IN LIMINE SAPIENTIÆ

UNIVERSITY
OF YORK

CENTRE FOR HEALTH ECONOMICS
HEALTH ECONOMICS CONSORTIUM

QALYs versus HYEs
A Theoretical Exposition

by A. J. Culyer and Adam Wagstaff

DISCUSSION PAPER 99

University of York
Centre for Health Economics

QALYS VERSUS HYES
A THEORETICAL EXPOSITION

by
A.J. Culyer and Adam Wagstaff

September 1992

THE AUTHORS

A.J. Culyer is Professor of Economics in the Department of Economics and Related Studies at the University of York.

Adam Wagstaff is Lecturer in Economics in the School of Social Sciences at the University of Sussex.

ACKNOWLEDGEMENTS

This paper was written whilst Adam Wagstaff was visiting the Centre for Health Economics at the University of York and the Institute for Medical Technology Assessment (IMTA) at Erasmus University Rotterdam. The support of both institutions is acknowledged. The IMTA support was funded by a grant from Merck Sharpe and Dohme for research on 'methodological issues in economic evaluation'.

FURTHER COPIES

Further copies of this document are available (at price £2.00 to cover the cost of publication, postage and packing) from:

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

Please make cheques payable to the University of York. Details of other papers can be obtained from the same address, or telephone York (0904) 433648 or 433666.

ABSTRACT

There has been a vigorous dispute about quality adjusted life years (QALYs) and it has been argued that they are an inappropriate measure of patient utility and that a more efficient approach is to measure outcomes in terms of health year equivalents (HYEs). This paper explores the theoretical underpinning of this debate.

It explores the claim that QALYs are liable to misrepresent consumer preferences and hence lead to decision-makers choosing options which are not those preferred by the public. It also considers the claim that HYEs do not suffer from this defect. We argue that none of the examples offered to date demonstrate the alleged tendency of QALYs to misrepresent preferences. We also suggest that although QALYs *may* misrepresent preferences in a way that HYEs do not, since they require that the individual's utility function be additively separable over time, there is no evidence to date that QALYs do so.

1. Introduction

Quality-Adjusted Life-Years (QALYs), as a species of outcome measure, have filled a yawning gap in the range of instruments available to epidemiologists, health service managers and policy makers, by providing a quantitative (if incomplete) algorithm for assessing the relative benefits of alternative health care programmes and highlighting several key value judgments that were previously all too frequently buried in less formal ways of assessing priorities (of which the worst was probably that form of noisy special pleading commonly called "shroud waving"). Mature statements by two intellectual leaders are Williams (1985) and Torrance (1986).

QALYs have also, however, been subject to criticism. In this paper we consider one of these: that QALYs are liable to misrepresent consumer preferences and hence lead to decision-makers choosing options which are not those preferred by the public. This argument, which, if true, clearly undermines the QALY approach substantially, has been set out by Gafni (1989), Mehrez and Gafni (1989) and Gafni and Zylak (1990). These authors also set out an alternative to QALYs - Healthy Years Equivalents (HYEs) - which, according to them, do not suffer from this defect.

We argue in this paper that the criticisms of Gafni and colleagues are misguided. We conclude that in none of their examples do they demonstrate the alleged tendency of QALYs to misrepresent preferences. We also suggest that although QALYs *may* misrepresent preferences in a way that HYEs do not, there is no evidence to date that QALYs do so.

2. QALYs and HYEs compared

QALYs and HYEs are both intended to capture the not only the individual's remaining life expectancy, but also the quality of life in each remaining year.

2.1 QALYs

An individual's QALY score is calculated by weighting each expected remaining life-year by the expected quality of life in the year in question and then computing the discounted sum of these values. The "qualities" are calculated as probability-weighted averages of the quality-of-life scores associated with each of the possible health states in which the individual may be expected to be found. These scores are measured on a scale ranging from (typically) zero (corresponding to "death") to one ("perfect health"). They may be derived in one of two ways. The first, which we call the W-type QALY [cf. Williams (1985)], seem to be ratio scales and are based on a sample of subjects' relative valuations of descriptive states. The second, which we call the T-type QALY [cf. Torrance (1986)], seem to be interval scales and are based on what Torrance has termed the "gold standard" procedure of the standard gamble (SG).

T-type QALY scores, on which we focus in this paper, are often described as "utility" measures, since they are typically derived by experimental methods rooted in the utility axioms of the von Neumann-Morgenstern basis for the SG and are intended to represent the subject's preferences. For example, in order to establish an individual's utility score for state Q^1, the individual would be offered a choice between Q^1 with certainty and a gamble involving perfect health with probability p and death with probability (1-p). The probability p is varied until the individual is indifferent between the two alternatives. Denoting this probability by p^*, we have

(1) $$u(Q^1) = p^* \cdot u(P) + (1-p^*) \cdot u(D)$$

where $u(Q^1)$, u(P) and u(D) are the utilities associated with being in the states Q^1, 'perfect health' and 'dead'. If we set $u(P) \equiv 1.0$ and $u(D) \equiv 0.0$, then the utility associated with state Q^1 can be measured as p^*.

Suppose, for simplicity, that an individual occupies just one health state each year and that it is known with certainty which state this will be. These states are denoted by Q_t $(t=1,\ldots,T)$. Suppose too that the utility scores for these states have been established. Then the individual's QALY score is calculated as

(2) $$\text{QALY} = \Sigma_{t=1}^{T} [1/(1+i)]^{t-1} u(Q_t),$$

where $u(Q_t)$ is the "utility" associated with being in state Q_t for one year and i is the rate of discount.

2.2 HYEs

HYEs have their roots even more firmly in utility theory. Denote by U the utility associated with the health profile $(Q_1,Q_2,\ldots,Q_T)$. Thus

(3) $$U = U(Q_1,Q_2,\ldots,Q_T).$$

Now consider an alternative profile, which gives rise to the same utility U, but which entails H years of perfect health followed by death. Let P and D represent the states 'perfect health' and 'dead' respectively. Then H is defined implicitly by the equation

(4) $$U = U(P_1,P_2,\ldots,P_H,D_{H+1},D_{H+2},\ldots,D_T),$$

so that H is the number of years of perfect health which give rise to a utility level U.

Mehrez and Gafni suggest a two-stage procedure to compute HYEs. In the first stage the SG technique is used to obtain a utility score for the health profile $(Q_1,Q_2,\ldots,Q_T)$. Thus individuals are offered a choice between $(Q_1,Q_2,\ldots,Q_T)$ with certainty and a gamble involving T periods of perfect health with probability p and immediate death with probability (1-p). As before, the probability p is varied until the individual is indifferent between the two alternatives. Denoting the indifference probability by p', we have

$$(5) \qquad U(Q_1,\ldots,Q_T) = p'\cdot U(P_1,\ldots,P_T) + (1-p')\cdot U(D_1,\ldots,D_T).$$

If we set $U(P_1,\ldots,P_T) \equiv 1.0$ and $U(D_1,\ldots,D_T) \equiv 0.0$, then the utility associated with the health profile $U(Q_1,\ldots,Q_T)$ can be measured as p'. In the second stage individuals are offered another choice, this time between H years in perfect health with certainty and a gamble involving T years in perfect health with probability p' and immediate death with probability (1-p'). The value of H is varied until the individual is indifferent between the two alternatives. We thus have

$$(6) \qquad U(P_1,\ldots,P_{H'}) = p'\cdot U(P_1,\ldots,P_T) + (1-p')\cdot U(D_1,\ldots,D_T),$$

where H' is the number of HYEs.

2.3 When do QALYs and HYEs give different results?

The obvious question arises: under what circumstances will the number of QALYs and the number of HYEs be the same? A necessary condition for the number of QALYs to be equal to the number of HYEs is that the utility function U(.) be additively separable over time. This is, in fact, implicitly assumed to be the case in the QALY approach. If this restriction is imposed on U(.), eq (3) becomes

$$(3') \qquad U = \Sigma_{t=1}^{T} [1/(1+i)]^{t-1} u(Q_t),$$

where u(.) is a utility function common to all periods. Hence in this case the number of QALYs is equal to U. Under intertemporal additive separability, eq (4) becomes

$$(4') \qquad U = u(P)\cdot\Sigma_{t=1}^{H} [1/(1+i)]^{t-1}$$

$$= \Sigma_{t=1}^{H} [1/(1+i)]^{t-1}$$

if we assume $u(P) \equiv 1.0$. But, in general, the RHS of (4') will be less than H, since one would expect $i > 0$. The implication is that, if the utility function is additively separable, H will, in general, be greater than the number of QALYs. It is apparent from eq (4') that, in order for the numbers of QALYs and HYEs to coincide, we require not only intertemporal additive separability but also a zero discount rate.

It might be argued that the relevant question is not "Under what circumstances will the numbers of QALYs and HYEs be the same?", but rather "Under what circumstances will the two methods generate the same rankings of projects?". For good or for bad, much of the interest to date in QALYs has, after all, been in QALY league tables. Inspection of eq (4') above reveals that under intertemporal additive separability the two methods will produce identical rankings: U is a monotonically increasing function of H. The crucial difference, then, between QALYs and HYEs is that the former assume intertemporal additive separability whilst the latter do not.

2.4 QALYs and HYEs in the assessment of contrast media

The roles of intertemporal additive separability and discounting are shown clearly in Gafni and Zylak's (1990) paper on the use of QALYs and HYEs in calculating the health gains from ionic and non-ionic contrast media. They seek to show that the QALY calculations reported by Goel et al. (1989) are unreliable because they fail to represent correctly patient preferences. They then go on to report some HYE calculations which they claim overcome these defects.

Table 1 shows the probabilities and quality-of-life scores which appear to have been attached by Goel et al. to the four possible outcomes associated with each contrast medium. Goel et al. assume a zero discount rate and obtain the QALY figures indicated. Gafni and Zylak employ HYEs instead of QALYs and claim that because of this they obtain a different result.

The difference in the result can easily be shown to stem not from Gafni and Zylak's use of HYEs but rather from their making different assumptions about the effects on the quality of life of the tests themselves and of the development of minor reactions. If the QALY calculations are re-done using the same set of assumptions as employed in the HYE approach, the results are precisely the same.

Note from table 1 that Gafni and Zylak assume 21 HYEs for the major reaction outcome. This result is obtained by multiplying the quality-of- life score used by Goel et al. by 30 (the assumed length of life after the test). To be consistent, Gafni and Zylak ought to have adopted the same approach when computing the HYEs associated with the other two non-fatal outcomes. In fact they assigned HYE values to these two outcomes which are not equivalent to the quality-of-life figures used by Goel et al. When the two sets of calculations are re-done using the same assumptions the two approaches give precisely the same result (table 1 shows what happens if one uses the assumptions of Goel et al. throughout). This is to be expected given our earlier result concerning the equivalence of QALYs and HYEs when a zero discount is imposed and given that the assumptions used are the same in both approaches.

The earlier discussion indicated that the QALY and HYE approaches will generate different numbers - but not different rankings - if discounting is introduced. Table 2, which shows the effect of discounting (a 10% rate has been used) but using the same assumptions throughout, confirms this. The QALYs and HYEs associated with the two procedures are now different, but the ordering of the media are unchanged. From the discussion above, it is also clear that for the two methods to generate different rankings, one would have to drop the assumption of intertemporal additive separability in computing HYEs.

3. Do QALYs misrepresent preferences?

Mehrez and Gafni (1989) and Gafni (1989) have argued that QALYs tend to misrepresent preferences but HYEs do not. Gafni, for example, suggests that "QALYs do not stem directly from the individual's utility function and thus, at best, only partly reflect an individual's preferences". He suggests that this "might lead to the choice of the non-preferred alternative due to the misrepresentation of the individual's preferences". HYEs, according to Gafni, do not suffer from this defect.

This claim is misleading. The crucial issue again turns out to be whether the individual's utility function is additively separable over time. If it is, as advocates of QALYs assume, the claims of Mehrez and Gafni regarding QALYs are without foundation.

3.1 Mehrez and Gafni's painful treatment example

Mehrez and Gafni give an example in which an individual faces two scenarios: scenario A in which he undergoes treatment for three months during which his health state is Q_A^1, followed by perfect health for 10 years; and scenario B in which he receives no treatment and enjoys his current less-than-perfect health for eight more years. The individual is known to prefer scenario B to A. The utility score associated with state Q_A^1 is assumed to be 0.00 (ie equivalent to being dead) and that associated with Q_B^1 is assumed to be 0.95. The rate of discount is assumed to be 0.10. Mehrez and Gafni point out that the QALY approach would then select scenario A (the QALY total is higher - 5.917 compared to 5.258), thus going against the individual's preferences.

This example is unconvincing. By making an assumption about the rate of discount, the authors implicitly invoke the assumption of intertemporal additive separability. Yet they make an assumption about the individual's ordering of the two scenarios which is incompatible with their assumptions about the values of the discount rate and utility scores. Since additive separability implies that the QALY figures can legitimately be interpreted as utility scores, the individual *must* prefer scenario A to B, if, as is assumed, the discount rate is 0.10 and the utility scores associated with Q_A^1 and Q_B^1 are 0.00 and 0.95. This implies that if the individual really does prefer scenario B to A , and if his utility scores really are 0.00 and 0.95, then his utility function cannot be additively separable, as Mehrez and Gafni implicitly assume.

3.2 Mehrez and Gafni's second example

The second example involves two health states, Q^1 and Q^2, and two scenarios, 1 and 2. In scenario 1 the individual enjoys 5 years in state Q^1 and then dies, whilst in scenario 2 he enjoys 10 years in state Q^2 before dying. Mehrez and Gafni employ the standard gamble to elicit utility scores, obtaining hypothetical values of 0.65 for scenario 1 and 0.50 for scenario 2. They conclude that the individual prefers scenario 1 to scenario 2. They then use these utility scores to compute the QALYs associated with each scenario and, on an assumption of a zero discount rate, obtain QALY totals of 3.25 (=0.65x5) and 5.00 (=0.50x10). They conclude that the QALY approach results in the selection of the non-preferred alternative and hence misrepresents preferences.

This example is also unconvincing for two reasons. First, notwithstanding Mehrez and Gafni's claims to the contrary, it is not possible to infer anything about the individual's ordering of the two scenarios if the utilities from the standard gamble are interpreted as utility scores for the two scenarios. Consider the SG for scenario 1. The individual faces a choice between scenario 1 and another scenario in which he faces either 5 years in perfect health (the probability of this happening being p) or death (the probability of this being (1-p)). Indifference between these two scenarios occurs where $p=0.65$. We thus have

$$(7)\qquad U(Q_1{}^1,\ldots,Q_5{}^1,D_6,\ldots,D_{10})$$

$$= 0.65\cdot U(P_1,\ldots,P_5,D_6,\ldots,D_{10}) + 0.35\cdot U(D_1,\ldots,D_{10}).$$

Mehrez and Gafni implicitly set $U(P_1,\ldots,P_5,D_6,\ldots,D_{10})$ equal to 1 and $U(D_1,\ldots,D_{10})$ equal to 0 to obtain

$$U(Q_1{}^1,\ldots,Q_5{}^1,D_6,\ldots,D_{10}) = 0.65.$$

Likewise for scenario 2 we have

$$(8)\qquad U(Q_1{}^2,\ldots,Q_{10}{}^2) = 0.50\cdot U(P_1,\ldots,P_{10}) + 0.50\cdot U(D_1,\ldots,D_{10}).$$

If we set $U(P_1,\ldots,P_{10})$ equal to 1 and $U(D_1,\ldots,D_{10})$ equal to 0 we obtain

$$U(Q_1{}^2,\ldots,Q_{10}{}^2) = 0.50.$$

From this Mehrez and Gafni deduce that scenario 1 is preferred to scenario 2. This is clearly wrong, the reason being that the two utility scores are measured on different scales. Although the zero point is the same on both scales, the upper fixed point (ie 1.0) is equal to $U(P_1,\ldots,P_5,D_6,\ldots,D_{10})$ on one scale and $U(P_1,\ldots,P_{10})$ on the other. The figures of 0.65 and 0.50 are, in other words, measured in different units and are therefore non-comparable.

In order to be able to say something definite about the individual's rankings of the two scenarios in this example, we require information on the relative values of the reference utilities, $U(P_1,\ldots,P_5,D_6,\ldots,D_{10})$ and $U(P_1,\ldots,P_{10})$. Evidently, for scenario 1 to be preferred to scenario 2 we require that the ratio of $U(P_1,\ldots,P_5,D_6,\ldots,D_{10})$ to $U(P_1,\ldots,P_{10})$ be greater than 0.77 (=0.50/0.65). All that can be said *a priori* is that $U(P_1,\ldots,P_5,D_6,\ldots,D_{10})$ is likely to be smaller than $U(P_1,\ldots,P_{10})$, and therefore the ratio of $U(P_1,\ldots,P_5,D_6,\ldots,D_{10})$ to $U(P_1,\ldots,P_{10})$ is likely to be less than one. Therefore nothing can be concluded *a priori* about the individual's ordering of the two scenarios.

The second flaw in Mehrez and Gafni's argument lies in their subsequent use of the utility scores to calculate QALY scores: one moment they interpret the utility scores as indicating the utilities associated with the two scenarios; the next they interpret them as indicating the utilities associated with the two health states. This is inconsistent: the scores cannot indicate both things simultaneously. The first interpretation is perfectly legitimate but, as has been shown above, the scores cannot be compared since they are measured in different units. The second interpretation is, however, the more obvious interpretation to persons familiar with QALYs. Its legitimacy rests, however, on the validity of the assumption of additive separability. If we impose additive separability and a zero discount rate, eq (7) becomes

(7') $\quad u(Q^1)\cdot 5 = 0.65\cdot u(P)\cdot 5 + 0.35\cdot u(D)\cdot 5$

where $u(Q^1)$, $u(P)$ and $u(D)$ are the utilities associated with being in the states Q^1, 'perfect health' and 'dead' for one year. If we set $u(P) \equiv 1.0$ and $u(D) \equiv 0.0$, then the utility stream associated with scenario 1 is 3.25 and the utility associated with being in Q^1 for one year only is 0.65. By analogy the utility stream associated with scenario 2 is 5.00 and the utility associated with being in state Q^2 for one year is 0.50. The implication is that the individual thus prefers a year in state Q^1 to a year in Q^2 but insufficiently so to prefer scenario 1 to scenario 2.

Whether this correctly represents his 'true' preferences (which, in this case, as indicated above, are unknown without further information) depends, obviously, on the validity of the assumption of additive separability. If this assumption is invalid, the QALY approach *may* get the rankings of the two scenarios wrong, but since not all non-separable utility functions are likely to generate this result, it is by no means certain.

Mehrez and Gafni do, in fact, address the point that the utility scores derived from a SG might alternatively be interpreted as the utilities associated with being in a particular state rather than as utility scores for a scenario. They claim, however, that this leaves their argument unaffected. Suppose, they suggest, an individual is known to prefer scenario 1 to scenario 2. Suppose too we measure the utilities for each state (ie being in each state for one year) at 0.60 and 0.40. Then it is clear that the QALY approach will choose scenario 2 to scenario 1, even though the individual prefers scenario 1.

This argument, like the previous arguments, hinges on the issue of intertemporal additive separability. If the individual's utility function is additively separable, if the discount rate is zero, and if the utilities associated with states Q^1 and Q^2 are indeed 0.60 and 0.40, the individual cannot prefer scenario 1 to 2. Under these assumptions the QALY approach correctly represents the individual's preferences. It would also do so if the discount rate were non-zero, although obviously a non-zero discount rate would have to be entered into the QALY calculation. The key issue, then, is additive separability: if this assumption is invoked but is invalid, the QALY approach *may* misrepresent preferences.

4. Conclusions

Our conclusions can be summarized briefly. The crucial distinction between QALYs and HYEs is that the former assume intertemporal additive separability of the utility function, whilst the latter do not. If intertemporal additive separability *is* acceptable, the two approaches will give identical rankings of projects, since the QALY total is a monotonically increasing function of the number of HYEs. Indeed, if a zero discount rate is used, the numbers of QALYs and HYEs will be identical.

We have argued that intertemporal additive separability also lies at the heart of Mehrez and Gafni's examples involving QALYs and preference reversals. Such reversals can occur only if the assumption of intertemporal additive separability invoked in the QALY approach is untenable. If this assumption is acceptable, the various assumptions made by Mehrez and Gafni in their examples are simply mutually inconsistent and do not entail preference reversals.

When Mehrez and Gafni claim, then, that QALYs do not derive from individuals preferences and hence tend to misrepresent these preferences, they are implicitly claiming that the assumption of intertemporal additive separability is, at least in the health context, unacceptable. In fact, they are implicitly claiming more than this, since non-separability is a *necessary* condition for such reversals, not a *sufficient* condition. Surprisingly, they offer no empirical evidence to show that people do, in practice, violate the assumption of intertemporal additive separability, let alone that they violate it in such a way as to make preference reversals commonplace. What they offer instead is a series of examples in which it is *assumed* that such reversals occur!

We conclude, therefore that, unless there is evidence that the assumption of intertemporal additive separability *is* frequently violated in such a way as to make use of QALYs misleading, the principle of parsimony of assumptions suggests that the profession may as well stick with QALYs rather than adopt the more complex methodology of HYEs.

Table 1: QALYs and HYEs associated with contrast media

	Goel et al.				Gafni & Zylak		Gafni & Zylak revised	
	Ionic probability	Non-ionic probability	Ionic QOL	Non-ionic QOL	Ionic HYEs	Non-ionic HYEs	Ionic HYEs	Non-ionic HYEs
No reaction	0.949875	0.994988	1.00000	1.00000	29.99800	29.99900	30.00000	30.00000
Minor reaction	0.050000	0.005000	1.00000	1.00000	29.98000	29.98000	30.00000	30.00000
Major reaction	0.000100	0.000010	0.70000	0.70000	21.00000	21.00000	21.00000	21.00000
Fatal reaction	0.000025	0.000003	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
QALYs/HYEs			29.99835	29.99984	29.99545	29.99874	29.99835	29.99984
Difference				0.00148		0.00329		0.00148

Table 2: QALYs and HYEs with discounting (i=10%)

	Goel et al.		Gafni & Zylak	
	Ionic QALYs	Non-ionic QALYs	Ionic HYEs	Non-ionic HYEs
No reaction	10.36961	10.36961	29.00007	29.00007
Minor reaction	10.36961	10.36961	29.00007	29.00007
Major reaction	7.25872	7.25872	10.31534	10.31534
Fatal reaction	0.00000	0.00000	0.00000	0.00000
QALYs/HYEs	10.36904	10.36955	28.99747	28.99981
Difference		0.00051		0.00233

References

Gafni, A., 1989, The quality of QALYs (quality-adjusted life-years): do QALYs measure what they at least intend to measure?, Health Policy 13, 81-83.

Gafni, A. and C.J. Zylak, 1990, Ionic versus nonionic contrast media: a burden or a bargain?, Canadian Medical Association Journal 143, 475-478.

Goel, V., R.B. Deber and A.S. Detsky, 1989, Nonionic contrast media: economic analysis and health policy development, Canadian Medical Association Journal 140, 389-395.

Kind, P., R. Rosser and A.H. Williams, 1982, Valuation of quality of life: some pyschometric evidence, in: M. Jones-Lee, ed., The value of life and safety: proceedings of a conference held by the Geneva Association (North-Holland, Amsterdam).

Mehrez, A. and A. Gafni, 1989, Quality-adjusted life-years, utility theory and healthy-year equivalents, Medical Decision Making 9, 142-149.

Torrance, G.W., 1986, Measurement of health state utilities for economic appraisal, Journal of Health Economics 5, 1-30.

Williams, A.H., 1985, Economics of coronary artery bypass grafting, British Medical Journal 291, 326-329.