LEADING ARTICLE

Pharmacoeconomics 2001; 19 (4): 323-335
1170-7690/01/0004-0323/$22.00/0

On Individual Preferences and Aggregation in Economic Evaluation in Healthcare

Bengt Liljas[1,2] and *Björn Lindgren*[1,3,4]

1 LUCHE (Lund University Centre for Health Economics), Lund, Sweden
2 Program on Economic Evaluation of Medical Technology, Department of Health Policy and Management, Harvard School of Public Health, Boston, Massachusetts, USA
3 Department of Economics, Lund University, Lund, Sweden
4 Department of Community Medicine, Lund University, Malmö, Sweden

Abstract

For practical reasons, in order to carry out economic evaluations of collective decisions, total costs will generally be compared with total benefits; hence, individuals' willingness to pay (WTP) or quality-adjusted life-years (QALYs) have to be estimated at an aggregate level. So far, aggregation has usually been done by taking the individuals' mean WTP or the unweighted number of QALYs. Since the aggregation process is closely related to the way that income, health and/or utility of different individuals are compared and weighted, it also has significant equity implications. Thus, the explicit (or, more often, implicit) assumptions behind the aggregation process will largely affect how health and welfare are distributed is society. The aggregation problem in economic evaluation is certainly not trivial, but is seldom addressed in current practice.

This paper shows the underlying assumptions of aggregate cost-benefit analysis (CBA) and cost-effectiveness analysis/cost-utility analysis (CEA/CUA), and it emphasises the particularly strong assumptions which have to be made when QALYs are interpreted as utilities in the welfare economics sense. Naturally, the appropriate method to choose depends on what is to be maximised: welfare or health. If decisions of resource allocation are to be based on economic welfare theory, then CBA should be preferred. However, if QALYs are interpreted as measures of health, rather than as utilities, then CEA/CUA would be appropriate.

Every time an individual purchases a good produced for the market, he or she informs the market that he or she is willing to pay at least the price of the good. Such information is repeatedly produced by the day-to-day behaviour of large numbers of consumers. It would certainly be possible, of course, at least in principle, to make evaluations of collective decisions in the same type of manner, calculating gains and losses for every individual affected by the collective decision. However, it would probably be prohibitively expensive. Thus, for practical reasons, in order to carry out economic evaluations of collective decisions, total costs will generally be compared with total benefits. Hence, individuals' willingness to pay (WTP) or quality-adjusted life-years (QALYs) have to be estimated at an aggregate level.

The aggregation problem in economic evaluation is not trivial; however, it is seldom addressed in current practice. Since the aggregation process

is closely related to the way that income, health and/or utility of different individuals are compared and weighted, it also has significant equity implications. Thus, the explicit (or, more often, implicit) assumptions behind the aggregation process will largely affect how health and welfare are distributed in society. The aim of this study is therefore to survey the underlying assumptions of aggregate cost-benefit analysis (CBA) and cost-effectiveness analysis or cost-utility analysis (CEA/CUA). Doing so, we especially examine the interpretation of QALYs as preference-based utility measures. With this delimitation we do not, however, imply that economic evaluations and QALYs need to be based on economic (welfare) theory. However, when such claims are being made, we strongly believe that it is important to be aware of the underlying assumptions of those claims. Other foundations than economic welfare theory exist, such as the extra welfarist approach and the decision-maker's approach, but these will not be discussed at any great length here. The readers are instead referred to the articles by, for example, Culyer,[1] Nord et al.,[2] and the review by Brouwer and Koopmanschap,[3] and the references therein. We concentrate on the principle, theoretical issues involved rather than the practical, empirical problems. Although this is a somewhat technical area, we have attempted to keep the presentation as nontechnical as possible, even though we thereby risk losing some exactness.

Section 1 outlines the historical development of WTP and QALYs. In section 2, WTP and QALYs are related to individual preferences and, by summarising existing literature, it is shown under what assumptions QALYs can be interpreted as an individual utility function. Section 3 is devoted to principles for aggregating individual WTPs and QALYs; it introduces the concept of the social welfare function (SWF) and derives the necessary conditions behind aggregating WTPs and QALYs. Section 4 holds a closing discussion, also containing some remarks on the empirical problems.

1. Historical Development of Willingness to Pay (WTP) and Quality-Adjusted Life-Years (QALYs)

CBA was introduced as a tool for improving decision-making in situations where markets fail, e.g. because of the existence of public goods, external effects or natural monopolies, and/or where markets have become politically excluded as instrument for the allocation of scarce resources i.e. in the public sector.[4,5] Large scale projects, such as the building of bridges or airports, were among the first investments to be assessed (see, for example, the survey presented in Warner and Luce[6]). The purpose in these analyses was to simulate market information by calculating all expected benefits of a project and all expected costs. In principle, the CBA is based on economic theory, and benefits are typically being estimated in monetary terms (directly or indirectly) as the individuals' maximum WTP. Societal welfare is maximised by undertaking those interventions only where the net present benefits minus the net present costs are larger than zero.[7]

WTP as a measure of the benefits of a project has a long history and is used in many other fields of applied economics, especially in the environmental field (see Mitchell and Carson[8] for a review of empirical WTP studies). In the health area, pioneering work was undertaken by, *inter alia*, Weisbrod[9] and Klarman.[10] However, many non-economists seem to have resisted placing monetary values on the benefits of healthcare, and relatively few WTP studies have been performed in the health area.[11-13] Thus, in order to be able to communicate with the healthcare community, alternative methods of evaluating benefits have been derived, the most prominent one being QALYs. The QALY approach typically produces a summary measure of the length of life and the quality of life for each year of life, where quality of life, ranging from 0 to 1, is assumed to be independent of when and where the relevant life-year is placed in the individual's life span. This concept was developed in the late 1960s and early 1970s.[14-16] In their much cited paper, Weinstein and Stason[17] argued that the QALY

was the most relevant outcome measure for evaluating healthcare interventions. Economic evaluation using QALYs as an outcome measure is nowadays mostly known as CUA in Europe, while in the US it has remained a subset of CEA.[18]

Early researchers do not seem to have been very concerned with the relation between the outcome measure and economic (welfare) theory or individual preferences. Instead they seem to have used a rather straightforward operations research or system analytical approach. In fact, several of the early papers on the QALY concept were typically published in operations research journals such as *Management Science* or *Operations Research* rather than in economic journals. Changes in health were considered to be the output of the healthcare sector, and scarce healthcare resources were to be allocated in order to maximise total health gains. In principle, the necessary condition for optimum would then be that the marginal cost per QALY gained is the same in all healthcare activities.[19] An exception would be when there are some restrictions, for instance the total number of kidney donors in renal disease.

Naturally, health and welfare are related concepts, but for most people health is but one factor, although an especially important one, influencing the welfare of the individual. Furthermore, the impact on welfare from a change in health may depend, *inter alia*, on the income, age, gender, family situation, attitudes to risk-taking and time preferences of the individual. Moreover, individuals may not be indifferent between healthcare measures producing the same health gain, because of other characteristics of the healthcare services that the individual may find positive or negative.[20] However, even though the early researchers did not claim that the concept had any clear theoretical basis, but considered QALYs to be a rather practical measure of health, suitable for planning purposes,[19] several writers since then have had the ambition to derive under which assumptions QALYs might be interpreted in terms of individual preferences for health rather than as an 'objective' measure of health *per se* (see section 2.3). When used in economic evaluations, this subjective interpretation of QALYs would then focus on the maximisation of economic welfare rather than total health gains. The solution of the 2 maximisation problems (i.e. QALYs interpreted as welfare versus health) will yield the same solution only under rather implausible assumptions. As long as there is a trade-off between health and other components of the individual's utility function (which there is according to economic theory), these assumptions are violated, and the solutions are thus different.

2. WTP, QALYs and Individual Preferences

2.1 Ordinal and Cardinal Utility

In neoclassical economic theory, individual preferences are assumed to be completely subjective. Furthermore, for most analytical purposes, preferences only have to be ordinal. Thus, the individual is assumed to be able to rank bundles of goods according to their desirability, but not to express how much more he or she prefers one bundle to another. An individual's preferences can then be represented by an ordinal utility function.

However, when the individual's decision, for example, involves risk, which is one of the fundamentals for the QALY model (see section 2.3), a stronger assumption than 'ordinality' is needed. When preferences are cardinal, the individual is assumed to not only be able to state which bundle of goods that he or she prefers, but also how much more desirable one bundle of goods is in comparison with another (i.e. differences in desirability). A cardinal utility function meets these requirements.[21,22] Still, preferences are assumed to be subjective and utilities not comparable among individuals (a condition which will be relaxed in section 3.3 on aggregating individuals' QALYs).

2.2 WTP and Individual Preferences

From the very beginning the concept of WTP was derived from economic welfare theory and, hence, from the assumptions about individual preferences usually made in the economic theory of

individual behaviour. The underlying assumptions of the individual's preferences (see Varian,[23] for a formal presentation) require that they are: (i) complete (any commodities can be compared); (ii) transitive (if x is preferred to y, and y is preferred to z, then x is also preferred to z); (iii) continuous (if y is preferred to z and x is close enough to y, then x is also preferred to z); (iv) monotone (at least as much of every commodity, and more of one, is better); and (v) convex (averages are preferred to extremes i.e. the more the individual has of x, the less willing he or she is to give up y for additional x).

The individual's utility function based on individual preferences can be defined broadly here, and is in this context usually assumed to depend on the prices of commodities, income and the level of health. In this case, the welfare change of a new policy (that, for example, affects the individual's level of health) is simply the difference in the (ordinal) utility between the state before and after the policy change. If this difference is positive, the change in policy has been beneficial to the individual. This is, of course, a very general measure and does not give decision-makers a very useful guide. A monetary measure of this change would be more helpful. One such measure is the compensating variation, which is defined as the maximum amount of money that can be taken from an individual (or the maximum amount that he or she would be willing to pay) after such a change, while leaving him or her as well off (i.e. having the same utility) as before the change.

2.3 QALYs and Individual Preferences

As emphasised previously, the QALY approach was introduced in a rather pragmatic way, without any formal deduction from reasonable assumptions of individual preferences. Ever since it was developed in the late 1960s, several attempts have been made to examine under which assumptions about individual preferences QALYs may be based in common neoclassical economic utility theory.

Using some simplifying assumptions, Pliskin et al.[24] derived under which conditions QALYs are a valid cardinal utility function. In doing so, they first accepted the assumptions of expected utility theory.[25] These assumptions, called the von Neumann-Morgenstern axioms (see, for example, Baumol[26] for a formal presentation), are: (i) independence (lotteries with indifferent prizes are indifferent); (ii) desire for high probability of success (the lottery with the greater probability of a favourable outcome is preferred); and (iii) compound probabilities (whether an outcome of a lottery results in the participation in yet another lottery does not matter – only the net probabilities of receiving prizes matter). The expected utility function resulting from these assumptions is a linear combination of the uncertain outcomes and the associated probabilities of achieving them.

Pliskin et al.[24] then assumed that health states worse than death do not exist. Several empirical studies have shown, however, that for many people there are, indeed, health states worse than death (see, for example, Kind et al.[27]). Utility was assumed to only depend on health and the number of life years. Furthermore, health status was constant over time (Broome,[28] however, generalised the model to include variable health states without further restrictive assumptions), and there was no discounting of QALYs (even though Broome[28] and Johannesson et al.[29] later showed how discounting could be entered into the model).

Three assumptions regarding the shape of the utility function were identified by Pliskin et al.[24] to make QALYs a valid cardinal utility function. First, for a given number of life-years, the individual's preferences over health states are independent of the constant number of life years, and vice versa. For example, if the individual is indifferent between living 5 years in moderate health, and having a 50% likelihood of living 5 years in either perfect health or poor health, then he or she should also be indifferent between these choices for all other durations. Secondly, the proportion of remaining life-years the individual is willing to give up for an improvement in health does not depend on the absolute number of remaining life-years involved. This can be illustrated by the case when the individual is indifferent between 5 years in perfect health and 10

years in moderate health. If so, he or she should also be indifferent between 10 years in perfect health and 20 years in moderate health. Thirdly, the individual must be risk-neutral over life-years (for all health states). Risk neutrality implies that, for instance, 5 additional years in perfect health with certainty is perceived by the individual as giving equal utility as a 50% likelihood of either 10 or zero additional years in perfect health. The first 2 assumptions are technically known as mutually utility independence and constant proportionally trade-off. Bleichrodt et al.[30] recently showed that, given risk neutrality, utility functions starting from the origin in the QALY weight/time space implies that these 2 assumptions automatically are satisfied.

With the assumptions outlined above, risk-neutral QALYs are then simply defined as the product of the utility of health and the number of life-years. There also exists a risk-adjusted QALY model in which constant proportional risk posture (or constant relative risk-aversion) must hold instead of risk-neutrality. This model has, however, empirically not gained as much attention.

2.4 Comparing Individual Preferences Measured as WTP and QALYs

So, what measure of individual preferences in economic evaluations invoke the most limiting assumptions? In principle, the individual WTP measure rests on a rather limited set of assumptions of individual behaviour, and is not affected by the choice of utility framework. In practice, there are a number of problems in the eliciting of individual preferences, following the WTP approach (see, for example, Mitchell and Carson[8]). Also, how well hypothetical WTP corresponds with real WTP is debated (but is not limited to CBA, as similar problems are present also for CEA/CUA), and the development of calibration techniques for hypothetical WTP is a current research area.[31]

When it comes to QALYs, however, a number of rather restrictive assumptions, that normally are not made when analysing individual behaviour in economic theory, need to be made. The utility of a given health status, for instance, is most often assumed to depend on many other circumstances,[32] rather than just the number (and quality) of the individual's remaining life-years (e.g. consumption of other goods, income, age, gender, family situation). Hence, if QALYs were based on individual preferences, the individual would behave in ways which would be considered as exceptions to the rule in other fields of economics in general – or health economics in particular – or not in accordance with empirical evidence (see section 4). The different assumptions of WTP and QALYs are presented in table I.

3. Aggregating WTPs and QALYs

If all the restrictions and assumptions of WTPs and QALYs outlined previously were met, then in-

Table I. The different assumptions needed when measuring individual preferences as either willingness to pay (WTP) or quality-adjusted life-years (QALYs) in economic evaluations

	WTP[a]	QALYs
Type of utility function	Ordinal	Cardinal
Other assumptions	1. The individual's preferences are: (i) complete; (ii) transitive; (iii) continuous; (iv) monotone; and (v) convex	1. The individual's preferences are: (i) complete; (ii) transitive; (iii) continuous; (iv) monotone; and (v) convex 2. Acceptance of the axioms of expected utility theory: (i) independence; (ii) desire for high probability of success; and (iii) compound probabilities 3. Health states worse than death do not exist 4. Utility is only dependent on health and the number of life years 5. Risk neutrality or constant relative risk-aversion 6. Mutual utility independence 7. Constant proportional trade-off

a For WTP assessed under risk, the axioms of expected utility theory must also hold.

dividual preferences would be correctly measured by WTPs and QALYs. If this were the case, then under what additional conditions would it be meaningful to aggregate individual WTPs and individual QALYs to some total measure of the benefits of a healthcare programme? This issue is addressed in sections 3.1 to 3.4.

3.1 The Social Welfare Function

In order to aggregate individual WTPs or individual QALYs to a total measure of benefits, some rules for aggregation have to be adopted. Several options are available, but in current practice the chosen rule is seldom explicitly stated. There has been quite a bit of discussion in the economic literature in general, however, about the aggregation of individual utilities. The SWF, a concept introduced by Bergson,[33] Lange[34] and Samuelson,[35] has played an important role in the discussion, and it is also applicable in this particular context.

The SWF specifies the rules for aggregating individuals' utilities into social welfare. The problem of maximising social welfare, in general, is to maximise the utility functions of all the individuals in society, subject to the society's budgetary constraint. The solution of this maximisation problem can easily be proven to be Pareto efficient i.e. being an allocation of resources in which no individual can be made better off without anyone else being worse off.[23] However, making some logical requirements on the underlying social preference relation, Arrow[36,37] showed that the only possible way to use the SWF is to assume that the individual utility functions are cardinal. In general, however, the higher the ability (and information) of the decision-maker to measure and compare individuals' welfare and utilities, the larger the set of feasible SWFs.[38-41] Furthermore, the optimisation problem is very general and does not specify the functional form (i.e. the aggregation process of the individuals' utility functions) of the SWF.

A large number of different SWFs exist. At one end of the spectrum is the utilitarian SWF. This is the classical SWF, and is most often the basis for WTP and QALY aggregation. In this case, all utilities are weighted equally and, hence, reduced inequality does not increase social welfare. Harsanyi[42] showed that a SWF based on utilitarianism can be formed only if the individual utility functions as well as the SWF satisfy the von Neumann-Morgenstern axioms, and that individual indifference between 2 alternatives also implies social indifference. At the other end of the spectrum is the Rawlsian SWF.[43] In this case, social welfare depends only on the lowest individual utility. Between these 2 extremes lies the case where some weight is attached both to all the individual utilities and to its distribution.[44] Here, both efficiency and equity aspects are taken into account. Empirically, the SWF can be evaluated at a number of different levels for the efficiency-equity trade-off in order to assess the sensitivity of the decision-maker's concern for equity. Typical, but arbitrary, indifference curves for these 3 versions of the SWF are illustrated in figure 1.

It should be observed, however, that the SWF does not have to be individualistic in the sense that social welfare is a function of individual utilities. The concept is more general than this. It could be any ordering of alternative social states; in particular, it could be defined over individual characteristics such as individual incomes, health, etc., without reference to individuals' utilities.

3.2 How to Aggregate Individuals' WTP

Much attention has been devoted to the issue of aggregating WTP in CBA (see Johansson[45] for a formal presentation). Basically, the change in social welfare is calculated by the weighted sum of the individuals' compensating variation (or maximum WTP), adjusted for their marginal utility of income (i.e. the individual's value, or utility, of the last unit of income). If a new policy increases the individuals' health, the welfare measure of this change is thus the sum of the individuals' maximum WTP (adjusted by their relative marginal utility of income) multiplied by the society's (equity) weights placed on these individuals' respective marginal utilities. Hence, the welfare change is the individuals' maximum WTP multiplied by the social

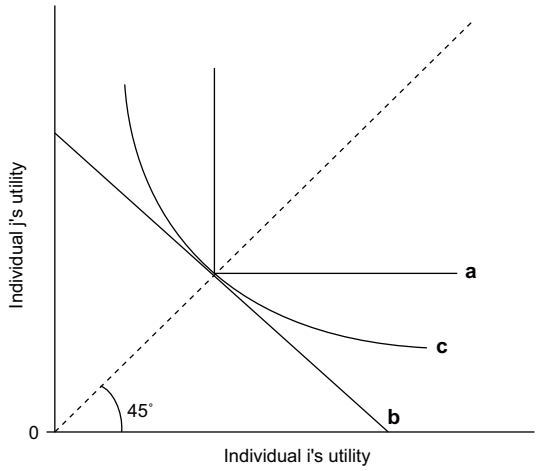

Fig. 1. Typical, but arbitrary, indifference curves for 3 versions of the social welfare function (SWF). (**a**) stands for the Rawlsian SWF, (**b**) for the utilitarian SWF and (**c**) for a SWF that lies in between these 2 extremes and, thus, reflects some trade-off between efficiency and equity.

marginal utility of individual income. Thus, as long as society weighs all individuals equally (i.e. the distribution of income is of no concern), and the individuals' marginal utility of income is the same, unweighted WTP is a correct measure of changes in economic welfare.

If these assumptions do not hold, an investment that meets the Kaldor-Hicks' welfare criterion[23] [defined as a change that makes the gainers so much better off that they can at least (hypothetically) compensate the losers], usually (assuming no transaction costs) will increase the economic welfare to society. This is so because, for example, an investment in which individual i gains $2 and individual j loses $1 might not be indifferent to society if the marginal utility of income and/or the equity weight placed on the individuals' utilities, is higher for individual j, and the compensation actually never takes place (i.e. it remains hypothetical). However, if the government could redistribute incomes over individuals so that the social marginal utility of income became equal for all individuals, then it would be possible to maximise the social welfare by aggregating the individuals' maximum WTPs without worrying about these assumptions. Thus, it might be possible to convert a potential Pareto improvement (i.e. an investment that meets the Kaldor-Hicks' welfare criterion) to an actual Pareto improvement (i.e. an investment that leads to at least 1 individual becoming better off without anyone else being worse off).

3.3 How to Aggregate Individuals' QALYs

There have been few attempts to explicitly incorporate QALYs (interpreted as utilities) in an SWF.[46-49] An exception is Bleichrodt,[50] who distinguished between the conditions for aggregation (i) within the utilitarian framework and (ii) in SWFs in which equity in the utility of health is of concern.

For the utilitarian SWF, *changes* in individual QALYs (or utilities) must be cardinally comparable among individuals (cardinal unit comparability) but not absolute levels of individual QALYs (or utilities). Furthermore, anonymity (meaning that the social preferences are independent of who gets the 'final' QALY) must also hold. A healthcare programme would then be preferred from the societal point of view if the total gains in utility of those who would gain utility from the programme are larger than the total losses of those who would lose utility from the programme – no matter who the winners or losers might be, or whether those who gained most were the better off also without the programme. Thus, changes in utilities are not given any distributional weights. It should be observed that there is no requirement that individual QALYs should be measured on a scale between 0 and 1.

For SWFs in which equity in the utility of health is of concern, absolute levels of individual QALYs (utilities) must be cardinally comparable (cardinal full comparability). A healthcare programme would be preferred from the societal point of view if the sum of equity weighted gains in utilities among the winners is greater than the sum of equity weighted losses in utilities among the losers. In this case, a cardinal measure of differences in utility is not sufficient. Since the distribution of the utility of health is of concern in the aggregation procedure, measurement also requires that absolute (the initial) lev-

els of individual utility of health can be compared. It should be observed that this does not necessarily mean that individuals must have identical preferences over health states. Individual preferences may differ, but they should be measurable in such a way that, for instance, the utility of perfect health for individual A is cardinally comparable with the utility of perfect health for individual B.

The necessary conditions for the aggregation of individual QALYs are very restrictive in comparison with what is being considered reasonable assumptions in mainstream economics. This is true in particular for the SWF in which equity in the utility of health is of concern. At the same time, there is evidence that people do care about equity in the health area. An experimental attempt to investigate individuals' willingness to trade efficiency for equity of QALYs between 2 hypothetical groups was done by Johannesson and Gerdtham.[51] They found that the respondents' mean marginal willingness to trade was significant (they were, on average, willing to give up 1 QALY in the group with more QALYs to gain 0.5 QALYs in the group with fewer QALYs), but that the size of the difference in the number of QALYs between the 2 groups was insignificant. In an experiment of similar design, Andersson and Lyttkens[52] found an even higher inequality aversion; the respondents were here willing to give up 1 QALY in the group with more QALYs, to gain 0.11 to 0.35 QALYs in the group with fewer QALYs. They also found, contrary to Johannesson and Gerdtham,[51] that the size of the difference in the QALY distribution had a significant impact on the trade-off.

The aggregation problem would, of course, be slightly easier if all individuals valued each arbitrarily chosen health state in exactly the same way. Still, absolute levels of utility must be cardinally measurable even in this case. Furthermore, there is ample empirical evidence that individuals differ considerably in their preferences for specific health states.[53-55] To be fair, when the utility of immediate death is assigned the value zero and perfect health the value 1 in QALY-based CEA/CUA, this should not be interpreted as positive statements about what individual preferences really are but as normative statements about how individual utilities of health states should be treated. Thus, the utility of perfect health and the utility of immediate death, respectively, should be treated as if they were the same for everyone. For other health states, individual utilities are assumed to be cardinally measurable, but not necessarily identical. In the aggregation procedure they are simply summed and averaged in the utilitarian SWF, but weighted before summation in the SWFs in which equity is of (special) concern. However, since some regard has already been taken to equity by giving the same utility to everyone for the health states immediate death and perfect health, the former is not quite a utilitarian SWF and should, hence, perhaps be referred to as a semi-utilitarian SWF. To assign predefined real values to 2 outcomes, X and Y, such that everyone prefers X to Y in order to create a consistent scaling procedure for social welfare judgments based on von Neumann-Morgenstern utilities, was first suggested by Hildreth.[56] Obviously, the QALY approach as usually practised, is a special case of this general idea.

3.4 Comparing the Aggregation of WTP and QALYs

So, which aggregation process imposes the most restrictive assumptions – the one for WTP or QALYs? When making this comparison we will leave the assumptions of the underlying SWF out, as they are the same for both WTP and QALYs.

For WTP, we have seen that the underlying assumptions are relatively few. Taking a Paretian perspective, the 'only' assumption that needs to be made is that the marginal utility of income has to be the same for all individuals. However, if society is to maximise potential Pareto improvements, not even this assumption is needed. The issues involved in aggregating individual QALYs have recently gained some attention, and it has been shown that individual utilities need to be both cardinal and comparable across individuals. It is completely clear that the conditions for aggregating individual QALYs – in addition to the conditions for individual QALYs

Table II. The assumptions needed [on top of those for measuring individual preferences as willingness to pay (WTP) or quality-adjusted life-years (QALYs)] for aggregating WTP and QALYs in a utilitarian framework, as well as in other social welfare functions (SWFs)

	WTP	QALYs
Utilitarian SWF	1. If a Paretian perspective: equal marginal utilities of income 2. If a potential Pareto perspective: no further assumptions	1. Cardinal unit comparability (i.e. changes in individual QALYs must be cardinally comparable among individuals) 2. Anonymity (i.e., that the social preferences are independent of who gets the 'final' QALY)
Other SWFs	1. The individuals' marginal utilities of income have to be estimated (which is not an assumption *per se*) 2. The society's equity weights of the individuals' utilities have to be estimated (which, again, is not an assumption *per se*)	1. Cardinal full comparability (i.e. absolute levels of individual QALYs must be cardinally comparable)

to be consistent with individual preferences – are very restrictive (see table II for a comparison).

4. Discussion

Individual WTPs in CBA are based, at least in principle, on exactly the same, fairly weak, assumptions about individual preferences which are commonly made in economic theory. In contrast, some very strong assumptions about individual preferences have to be made in order for QALYs to be interpreted as utilities.[57,58] The reason for the difference in assumptions between WTP and QALYs is (partly) that income (from which the individual's WTP is deducted from) is just one argument in the individual's utility function, whereas the number of QALYs is a utility function by itself. Mutual utility independence, constant proportional tradeoff and risk neutrality are assumptions not normally made in economic theory and they are also often contradicted by experimental evidence. See, for example, Bleichrodt and Johannesson[59] and Krabbe and Bonsel[60] for experimental tests of the first 2 assumptions, Stiggelbout et al.[61] for an experiment on the third assumption, and Shoemaker[62] for a review of experimental tests on expected utility theory. Moreover, it should be mentioned that other economic foundations than expected utility theory have been tested. In the study by Bleichrodt and Quiggin,[63] the prospect theory (which can be seen as a generalisation of the expected utility model) of Kahnemann and Tversky[64] was applied as an alternative. This was proven to increase the descriptive validity of the model. Furthermore, McNeil et al.[65] found that individuals were not willing to trade length of life for quality of life if the remaining years of life was less than 5. When the remaining years of life were less than 5, it was only length, rather than quality, of life that affected the individual's utility. Sackett and Torrance[66] noticed that the value that individuals attach to different health states is highly affected by the duration they would have to remain in that health state.

In our view, rather than examining under what strong assumptions QALYs can be interpreted as utilities, it would be more interesting to examine whether there exists a health index based on plausible assumptions about individual preferences. However, according to the analysis of Hougaard and Keiding,[67] such an index quickly runs into the same issues, and is thus not easily achievable. In a Lancaster or Grossman framework (in which health and other commodities are produced by using own time and market goods), Hougaard and Keiding[67] show that the health index (or QALY) must be completely individual-specific, or preferences have to be identical, production technologies the same, and the initial endowments of health and wealth equal.

Furthermore, while the consistent aggregation of individual WTPs boils down to an assumption regarding the individuals' marginal utilities of income, the aggregation of QALYs requires an assumption of cardinal unit comparability of individual utilities. Bleichrodt and Quiggin[68] showed under what conditions CEA/CUA is consistent with CBA for the more general (and realistic) utility function consisting of both consumption and health status. It turns out that very restrictive assumptions have to be made, e.g. that: (i) the individual's utility is multiplicative in the utility of consumption

and the utility of health status (implying that improvements in health are valued higher at high levels of consumption, which has obvious distributional implications); and (ii) the utility of consumption is constant over time (which in this case implies that consumption is constant). The latter assumption is a much stronger assumption than the former one.

In fact, since Robbins' methodological critique,[69,70] disputing the scientific status of interpersonal comparisons of utilities, many economists seem to have come to the conclusion that no interpersonal comparisons whatever are allowed or recommended. Others, such as Sen,[71,72] have criticised their fellow economists for avoiding one of the most important political issues i.e. inequalities in individuals' opportunities. Sen has instead argued for moving away from a utility-based SWF to interpersonal comparisons based on non-utility information, such as functionings and capabilities.

Taking all these problems into consideration, the cautious economist may feel that he or she could just as well leave the tricky normative issues in health and healthcare to others and concentrate on the issues of positive health economics. However, decisions about allocating scarce healthcare resources will be made whatever the economist decides for him or herself. The challenge to provide decision-makers with information for improving the allocation of resources, both in terms of efficiency and equity, would still be there. Portney[73] recently raised a similar argument as a response to several critics of the contingent valuation method in the environmental field (who argued that it was flawed and thus should be abandoned by economists, rather than further developed).

What options would then be left for economic evaluation? If you are a traditional Paretian welfare economist, claiming that all decisions about resource allocation should be based on individual utilities, the obvious answer would be to increase the use of WTP-based CBA. This idea was recently promoted by Kenkel,[74] who argued that the customers of economic evaluations should be given (what he perceives to be) the correct economic evaluation (i.e. CBA) rather than what they demanded (CEA/CUA). This approach does not necessarily completely resolve the issue of equity considerations, however. In practice, most CBA done so far have taken the unweighted mean of the concerned individuals' WTPs as a measure of health benefits (i.e. a utilitarian approach). From an economic welfare perspective, the procedure rests on the assumption that the marginal utilities of income for all individuals in question are equal. So, few (if any) cost-benefit studies have in reality tried to take equity into account.

If you are concerned with individuals' health and the distribution of health among individuals rather than just the utilities which individuals may derive from their health (or the distribution of utilities), you may choose to treat QALYs as an 'objective' measure of health in its own right, or at least as a commonly accepted way of measuring outcomes in healthcare. Maximising the number of equity-weighted QALYs would then be quite an appropriate goal for the healthcare sector (or society at large). This is in line with Sen's[71,72] advocacy for interpersonal comparisons in the non-utility space, for which there is very much to be said, in our opinion. For instance, according to the Swedish Health Care Act of 1984, the main objectives of the Swedish healthcare system are the health of individuals and the distribution of health in the population. This trade-off between efficiency and equity has recently been confirmed in a survey to Swedish politicians.[75] Furthermore, see Olsen[76] for a discussion on the priority-setting in the Norwegian healthcare system (which is based on severity level and the capacity to benefit).

Sen's view is also shared by many health economists (see, for instance, Culyer,[1] Wagstaff,[47] and Williams[77]). Since this view is obviously not founded in economic welfare theory, it has been called the 'non-welfarist' (or 'extra-welfarist') view on economic evaluation in healthcare.

In principle, it would be possible to take equity considerations into account in 'extra welfarist' CEA/CUA, e.g. using a SWF in which utilities are replaced by health states.[47] In practice, however, aggregation has (implicitly) applied a utilitarian ap-

proach, with the same weights given to all individuals' health states (or rather, changes in health states), as measured by QALYs. Thus, so far, there exist few, if any, published CEA/CUA studies that have taken equity in health into account.

Another issue of debate, which concerns both methods, is how to deal with altruism i.e. how (and if) should the values individuals place on other individuals' health and/or utility be incorporated into the analysis. This has been touched upon in the case of CBA in healthcare a number of times,[78-80] but is still an interesting topic for further research within CEA/CUA. Labelle and Hurley[81] discussed this issue within the CEA/CUA framework, and stated that which externalities to include depends on what should be maximised: the SWF or the community's health (i.e. what interpretation of QALYs that should be made; as utilities or as health).

5. Conclusions

As a measure of individual preference in economic evaluations, the WTP is based on relatively few weak assumptions. The aggregation process for individual WTPs also requires relatively limited weak conditions. In contrast, QALYs have a number of restrictive assumptions in order to be interpreted as utilities. Furthermore, the conditions for aggregating individual QALYs are also very restrictive, beyond those normally made in economic theory.

So, if researchers want their analyses to be based on economic welfare theory, CBA should be preferred. However, if QALYs are interpreted as commonly accepted measures of health rather than as utilities, and if the objective of the healthcare system is to maximise health rather than welfare of the population, then there is a strong rationale for using CEA/CUA.

The challenges to the practitioners of both the WTP (or CBA) and the QALY (or CEA/CUA) approach are similar. The approaches as used in practice must be valid and reliable in order to be of any value for healthcare decision-makers. Conceptual issues and practical problems seem to abound,[18,82,83] and the need for future research and development is substantial, indeed.

Acknowledgements

Financial support for this study was provided in part by grants from the Swedish Medical Research Council and the Swedish Council for Social Research. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing and publishing the report. We would like to thank Fredrik Andersson, Sören Höjgård and Carl Hampus Lyttkens (Lund University), and Magnus Johannesson (Stockholm School of Economics) for their comments on earlier versions of this paper.

References

1. Culyer AJ. The normative economics of health care finance and provision. Oxford Rev Econ Policy 1989; 1: 34-58
2. Nord E, Pinto JL, Richardson J, et al. Incorporating societal concerns for fairness in numerical valuations of health programmes. Health Econ 1999; 8: 25-39
3. Brouwer WBF, Koopmanschap MA. On the economic foundation of CEA. Ladies and gentlemen, take your positions! J Health Econ 2000; 19: 439-59
4. Drèze J, Stern N. The theory of cost-benefit analysis. In: Auerbach AJ, Feldstein M, editors. Handbook of public economics. Vol. II. Amsterdam: North-Holland, 1987: 909-89
5. Prest AR, Turvey R. Cost-benefit analysis: a survey. Econ J 1965; 75: 683-735
6. Warner KE, Luce BR. Cost-benefit and cost-effectiveness analysis in health care. Ann Arbor (MI): Hospital Administration Press, 1982
7. Sugden R, Williams A. The principles of practical cost-benefit analysis. Oxford: Oxford University Press, 1978
8. Mitchell RC, Carson RT. Using surveys to value public goods: the contingent valuation method. Washington, DC: Resources for the Future, 1989
9. Weisbrod BA. Economics of public health. Philadelphia (PA): University of Philadelphia Press, 1961
10. Klarman HE. The economics of health. New York (NY): Columbia University Press, 1965
11. Backhouse ME, Backhouse RJ, Edey SA. Economic evaluation bibliography. Health Econ 1992; 1 Suppl.: 3-7
12. Elixhauser A, Luce BR, Taylor W, et al. Health care cost-benefit and cost-effectiveness analysis (CBA/CEA) from 1979 to 1990: a bibliography. Med Care 1993; 31: JS139-JS141
13. Elixhauser A, Halpern M, Schmier J, et al. Health care CBA and CEA from 1991 to 1996: an updated bibliography. Med Care 1998; 36: MS1-MS9
14. Fanshel S, Bush JW. A health-status index and its application to health-services outcomes. Oper Res 1970; 18: 1021-66
15. Bush JW, Fanshel S, Chen MM. Analysis of a tuberculin testing program using a health status index. Socioecon Plann Sci 1972; 6: 49-68
16. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health Serv Res 1972; 7: 118-33
17. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. N Engl J Med 1977; 296: 716-21

18. Drummond MF, O'Brian BJ, Stoddart GL, et al. Methods for the economic evaluation of health care programmes. 2nd ed. Oxford: Oxford University Press, 1997
19. Torrance GW. Health status index models: a unified mathematical view. Man Sci 1976; 22: 990-1001
20. McGuire A, Henderson T, Mooney G. The economics of healthcare: an introductory text. London: Routledge and Kesan Paul, 1988: 38-53
21. Baumol WJ. The cardinal utility which is ordinal. Econ J 1958; 68: 665-72
22. Harsanyi JC. Cardinal utility in welfare economics and the theory of risk-taking. J Pol Econ 1953; 61: 434-5
23. Varian HR. Microeconomic analysis. 3rd ed. New York (NY): Norton International Student Edition, 1992
24. Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. Oper Res 1980; 28: 206-24
25. von Neumann J, Morgenstern O. Theory of games and economic behaviour. 2nd ed. Princeton (NJ): Princeton University Press, 1947
26. Baumol WJ. Economic theory and operations analysis. 3rd ed. London: Prentice Hall, 1972
27. Kind P, Rosser R, Williams A. Valuation of quality of life: some psychometric evidence. In: Jones-Lee MW, editor. The value of life and safety. Amsterdam: North-Holland Publishing Company, 1982: 159-70
28. Broome J. Qalys. J Public Econ 1993; 50: 149-67
29. Johannesson M, Pliskin JS, Weinstein MC. A note on QALYs, time tradeoff, and discounting. Med Decis Making 1994; 14: 188-93
30. Bleichrodt H, Wakker P, Johannesson M. Characterizing QALYs by means of risk neutrality. J Risk Uncertainty 1997; 15: 107-14
31. Liljas B, Blumenschein K. On hypothetical bias and calibration in cost-benefit studies. Health Policy 2000; 52: 53-70
32. Grossman M. On the concept of health capital and the demand for health. J Pol Econ 1972; 80: 223-55
33. Bergson A. A reformulation of certain aspects of welfare economics. Q J Econ 1938; 52: 310-34
34. Lange O. The foundations of welfare economics. Econometrica 1942; 10: 215-28
35. Samuelson PA. Foundations of economic analysis. Cambridge (MA): Harvard University Press, 1947
36. Arrow KJA. A difficulty in the concept of social welfare. J Pol Econ 1950; 58: 328-46
37. Arrow KJA. Social choice and individual values. New Haven (CT): Yale University Press, 1951
38. Boadway RW, Bruce N. Welfare economics. Oxford: Basil Blackwell, 1984
39. Sen AK. On weights and measures: informational constraints in social welfare analysis. Econometrica 1977; 45: 1539-72
40. Sen AK. The possibility of social choice. Am Econ Rev 1999; 89: 349-78
41. Slesnick DT. Empirical approaches to the measurement of welfare. J Econ Lit 1998; 36: 2108-65
42. Harsanyi JC. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. J Pol Econ 1955; 63: 309-21
43. Rawls J. A theory of justice. Cambridge (MA): Harvard University Press, 1971
44. Atkinson AB. On the measurement of inequality. J Econ Theory 1970; 2: 244-63
45. Johansson PO. Evaluating health risks: an economic approach. Cambridge: Cambridge University Press, 1995
46. Butler JRG. Welfare economics and cost-utility analysis. In: Zweifel P, Frech III HE, editors. Health economics worldwide. Dordrecht: Kluwer Academic Publishers, 1992: 143-57
47. Wagstaff A. QALYs and the equity-efficiency trade-off. J Health Econ 1991; 10: 21-41
48. Williams A. Welfare economics and health status measurement. In: van der Gaag J, Perlman M, editors. Health, economics, and health economics. Amsterdam: North-Holland Publishing Company, 1981: 271-81
49. Williams A. QALYS and ethics: a health economist's perspective. Soc Sci Med 1996; 43: 1795-804
50. Bleichrodt H. Health utility indices and equity considerations. J Health Econ 1997; 16: 65-91
51. Johannesson M, Gerdtham UG. A note on the estimation of the equity-efficiency trade-off for QALYs. J Health Econ 1996; 15: 359-68
52. Andersson F, Lyttkens CH. Preferences in equity in health behind a veil of ignorance. Health Econ 1999; 8: 369-78
53. Johannesson M, Johansson PO. Is the value of a QALY gained independent of age? Some empirical evidence. J Health Econ 1997; 16: 589-600
54. Loomes G, McKenzie L. The use of QALY's in health care decision making. Soc Sci Med 1989; 28: 299-307
55. Weinstein MC. A QALY is a QALY is a QALY – or is it? J Health Econ 1988; 7: 289-90
56. Hildreth C. Alternative conditions for social ordering. Econometrica 1953; 21: 81-9
57. Carr-Hill RA. Assumptions of the QALY procedure. Soc Sci Med 1989; 29: 469-77
58. Gafni A. Proper preference-based outcomes measures in economic evaluations of pharmaceutical interventions. Med Care 1996; 34: DS48-DS58
59. Bleichrodt H, Johannesson M. The validity of QALYs: An experimental test of constant proportional tradeoff and utility independence. Med Decis Making 1997; 17: 21-32
60. Krabbe PFM, Bonsel GJ. Sequence effects, health profiles, and the QALY model: in search of realistic modeling. Med Decis Making 1998; 18: 178-86
61. Stiggelbout A, Kiebert GM, Kievit J, et al. Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. Med Decis Making 1994; 14: 82-90
62. Shoemaker PJH. The expected utility model: its variants, purposes, evidence, and limitations. J Econ Lit 1982; 20: 529-63
63. Bleichrodt H, Quiggin J. Characterizing QALYs under a general rank dependent utility model. J Risk Uncertainty 1997; 15: 151-65
64. Kahnemann D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979; 47: 263-91
65. McNeil BJ, Weichselbaum R, Pauker SG. Speech and survival: tradeoffs between quality and quantity of life in laryngeal cancer. N Engl J Med 1981; 305: 982-7
66. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. J Chronic Dis 1978; 31: 697-704
67. Hougaard JL, Keiding H. On the welfare economic foundations of health status measures. Working paper. Copenhagen: Department of Economics, University of Copenhagen, 1999
68. Bleichrodt H, Quiggin J. Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost-benefit analysis? J Health Econ 1999; 18: 681-708
69. Robbins L. An essay on the nature and significance of economic science. London: Allen and Unwin, 1932
70. Robbins L. Interpersonal comparisons of utility: a comment. Econ J 1938; 48: 635-41

71. Sen AK. Equality of what? In: McMurrin S, editor. Tanner lectures on human values. Cambridge: Cambridge University Press, 1980
72. Sen AK. On economic inequality. 2nd ed. Oxford: Oxford University Press, 1997
73. Portney PR. The contingent valuation debate: why economists should care. J Econ Persp 1993; 8: 3-17
74. Kenkel D. On valuing morbidity, cost-effectiveness analysis, and being rude. J Health Econ 1997; 16: 749-57
75. Lindholm L, Rosen M. On the measurement of the nation's equity adjusted health. Health Econ 1998; 7: 621-8
76. Olsen JA. Theories of justices and their implications for priority setting in health care. J Health Econ 1997; 16: 625-39
77. Williams A. Intergenerational equity: an exploration of the 'fair innings' argument. Health Econ 1997; 6: 117-32
78. Johansson PO. Altruism and the value of statistical life: empirical implications. J Health Econ 1994; 13: 111-8
79. Jones-Lee MW. Altruism and the value of other people's safety. J Risk Uncertainty 1991; 4: 213-9
80. Jones-Lee MW. Paternalistic altruism and the value of a statistical life. Econ J 1992; 102: 80-90
81. Labelle RJ, Hurley JE. Implications of basing health-care resource allocations on cost-utility analysis in the presence of externalities. J Health Econ 1992; 11: 259-77
82. Gold MR, Siegel JE, Russell LB, et al., editors. Cost-effectiveness in health and medicine. New York (NY): Oxford University Press, 1996
83. Johannesson M. Theory and methods of economic evaluation of health care. Dordrecht: Kluwer Academic Publishers, 1996

Correspondence and offprints: *Bengt Liljas*, AstraZeneca, 221 87 Lund, Sweden.
E-mail: bengt.liljas@astrazeneca.com