Take a choice that affects only one person: a choice between alternative treatments for a single patient. Suppose no one else is affected by the results – no relative or dependant or anyone else. This implies that, if the treatments differ in cost, the difference will be paid by the patient. The benefit of each treatment is then simply a matter of how good it is for the one person, taking into account the cost to her.

It is commonly assumed that if a person prefers one of two alternatives to the other, then that one is better for her. I call this 'the preference-satisfaction theory of good'. I shall raise a question about it later. But for the time being let us take it for granted, and see how far it can carry us. It tells us that, if we know our subject's preferences, we know what is better for her. If, say, the question is whether she should be given a treatment that relieves

her pain but shortens her life, we only need to know whether she prefers a longer life with pain or a shorter life without.

If we knew the subject's preferences between alternative treatments, therefore, there would be no need to calculate qalys. But often we do not know these preferences. The subject may be too ill to express them. She may be too young to have any. Or we may be making a decision at a general level, so we do not know which particular people will be affected; we might, for instance, be wondering which sort of treatment for a disease to concentrate resources on. In these cases, we will have to make a judgement about which of the alternatives the person prefers or would prefer. Qalys could be useful here. So this gives us a reason for analysing the structure of preferences, and its connection with qalys. The analysis will also be useful when we come to consider choices that affect the good of more than one person. The most thorough analysis I know is in Pliskin, Shepard and Weinstein (1980). But I do not find it perfectly satisfactory, for reasons given in the appendix. So my own simple analysis follows in section 3.

3. Quality adjustment factors in the representation of preferences

Suppose our subject is faced with a range of alternative lives. In each she lives for a number y of years beyond those she has already lived. (The number y varies between the alternatives.) In the first of these years the quality of her life is q_1, in the second q_2, and so on. Each life can be described by a vector of variable length $(q_1, q_2, \ldots, q_y)$. By a 'quality of life' I mean something like: confined to a wheelchair and in slight pain. A description like this could be filled out to any degree of detail. But since we are concerned with health, I mean particularly a description of the person's state of health. Imagine that other aspects of the quality of her life, such as her wealth, are held constant throughout this exercise.

Suppose our subject has preferences among these alternatives. I shall call her a *discounted-qaly maximizer* if and only if her preferences can be represented by a utility function of the form

$$V(q_1, q_2, \ldots, q_y) = v(q_1) + r_2 v(q_2) + \cdots + r_y v(q_y). \tag{1}$$

The constants $r_2, r_3, \ldots$ are her discount factors, and the subutility function v gives her quality adjustment factors $v(q)$ for each quality q. (The r's may all be one; my term 'discounted-qaly maximizer' includes people who discount at a zero rate.) Throughout this paper I shall assume that people are discounted-qaly maximizers. This is a big assumption: one of those I suggested may be acceptable as a first approach to the problem. The appendix derives it from more primitive conditions. This derivation does not justify the assumption, but it does show what it depends on. The most

dubious condition is that, in the person's preferences, qualities of life at different times are *strongly separable*. This is actually a necessary condition if the person is to be a discounted-qaly maximizer; a utility function has an additively separable form, as (1) has, if and only if the preferences are strongly separable [see, for example, Deaton and Muellbauer (1980, pp. 137–142)]. Strong separability means that a person's preferences about the qualities of her life in any particular group of years are independent of the qualities of her life in other years.

For a discounted-qaly maximizer, representing her preferences in the form (1) determines both V and v uniquely up to positive multiples. (They are 'ratio scales', that is to say.) The additively separable form of (1) determines V and v uniquely up to increasing linear transformations [see, for example, Broome (1991, p. 74)]. And the zero on the scale of adjustment factors is not arbitrary. It is assigned to the quality of life (if there is one) that the person would just as soon not live at all: that is, to the quality q^0 such that $(q_1, q_2, \ldots, q_y, q^0)$ is indifferent to $(q_1, q_2, \ldots, q_y)$ for all values of y and $q_1, q_2, \ldots, q_y$.

It is traditional to assign a factor of one to healthy life: $v(h) = 1$, where h is good health. This is arbitrary and simply sets the scale of factors. This scale makes no difference until we come to decisions that involve the good of more than one person, in section 7.

If we knew the discount factors and the function v for a person, we should be able to predict her preferences, and this would give us a basis for assessing the benefits of different treatments. I shall concentrate on the quality adjustment factors given by v. How can they be determined? In principle, a utility function of the form (1) could be fitted econometrically to a person's preferences. But this would require more data than are generally available in practice [see the informal exercise in Pliskin, Shepard and Weinstein (1980)]. So in practice more primitive methods are generally used [see Weinstein (1986)]. In this section I shall mention two. Each requires a heroic assumption to make it work.

One possible heroic assumption is that the person does not discount her future qalys: all her discount factors r_i are one. Let us call a person a *qaly maximizer* if she does not discount. Suppose a qaly maximizer is indifferent between living some number t years in good health and living ten years at some quality q. Then eq. (1) tells us that $10 v(q) = t v(h) = t$, since $v(h) = 1$. So $t/10$ gives us a measure of the quality adjustment factor $v(q)$. Here, then, is a method of estimating the adjustment factor $v(q)$: find out what value of t makes the person indifferent between t years of good health and ten years at quality q. Call this the 'time method'. Plainly it will give the right answer only if the person does not discount.

An alternative heroic assumption is that the person is *risk neutral* about discounted qalys. To state this assumption properly, we must first suppose

our subject is not only a discounted-qaly maximizer, but also conforms to expected utility theory. If she does, her preferences amongst gambles can be represented by the expected utility function

$$E(u(V(q_1, q_2, \ldots, q_y))) = E(u(v(q_1) + r_2 v(q_2) + \cdots + r_y v(q_y))), \qquad (2)$$

where E is the expectation operator and u is some increasing transformation of discounted qalys. The person is risk neutral if and only if u is linear. In that case she is an *expected-discounted-qaly maximizer*: between gambles, she always prefers the one that offers a greater expectation of discounted qalys.

Suppose a person is indifferent between, on the one hand, living ten years of life at quality q and, on the other, a gamble offering a chance p of living ten years in good health and a chance $(1-p)$ of dying immediately. Then eq. (2) tells us that

$$u(Rv(q)) = pu(Rv(h)) + (1-p)u(0) = pu(R) + (1-p)u(0), \qquad (3)$$

where R is $(1 + r_2 + \cdots + r_{10})$. If the person is risk neutral, so u is linear, it follows that $p = v(q)$. This gives us a second method for estimating the adjustment factor $v(q)$: find out what value of p makes the person indifferent in this choice. I call this the 'probability method'. It will give the right answer *only* if the person is risk neutral about discounted qalys.

A simple example will illustrate the danger of error here. Amanda is a qaly maximizer – she does not discount. For her V in (1) is simply her total qalys. But suppose she is risk averse about her total qalys. In choosing amongst gambles, she maximizes $E(u(V))$, where u is an increasing strictly concave transformation. For simplicity, let us concern ourselves only with lives whose quality is constant at q. For these lives, $V = yv(q)$. So Amanda maximizes $E(u(yv(q)))$. Let q' have an adjustment factor $v(q')$ of a half. This means that Amanda is indifferent between ten years of life at q' and five years of healthy life, both lived for sure. But suppose she is indifferent between ten years at q' and a gamble giving an 0.71 chance of ten years in good health and 0.29 chance of death. (This will be so if u happens to be the square root function.) The probability method would say the quality adjustment factor of q' is 0.71. But that would be wrong; the factor is a half. A quantity of qalys is obtained by multiplying a number of life years by a quality adjustment factor. But for Amanda, multiplying the number of her life years by 0.71, or in general by any factor obtained from the probability method, would not give the quantity of anything.

Writers on medical decision-making seem attached to the probability method, at least in principle. (It is said to be difficult to put into practice.) Weinstein (1986, p. 205), for instance, says it has high 'theoretical merit'. But

this method will work only if people are risk neutral about their discounted qalys, and on the face of it that seems an implausible assumption. It seems about as implausible as the assumption that people do not discount, which would make the time method valid. There is some empirical evidence that people are on average risk neutral [Miyamoto and Eraker (1985)], but it is too slight to place much trust in. I think we have little reason to trust the probability method.

4. A cardinal measure of good?

When a decision affects only one person, it only matters which alternatives are better or worse for the person, not how much better or worse. Only the *order* of good matters, not *amounts* of good. But we shall also need to think about decisions that affect several people. For that, we shall need to know about amounts of good for each person. (We shall also need to know how to weigh one person's good against another's, but I shall leave that question aside till section 7.) We shall need a *cardinal* scale of good. How can we find one? In particular, do a person's qalys measure her good cardinally? If one treatment produces more qalys for a given cost than another, it is often said to be better to direct resources towards the former rather than the latter, even though these treatments are for different diseases and different people [e.g. Gudex (1990)]. But this conclusion – in which different people's qalys are added and compared to assess overall good – cannot be valid unless each person's qalys measure her good cardinally. Is this a justifiable assumption? More generally, to allow for discounting, let us ask whether the discounted quantity of qalys, V in eq. (1), measures good cardinally.

The term 'cardinal' can be confusing. A utility function is often called cardinal if it is defined uniquely up to increasing linear transformations. Any additively separable utility function has this property, so the function V in (1) is cardinal in this sense. But whether V constitutes a cardinal measure of the person's good is another matter. To say V is a cardinal measure of good means it is an increasing linear transform of good. Is it?

Conventional wisdom is that this question cannot be answered by examining the structure of people's preferences. The preference-satisfaction theory of good says that a person's preferences will tell us the order of her good, but nothing about the amount of her good: they tell us when one alternative is better for the person than another, but not how much better.

For instance, think again about Amanda, who does not discount her qalys. For her, V in eq. (1) is simply a total of qalys. She is indifferent between ten years at quality q' and five years in good health, and the time method, which works for Amanda because she does not discount, tells us that her adjustment factor $v(q')$ is a half. Given that, it is tempting to assume that

good health is twice as good for Amanda as q': that, say, ten years in good health are twice as good for her as ten years at q'. After all, ten years in good health give her ten qalys, and ten years at q' give her five. But nothing in the preferences licenses this assumption. We cannot even assume that ten years in good health are twice as good for Amanda as five years in good health. Even though Amanda does not discount, we are not entitled to assume her good is proportional to the number of years she lives. We are not entitled to assume ten qalys are twice as good for her as five.

Or take Basil, who is an expected-discounted-qaly maximizer: he is risk neutral about qalys. Suppose Basil is indifferent between ten years at q'', and a gamble giving him equal chances of ten years in good health and dying immediately. Then the probability method, which works for Basil because he is risk neutral, tells us that his adjustment factor $v(q'')$ is a half. Once again we are not entitled to conclude from the preferences that ten years in good health are twice as good for Basil as ten years at q''.

All that is conventional wisdom. But some authors have taken a different view. They have thought that, although other preferences give no information about the amounts of a person's good, preferences about gambles do, at least for people who are rational. The basis of this idea is the supposition that, when faced with a choice of gambles, a rational person will always choose the alternative that has the greater expectation of good for her. Bernoulli (1738) took this view and Harsanyi [e.g. Harsanyi (1975, p. 600)] takes a similar one. I call it 'Bernoulli's hypothesis'. Now apply it to Basil. Basil is indifferent between ten years at q'' and an equal gamble on either ten years in good health or death. According to Bernoulli's hypothesis, if Basil is rational, these two alternatives must give him the same expectation of good. Therefore ten years at q'' must lie half way on his scale of good between death and ten years in good health. If we take the goodness of death to be zero, then ten years in good health are twice as good for Basil as ten years at q''.

In general, if Bernoulli's hypothesis is correct, quality adjustment factors properly derived by the probability method would provide a cardinal measure of the person's good. This thought may help to explain why medical decision theorists are attracted to the probability method. If qalys are to determine whether one use of resources is better or worse than another, they must provide a cardinal scale of good. Bernoulli's hypothesis promises to derive a cardinal scale from people's preferences alone, and it seems to imply that the right scale is to be found by the probability method.

Now, it is actually a mistake to think that Bernoulli's hypothesis supports the probability method. Whether or not this method is valid depends only on the form of a person's preferences, and Bernoulli's hypothesis has nothing to do with it. The probability method is valid if and only if the person is risk neutral about discounted qalys. If she is not, the method does not correctly

give a quality adjustment factor at all, let alone one that indicates good cardinally.

Nevertheless, Bernoulli's hypothesis can come into the argument a different way. In eq. (2), the transform u of discounted qalys V is defined as that which the person maximizes the expectation of. According to Bernoulli's hypothesis, the person maximizes the expectation of her good. Therefore, according to the hypothesis, u will measure her good cardinally: it will be an increasing linear transform of her good. Bernoulli's hypothesis implies, then, that the right cardinal measure of good is not discounted qalys V, but the transform $u(V)$. Pliskin, Shepard and Weinstein (1980) explicitly use this transform in their work. These authors are concerned only with choosing between alternative treatments for a single patient, and do not make comparisons between people. This means they themselves are not committed to the view that the transform measures good cardinally. But other authors [e.g. Torrance, Boyle and Horwood (1982)] do use the transform in comparisons between people. They thereby imply that it measures good cardinally.

However, Bernoulli's hypothesis is not very plausible, and qalys supply a good example of why not. Take Amanda once more, who maximizes

$$E(u(V(q))) = E(u(yv(q))).$$

Does u measure Amanda's good cardinally? Amanda is indifferent between ten years of life at q' and five years of healthy life. And she is also indifferent between ten years at q' and a gamble giving an 0.71 chance of ten years in good health and 0.29 chance of death. Bernoulli's hypothesis says that ten years at q' are 0.71 times as good for Amanda as ten years in good health. Is this plausible? Surely not. Surely it is more plausible to think that ten years at q' are half as good for her as ten years in good health, since ten years at q' are equally as good as five years in good health, and since Amanda does not discount her qalys. Surely the most plausible explanation of why she requires a chance as high as 0.71, rather than 0.5, to accept the gamble I described is that she is risk averse about her good, and inclined to avoid gambles. This conclusion is not forced on us by the form of Amanda's preferences, but it simply seems more plausible for her than Bernoulli's hypothesis does. It is much less plausible to suggest that ten years at q' are 0.71 times as good as ten years in good health. It is much more plausible that V measures Amanda's good cardinally than that u does.

I do not think, therefore, that we should rely on Bernoulli's hypothesis to determine a cardinal scale for good. [There is a discussion of this hypothesis in Broome (1991, pp. 142–148 and 213–222)]. I think we should stick to the conventional wisdom that preferences alone cannot give us a cardinal scale. Consequently, to find one, we shall need to give some independent consider-

ation to the structure of a person's good. This is a matter for ethics and not decision theory.

5. Qalys as a measure of good

I have examined in detail the structure of a person's good elsewhere [Broome (1991)]. In this paper I shall not try to offer a sophisticated analysis, but simply mention two assumptions that could serve to support the use of qalys as a cardinal measure of good. I do not insist on the truth of these assumptions, but I hope, once again, that they may be acceptable as a first approach to the question. At least they indicate the type of assumptions that are needed. They are ethical assumptions, not assumptions about the form of preferences.

Let us assume, first, that the goodness of a life for a person is the total of the good it brings her at each of the times in her life. One good reason to doubt this assumption is that, because it counts the *total* of good only, it gives no value to evenness in the distribution of good through life. It might reasonably be thought that a uniformly good life is better (or perhaps worse) than a life with ups and downs but the same total. On the other hand, one feature of the assumption is surely indubitable: in some way or other, the good that comes to a person at each of the different times in her life will enter into determining her overall good. Her overall good, that is to say, is a *function* of her good at different times (and perhaps of other things too). A second feature is more controversial: this function is *symmetrical*. But this I am willing to defend. We are concerned with the goodness of the person's life as a whole, and from the standpoint of a life as a whole all times must count equally. The good at one time must count in exactly the same way as the good at another. Both are equally much the person's good. For instance, in determining the overall goodness of a life, good that comes later in the life cannot count differently from good that comes earlier.

And let us assume, second, that the goodness of a person's life at any particular time depends only on the quality of her life at that time. In section 3, I explained the notion of a quality of life by means of an example: confined to a wheelchair and in slight pain. A description like this could be filled out to any degree. Presumably if it was filled out enough, it would encompass all aspects of the person's good, so it could not fail to determine how good her life is at the time. But I deliberately restricted the notion of quality to states of the person's health, and a description of her health alone will leave out many aspects of her good. I said that other aspects, such as her wealth perhaps, are supposed to be held constant. Given that they are, the quality of her life, even though it is defined in terms of health only, will determine how good her life is.

However, some other aspects of the person's good cannot be held constant,

because of the nature of the question we are considering. One is the person's age. And perhaps a given quality of life (in terms of health) is better for a person at one age than at another. Perhaps good health enables you to enjoy life more when you are young than when you are old. Another aspect that cannot be held constant is length of life. And perhaps some qualities are better for a person if her life is going to be short, and others if it is going to be long. Perhaps the chance of being active is more important in a short life, and absence of pain in a long one. A third aspect that cannot in practice be held constant is the person's wealth. Treatments have costs, which are often borne by people other than the patient, by taxpayers let us say. Paying the costs is bad for these people, and it diminishes the true quality of their lives. But it does not affect their health, and it will not show up in a notion of quality that is restricted to states of health. To deal with this problem, in theory we need a broader notion of quality, and in practice we need a way of comparing the value of wealth against the value of health. But in the meantime we can set the problem aside by confining ourselves to 'cost-effectiveness analysis': we can compare the improvements to people's health that can be achieved by different methods at a given cost. In this way we can hold the costs to the taxpayer constant and allow ourselves to compare states of health only. This will not help us to decide how much money should be spent on the health service, but it will help us to decide the best allocation of whatever money is spent.

Granted the two assumptions I have made, it follows immediately that the goodness of an outcome for a person is the number of life years it brings her, adjusted for quality. The quality of a particular year of life determines how good that year is for her. Adding up these amounts gives us the total of good in her life, and I have assumed that is equivalent to how good the outcome is for her. The appropriate quality adjustment in this calculation is given by how good the quality is for the person. Future years are not discounted.

6. Quality adjustment factors in the measure of good

Given the assumptions of section 5, we have found that quality-adjusted life years are the correct measure of a person's good. There remains, however, the problem of making the right adjustment for quality. The factor $v(q)$ I introduced in section 3 was defined in the course of representing a person's preferences by a utility function. Now we need something different: a measure of how good a particular quality of life is for a person.

Nevertheless, we can hope that the right adjustment factors might be derivable from preferences all the same. It depends on how a person's preferences are related to her good. We shall have to make an assumption about that. And I now want to disagree with the preference-satisfaction

theory. What a person prefers does not necessarily coincide with what is good for her. This is a commonly accepted proposition. It is commonly accepted that people discount their future good in forming their preferences. They sacrifice a greater amount of good coming in the further future for the sake of gaining a smaller amount in the nearer future. Parfit (1984, ch. 14) argues that this may be rational. Pigou (1932, pp. 24–26) thought it was not. But, rational or not, a person who discounts her good is not maximizing her good. Good that comes to her later cannot count less in her overall good than good that comes to her earlier. So, of two alternatives, she sometimes prefers the one that is less good for her.

However, though I doubt the full-blooded preference-satisfaction theory, one part of it may be a good enough approximation for our purposes. A person's preferences may not coincide with her good across time, but at a single time they may be close enough. It may be that, when a person forms her preferences amongst alternative lives of various qualities, she maximizes a discounted total of her good. Her good in any year is determined by the quality of her life in that year. So, in the notation of section 3, she maximizes

$$D(q_1, q_2, \ldots, q_y) = g(q_1) + \rho_1 g(q_1) + \cdots + \rho_y g(q_y),$$

where the ρ's are discount factors, and $g(q)$ is the goodness for the person of quality q. But this is a representation of the person's preferences in the same additively separable form as eq. (1). Under this form, the subutilities v are determined uniquely up to a positive multiple. So v must be a multiple of g. This is the conclusion we have been looking for since section 4. The quality adjustment factor $v(q)$, which can be determined from the person's preferences, is a positive multiple of the person's good. It is therefore an increasing linear transform of her good. That is to say, it supplies a cardinal representation of her good.

On the other hand, even this may be assuming too close a connection between preferences and good. I am impressed by a point made by Culyer (1990), that using qalys does not commit one to a narrow – he calls it 'welfarist' – conception of good. Qaly analysis assigns values to states of health, and leaves it open whether these values are determined by how people feel when they are in these states, by their preferences about them, or perhaps by some objective principles. All of these possibilities are consistent with the general idea I started with: that qalys are aimed at assessing good or benefit. We must simply allow for alternative conceptions of good.

If the adjustment factors are to be severed from preferences, then neither the time method nor the probability method can determine them. One other method is popular. It is simply to ask people how good the alternatives are. Call this the 'direct method'. Here is an example of the sort of question that may be asked [Torrance, Boyle and Horwood (1982, p. 1051)]:

The subject was asked to imagine being in these situations for a lifetime with everything else normal, or average. S/he was asked to place the most desired level ... at 100, the least desirable ... at 0, and the others in between in order of desirability, with ties allowed, and spaced such that the relative distance between the levels corresponds to her/his feelings about the relative differences in desirability.

If questions like this elicit sensible answers, they will do so whether or not the subject discounts or is risk neutral. In some ways, therefore, this could be a more reliable way of estimating the adjustment factors than either the time method or the probability method. (Oddly enough, Torrance, Boyle and Horwood, having obtained a scale of adjustment factors by the direct method, then choose to convert it by a formula to a different scale: one that, they believe, would have been obtained by the probability method. Such is the magnetism of the probability method.) However, the direct method is quite different from the others in that it asks for a judgement about goodness, rather than a preference. How much trust we can put in the method depends on the subject's qualifications for making the judgement.

7. Comparisons between people

Now I come to the question of aggregation and comparing the good of different people. Once more for the sake of a first approach to the problem, let us adopt the utilitarian principle that the goodness of an action (the choice of a particular treatment programme, say, or a particular allocation of resources) is the total resulting good of the people. I have already identified each person's good with her qalys. So the total good of the people is the total of their qalys. One action is better than another if and only if it leads to more qalys. This is the fundamental precept that guides the practical use of qalys. We have by now found some basis for it. However, there are still some things to worry about.

The first is a doubt about the utilitarian principle I assumed. This principle is only one part of utilitarianism, and it does not commit me to the rest. For one thing, it permits the broad conception of good I mentioned at the end of section 6, and it does not insist that quality adjustment factors should be derived from preferences. But it does rule out the egalitarian view that, for a given total of good, it is better to have this total more, rather than less, equally distributed. My own view [Broome (1991, ch. 9)] is that the value of equality is best understood differently, in a way that is consistent with the utilitarian principle. It is certainly true that qalys do not give value to equality, but they should not be expected to. Equality is an aspect of fairness, and I said in section 1 that fairness needs to be considered separately.

A second reason to be careful is that we are now putting different people's qalys together: we are making interpersonal comparisons of good. So far, I have argued that for a single person a qaly always represents the same amount of good; that is what it means to say that qalys measure good cardinally. Now we are assuming that a qaly to one person represents the same amount of good as a qaly to another. To put it differently: the same distance on different people's scales of quality adjustment factors always represents the same amount of good. The nought on each person's scale is the quality that makes life just not worth living. The one is good health. The assumption is that the distance between these two represents the same amount of good for each person.

This is quite implausible. The nought represents the same level of good for everyone, because it is the level that is just not good at all. But good health is plainly not equally good for everyone. Good health is only a state of good *health*, and nothing else in one healthy person's life may be good, whereas everything else in another's may be. Qalys to one person will represent more good than qalys to another. Prolonging the life of, say, a happy person will do more good than prolonging the life of an unhappy one.

This point has disturbing implications, and I sympathise with Torrance (1985, p. 17) when he recommends qalys because they are 'egalitarian within the health domain'. Qalys treat the gap between nought and one as the same for everyone, and this, Torrance says, means that 'each individual's health is counted equally'. However, looked at one way, what Torrance says is incorrect. Restoring an old person to health produces fewer qalys then restoring a young person to health, because the old person has fewer years to live. So, in a way, qalys count the health of old people for less than the health of young people. Torrance hopes to achieve fairness by using a particular scale for estimating benefit, but actually the scale he recommends, which values a year of healthy life equally for everyone, can defensibly be claimed to be unfair to the old. I think it is better to separate the question of good from the question of fairness, rather than try to adjust the scale of good to take account of fairness [see Broome (1988)]. The assumption that healthy life is equally good for everybody is just false.

8. Problems of existence

Now I am going to mention what I think is the most serious difficulty over using qalys to measure the goodness of alternative actions. I can do no more than mention it; it is examined in detail in Broome (1985).

There are two ways of bringing it about that years of good life are lived – of producing qalys, that is. One is to prolong a person's life or make it better. The other is to bring into existence a new life. A decision made in medicine will often do both. For instance, if a child is saved she will

probably later have children herself, who will enjoy good lives. What is to be done about this? Should one give equal value to qalys brought about by either method? Should they have a different value? Or what?

Traditionally, qalys brought about by creating a new life are not counted at all. This traditional procedure seems intuitively natural, but it encounters two problems, one practical and the other theoretical. The practical problem is that it leads to an anomaly at the borderline between creating life and prolonging life. A study by Boyle et al. (1983) attaches a high value to saving the life of a prematurely-born baby, because if the baby survives she will gain a whole lifetime of qalys. It seems a little odd that saving a baby should be valued so much higher than, say, saving a twenty-year-old. Kuhse and Singer (1988), commenting on this study, point out how particularly odd it would be unless a similar high value is attached to the life of an unborn foetus. But it is not at all clear how the traditional procedure should be applied to a foetus. It matters crucially in this procedure whether an action counts as prolonging the life of an existing person, or as bringing about the existence of a new person. So it matters crucially when a person comes into existence. Once she exists, all her future qalys will count; up till then, none of them. But the beginning of a person seems inherently vague, so it seems wrong to attach great importance to the moment when it occurs.

The theoretical problem is to find a sound justification for the traditional procedure in the first place. Philosophical support for it can be drawn from an argument of Narveson's (1967). Narveson argues that a benefit has to be a benefit to somebody, and that a person is not benefited by being brought into existence, even if her life is a good one. If, therefore, an action brings it about that someone exists who would otherwise not have existed, that person's well-being is not a benefit arising from the action. This is intuitively plausible, and it surely expresses the intuitive attraction of the traditional procedure. The qalys of new people are traditionally not counted in the calculations because they seem not to represent a benefit to anyone. Narveson's argument leads us to the following principle for evaluating two alternative actions: the better action is the one that is better for those people who will exist whichever action is done. This principle would support the traditional procedure. Unfortunately, however, it turns out to be unacceptable. I explained why in Broome (1985), taking my lead from Parfit (1984, Part 4). The most serious objection is that one can find examples of three alternatives A, B and C, where the principle says A is better than B, B better than C and C better than A. This is a logical contradiction [Broome (1991, pp. 11–12)]. A principle that implies a contradiction cannot be correct.

I think it may be possible to find a philosophically defensible way around the practical problem [see Broome (1985)]. But I suspect the theoretical problem is insoluble. I suspect the traditional procedure has no sound justification.

An alternative is to count in favour of an action all the qalys the action brings about, including those enjoyed by people it brings into existence. The value of saving a person, for instance, would include all the qalys of her descendants. This is the procedure that would be recommended by 'classical' utilitarianism, which evaluates an action by the total of all the good enjoyed by everyone who will live if that action is done. Classical utilitarianism has plenty of problems of its own, some of which are mentioned in Broome (1985). Since I have never seen it recommended for medical decision-making using qalys, I shall not dwell on these problems here.

For decisions that effect which people exist, no principles of evaluation have been found that are free from problems. Consequently, we have no unproblematic way of using qalys in those medical decisions that have such effects. This is a large fraction of all medical decisions. Moreover, the difficulties may spill over into other medical decisions too. If we doubt there should be a large difference between the value of saving a premature baby and the value of saving a foetus, that may make us doubt that the value of saving a baby is really all the qalys in the rest of her life. It may cast doubt on our whole way of using qalys.

9. Summary

I think that qalys have an important contribution to make to medical decision-making. When there is a choice to be made, qalys are in principle a measure of the benefits each of the alternatives will bring. The benefits are not everything that matters, and qalys do not take account of other considerations such as fairness. Furthermore, there are many minor and major assumptions implicit in the use of qalys to measure good. Many are implausible, and we must therefore be cautious. But qalys could be useful nonetheless.

However, there are some common misconceptions about how quality adjustment factors should be calculated. I think the popularity of what I called 'the probability method' is misplaced.

Finally, there are the intractable problems mentioned in section 8. These are serious and fundamental. They afflict the whole of decision-making in matters of life and death, and they remain unsolved.

Appendix

A.1. Derivation of the discounted-qaly maximizing eq. (1)

I shall work with continuous time; I used discrete time in the text simply for the sake of a simple presentation.

Take a person, and let Y be the greatest number of years she could live.

Let y be the number she actually does live. Let Q be the set of qualities of life. Let $q(t) \in Q$ be the quality of the person's life t years after her birth. So q is a function from $\{t : 0 \leq t \leq y\}$ to Q; I shall call q a 'quality function', and say it 'lasts' y years. The pair $\langle y, q \rangle$ defines a life. Assume that the person has a preference relation amongst such pairs.

Call a life of Y years a 'full life'. For a moment, think only of the preference relation amongst full lives. This may be thought of as a relation amongst those functions q that last Y years. The model we have now, confined to full lives, is an exact analogy of the model of uncertainty in Savage's (1972) version of expected utility theory. Times t in $\{t : 0 \leq t \leq Y\}$ are analogous to Savage's 'states'; qualities in Q to his 'consequences'; and quality functions q (lasting Y years) to his 'acts'.

Savage supposes there is a preference relation amongst acts, and he imposes seven postulates on it [Savage (1972, endpapers)]. Analogous postulates may be imposed on the preference relation amongst full lives, and they make good sense if they are. I shall not spell them out here. But Postulate 2, the 'sure-thing principle' [Savage (1972, p. 23)], needs mentioning because it is the most dubious. Re-interpreted, it is an assumption of strong separability between times. It says that the value the person attaches to the quality of her life during any particular period is independent of its quality at other times.

From his postulates, Savage deduces expected utility theory. The analogous postulates will imply the analogue of expected utility theory. That is: there will be a discount function r (playing the role of a probability function) defined on times, and a subutility function v defined on qualities, such that preferences are represented by

$$\int_0^Y r(t) v(q(t)) \, dt.$$

The function v is unique up to increasing linear transformations.

Now take a life $\langle y, q \rangle$ of any length y. Consider extending this life to Y years by adding a period of life of a constant quality q^0. That is to say, consider the life $\langle Y, q^+ \rangle$, where q^+ is defined by $q^+(t) = q(t)$ for $0 \leq t \leq y$ and $q^+(t) = q^0$ for $y < t \leq Y$. Now assume there is a quality q^0 such that, for all y and for all quality functions q that last y years, the person is indifferent between $\langle y, q \rangle$ and the extended life $\langle Y, q^+ \rangle$ I have just described. I shall call q^0 the *valueless quality*. This assumption, then, is that there is a quality of life that the person would just as soon not live as live, and furthermore that this quality is the same however long she has already lived, and whatever her life has so far been like. Put briefly: there is a quality of life that is *always valueless*.

The utility of the extended life $\langle Y, q^+ \rangle$ is, by the formula above:

$$\int\limits_0^y r(t)v(q(t))\,\mathrm{d}t + v(q^0)\int\limits_y^Y r(t)\,\mathrm{d}t.$$

Since v is unique only up to linear transformations, we may set $v(q^0)=0$. This now defines v uniquely up to multiples. Then the utility of $\langle Y,q^+\rangle$ is

$$\int\limits_0^y r(t)v(q(t))\,\mathrm{d}t.$$

And since, by our assumption, $\langle Y,q^+\rangle$ is indifferent to $\langle y,q\rangle$, this may be taken as the utility of $\langle y,q\rangle$. This is the continuous-time version of the discounted-qaly formula, eq. (1) in the text.

In summary, the following conditions are together sufficient for a person to be a discounted-qaly maximizer:

(1) The person has preferences that conform to the analogues of Savage's postulates.

(2) There is a quality of life that is always valueless for her.

A.2. Comments on Pliskin, Shepard and Weinstein (1980)

Pliskin, Shepard and Weinstein consider lives described by pairs $\langle y,q\rangle$, where y is the length of the life and q its quality. For them, q is not a function but a single quality, which is assumed constant throughout the life. They make two assumptions about a person's preferences over lotteries for such lives. The first is that they are 'mutually utility independent'. This means that the person's preferences about lotteries over qualities q, with y held constant, are independent of the constant value of y, and similarly that her preferences about lotteries over y are independent of q. The other assumption is that the preferences have 'constant proportional tradeoff'. This means that 'the proportion of remaining life years that one is willing to give up for an improvement in health status from any given level q_1 to any other level q_2 does not depend on the absolute number of remaining life years involved' (pp. 210–211). Pliskin, Shepard and Weinstein show that these two assumptions imply, amongst other things, that the person is an undiscounted qaly maximizer.

I find this analysis unsatisfactory for three reasons. [There is another good critique in Loomes and McKenzie (1990).] First, it offers no analysis of preferences over lives whose quality is not constant. Second, the assumption of mutual utility independence is hard to judge because it requires separability between lotteries. Compare my own assumption that qualities are separable between times. An assumption that *lotteries* over qualities are

separable between times would be much stronger. Indeed, by a theorem of Gorman's (1968), it would imply risk neutrality about qalys, which is surely implausible. This suggests to me that one ought to be cautious about assuming separability of lotteries. My own assumption that qualities are separable does much the same job better. Uncertainty is a complication rather than an essential part of the problem of valuing lives, and it ought not to be introduced into the analysis earlier than it need be. My third and most important objection is that constant proportional tradeoff is a highly specific assumption, which is out of place at the level of general theory. Furthermore, it is implausible anyway. It rules out any discounting of future qalys, and discounting is commonly taken to be a fact of life.

References

Baldwin, S., C. Godfrey and C. Propper, eds., 1990, Quality of life: Perspectives and policy (Routledge, London and New York).

Bell, J.M. and S. Mendus, eds., 1988, Philosophy and medical welfare (Cambridge University Press, Cambridge and New York).

Bernoulli, D., 1738, Specimen theoriae novae de mensura sortis, Commentarii Academiae Scientiarum Imperialis Petropolitanae 5. Translated by Louise Sommer as: Exposition of a new theory on the measurement of risk, Econometrica 22 (1954), 23–36.

Boyle, M.H., G.H. Torrance, J.C. Sinclair and S.P. Horwood, 1983, Economic evaluation of neonatal intensive care of very-low-birth-weight infants, New England Journal of Medicine 308, 1330–1337.

Broome, J., 1985, The economic value of life, Economica 52, 281–294.

Broome, J., 1988, Good, fairness and qalys, in: J.M. Bell and S. Mendus, eds., Philosophy and medical welfare (Cambridge University Press, Cambridge and New York) 57–73.

Broome, J., 1991, Weighing goods (Blackwell, Oxford and Cambridge Massachusetts).

Culyer, A.J., 1990, Commodities, characteristics of commodities, characteristics of people, utilities, and the quality of life, in: S. Baldwin, C. Godfrey and C. Propper, eds., Quality of life. Perspectives and policy (Routledge, London and New York) 9–27.

Deaton, A. and J. Muellbauer, 1980, Economics and consumer behavior (Cambridge University Press, Cambridge and New York).

Gorman, W.M., 1968, The structure of utility functions, Review of Economic Studies 35, 367–390.

Gudex, C., 1990, The qaly: how can it be used?, in: S. Baldwin, C. Godfrey and C. Propper, eds., Quality of life: Perspectives and policy (Routledge, London and New York) 218–230.

Harris, J., 1987, Qalyfying the value of life, Journal of Medical Ethics 13, 117–123.

Harsanyi, J.C., 1975, Can the maximin principle serve as a basis for morality? A critique of John Rawls's theory, American Political Science Review 69, 594–606.

Kuhse, H. and P. Singer, 1988, Age and the allocation of medical resources, The Journal of Medicine and Philosophy 13, 101–116.

Lockwood, M., 1988, Quality of life and resource allocation, in: J.M. Bell and S. Mendus, eds., Philosophy and medical welfare (Cambridge University Press, Cambridge and New York) 33–56.

Loomes, G. and L. McKenzie, 1990, The scope and limitations of qaly measures, in: S. Baldwin, C. Godfrey and C. Propper, eds., Quality of life: Perspectives and policy (Routledge, London and New York) 84–102.

Miyamoto, J.M. and S.A. Eraker, 1985, Parameter estimates for a qaly utility model, Medical Decision Making 5, 191–213.

Narveson, J., 1967, Utilitarianism and new generations, Mind 76, 62–72.

Parfit, D., 1984, Reasons and persons (Oxford University Press, Oxford and New York).

Pigou, A.C., 1932, The economics of welfare, 4th edn. (Macmillan, London).

Pliskin, J.S., D.S. Shepard and M.C. Weinstein, 1980, Utility functions for life years and health status, Operations Research 28, 206–224.

Savage, L.J., 1972, The foundations of statistics, 2nd edn. (Dover, New York).

Torrance, G.W., 1985, Measurement of health state utilities for economic appraisal, Journal of Health Economics 5, 1–30.

Torrance, G.W., M.H. Boyle and S.P. Horwood, 1982, Application of multi-attribute utility theory to measure social preferences for health states, Operations Research 30, 1043–1069.

Weinstein, M.C., 1986, Risky choices in medical decision making: A survey, The Geneva Papers on Risk and Insurance 11, 197–216.

Williams, A., 1985, The value of qalys, Health and Social Service Journal (18 July).