

## Analysing health outcomes

Jack Dowie London School of Hygiene and Tropical Medicine, London

### Abstract

*If we cross-classify the absolutist-consequentialist distinction with an intuitive-analytical one we can see that economists probably attract the hostility of those in the other three cells as a result of being analytical consequentialists, as much as because of their concern with “costs”. Suggesting that some sources of utility (either “outcome” or “process” in origin) are to be regarded as rights cannot, says the analytical consequentialist, overcome the fact that fulfilling and respecting rights is a resource-consuming activity, one that will inevitably have consequences, in resource-constrained situations, for the fulfilment of the rights of others. Within the analytical consequentialist framework QALY-type measures of health outcome have the unique advantage of allowing technical and allocative efficiency to be addressed simultaneously, while differential weighting of QALYs accruing to different groups means that efficiency and equity can be merged into the necessary single maximand. But what if such key concepts of the analytical consequentialist are not part of the discursive equipment of others? Are they to be disqualified from using them on this ground? Is it ethical for intuition to be privileged in ethical discourse, or is the analyst entitled to “equal opportunities” in the face of “analysisism”, the cognitive equivalent of “racism” and “sexism”?*

(Journal of Medical Ethics 2001;27:245–250)

**Keywords:** Health economics; utility; analysisism; rights; QALY

The vast majority of jobbing health economists are now involved in *economic* evaluations, undertaken as a complement or “add-on” (even afterthought) to an evaluation undertaken in some other—and, by implication, “non-economic”—way. The most common situation occurs when a clinical trial of an intervention is performed and the health economist is asked to address the question of whether, given that the outcomes indicate the new intervention is “effective”, it is also *cost-effective*. The contrast between “clinical effectiveness” and “cost-effectiveness”, now a feature of health care discourses in most countries, is accordingly created. In separating evaluation for *scientific* purposes on the one hand and *policy* purposes on the other, this distinction between “clinical” and “economic” evaluation greatly exacerbates the difficulty of conducting ethical debate in the health care context.

### Science and policy

Evaluation for scientific purposes is fundamentally different from evaluation for decision making purposes. The standards for deciding whether something is true (do we *know* this?) are quite legitimately very different from those appropriate for choosing between alternative actions (*should* we do this or that?). Broadly speaking, clinical evaluation seeks to relate outcome to process in a way that will enable us to test the hypothesis that the new intervention (process) will have better results (outcomes) than the old by *the standards of science*. Economic evaluation, on the other hand, relates outcome to process in a way that will help us decide what to do (for example, make the intervention available in the NHS), using *the standards and principles appropriate to policy decisions*.

Two key differences epitomise the conflict, both of which clearly affect outcome definition and assessment in health care.

In deciding whether something should be accorded the status of knowledge it is conventional in scientific research, such as randomised controlled trials (RCTs), to seek to avoid a type 1 error (not rejecting a false hypothesis) “at all costs” compared with a type 2 error (rejecting a true one). Since placing an *infinite* disutility on this outcome would be an “impractical” criterion for knowledge—we would rarely “know” anything—the conventional statistical tests are set at a level which is intended to rule out any “real” chance of the new intervention being accepted as better when it isn’t. But, as we well know, many patients, because of their clinical condition and prognosis, would be very willing to take new drugs or undergo innovative techniques that have not been shown to be effective or “safe” by the standards of science in clinical trials. They are quite rational in this attitude, because, in deciding what to do, it is the relative error burdens experienced in the real world that are relevant, ie what will follow if drug x is taken if it is actually not effective compared with what will happen if it is not taken if it is actually effective. It is the real world relationship (ratio) of the consequences of false positive and false negative errors that matter to decision makers. That ratio will almost never be anything like the infinite one which is imposed—legitimately—in the pursuit of “the truth”.<sup>1</sup>

The second key difference lies in attitude to time. Science, legitimately using *eternal* standards in its pursuit of knowledge, does not even contemplate discounting time. Decision makers, however, must always consider doing so. Patients and populations

under threat of mortality or continuing morbidity cannot afford to place equal value on future days and months, let alone future centuries.

### Intuition and analysis

If we concentrate too much on the simple fact that economic evaluations are concerned with costs and cost-effectiveness, whereas clinical evaluations aren't, we are in danger of overlooking a more profound way in which health economists differ from most non-economists involved in health care discourses. While everyone is in favour of "evaluation" in some abstract way, the actual implementation of this task can be tackled in more or less analytical ways<sup>2</sup> and hence more or less "transparently". The transparent and analytical approach to evaluation characteristic of the economist's approach has probably been as much responsible for their treatment (by some) as ethical outcasts, as their insistence that costs be taken into account.

It is often useful (and certainly tempting) to contrast extreme positions, even if they are knowingly unreasonable caricatures. Here are two relevant caricatures talking about each other in the health context.

*The ethical intuitionist:* One can't be truly ethical in assessing whether something is fair and just if one is "cold and calculating" and engaged in "measuring the unmeasurable"—such as the value of life or health to human beings. A judgment, decision or outcome in this area either "feels right" or it doesn't and no amount of analysis, especially quantitative and mathematical-statistical analysis can get around that fact. Indeed, bringing these things out into the open and analysing them with such approaches and techniques undermines and destroys the fundamentally intuitive basis of ethical behaviour and standards. It does so partly because the average reasonable person does not have the competencies necessary to comprehend fully what is going on inside the "black box" analysis, but much more because human beings intuitively understand that the underlying complexities and clashes cannot be resolved, or even greatly illuminated by, what claims to be "rational" analysis.

*The ethical analyst:* One can't be truly ethical in assessing whether something is fair and just without being highly analytical and detached. It is essential to measure and calculate, in open and defensible ways, precisely in order to establish whether what we have done, or are doing, is fair and just. No amount of human intuition can get round the fact that human beings have an almost infinite capacity to interpret things in ways which favour themselves or those to whom they relate closely. Society can function only if fairness and justice not only exist between individuals and groups, but are *seen* to exist. This criterion of transparency can be met only if we move to a level of analysis, including measurement and calculation of outcomes, which is requisite to the task. Moreover, it has to be pointed out that the intuitions of the intuitionists will

necessarily reflect *implicit* quantification of the unquantifiable and *implicit* calculation of the incalculable—so the major difference between us is that they can inspect our measurements and calculations, whereas we can't see into their even blacker black box. Basically if the central ethical questions are not addressed "full frontally" they are really not being addressed at all.

Most economists would feel that, in fulfilling their professional duty of raising the analytical level of debate about decisions and policies they are, above all, increasing their transparency. Dismal and unrewarding as this task is, they are simply uncovering the hard, even tragic choices, that are inherent in the operation of a resource-constrained health care system. Most health economists see themselves as having the same motivation as health care practitioners—to save lives, reduce mortality and improve the health-related quality of life of fellow human beings—by helping to ensure that resources are not wasted by decisions which lead to worse outcomes than could be achieved.

The intuitive-analytical distinction is not to be confused with that between the major schools of thought in ethics—the absolutists (among whom the deontologists are the majority) and the consequentialists (of whom utilitarians of various sorts constitute the vast bulk). One can be an analytical deontologist or an intuitive utilitarian. If we cross-classify the two distinctions we can suggest that economists attract the hostility of the inhabitants of the other three cells of the 2 x 2 table as a result of being analytical consequentialists.

### Clinical effectiveness and cost-effectiveness

Clinical researchers evaluate a new drug by its clinical effectiveness, in other words its outcome in terms of its "effect" on patients. A drug or intervention is more effective if it has a better outcome. The outcome may vary from an *intermediate* biophysical marker, such as a cell count or blood pressure measure, to a "final" *disease-related* state such as mortality from cancer or myocardial infarction. Recently there has been some movement towards the use of final outcomes that reflect the *health-related* quality of life of the patient as well as mortality or life expectancy, such as Quality Adjusted Life Years (QALYs).

Economic evaluators suggest there is something missing in this notion of clinical effectiveness, even if it is extended to cover a wider range of health-related outcomes in the individual. If clinical trials are funded to determine whether or not an intervention should be introduced, there are actually *two* sorts of relevant outcome to consider. One is the effect on the patient receiving the intervention. The other is the effect on other patients in the system who do not receive the benefits they could otherwise have received had these resources not been so deployed. In a resource-constrained publicly-funded health care system this cost, while conventionally phrased and measured in terms of money (financial cost) is more appropriately

conceptualised as the *benefits forgone* by that patient (or those patients) who would have benefited most from having the resources devoted to them instead. Only if this *opportunity cost* is less than the benefit received by those who do receive the resources are we in an optimal—and it can be argued, ethical—situation as far as outcomes are concerned. So the economist's conclusion, though rarely stated this bluntly, is that the only comprehensive form of evaluation in this context is a cost-effectiveness one. Any other form (for example clinical evaluation) can only be regarded as a limited and partial evaluation.

The response to this will usually be along the lines that one should, in a clinical study, seek only to establish the *scientific basis* for clinical or health service practice and leave the things that can't be done "scientifically" to the decision or policy maker (for example clinician, health authority, National Institute of Clinical Excellence). These things will include questions of cost, however conceptualised. In terms of what was said earlier, however, this is implying that we should analyse *one part* of the outcome question at an extremely rigorous level and then hand over the results and the rest of the decision making to people operating at a highly intuitive level. But who needs to know the effect of a drug on the blood pressure of a patient with great rigour and precision without knowing, with equivalent rigour and precision, the effect on other patients' health and health care as a result of the resources being used in providing it to the first patient? The economist finds it hard to come up with a reason for not saying "nobody" and suggests, further, that failing to address this latter question equally analytically compromises the quality of decision making and its ethical status.

### Outcome and process utility

Even if we accept all this, the various outcome consequences still have to be identified and measured in a way that has operational potential. So while economists agree that the task of evaluation should be undertaken analytically and must embrace the concept of opportunity cost in the search for optimality, the *maximand*—the thing to be maximised in this search for optimality—is still open. It is here where many of the relevant debates *within* health economics—echoing those outside it—are located.

I select just one of these debates for attention here. On one side of it are those economists who suggest that policy makers in a resource-constrained publicly funded health care system should be allocating the health care budget in such a way as to maximise the *health* or *health gain* achieved. This will still require definition and measurement of health and the development of methodologies to establish the bases for such measurement, for example the relative weights to be attached to different health limitations and the relative importance to be attached to treating different groups of people. But within this position all are united in agreeing that the *health budget*

should be used to produce *health outcomes*. The purpose—and duty—of the health care system is to maximise the health gain of the population from the finite resources available to it.

To those on the other side of the debate the purpose is to maximise the amount of health care *services* provided, not the amount of health generated by them. While most health gain measures may incorporate "normal care" satisfactorily, their use will not necessarily lead to the supply of all the services that people want and expect—or supply them in the way they want them provided. In the extreme a service may be almost purely symbolic but still valued.<sup>3</sup> In other words, resources may be seen as properly used in a way which is deliberately non-productive in health or health gain terms—symbolically—in order to satisfy some "right" or "duty" or "obligation". Economists supporting this view could be interpreted as seeking to bring within the consequentialist framework actions of (only apparently) "wasting" resources in a display of affirmation of the human spirit, community spirit, or whatever.

### Very strange road

While tempting, for the majority of economists the second position takes us down a very strange road indeed. What services should we provide? "All that are wanted or demanded" cannot be the answer, because either will exceed the resources available and the question of allocating those resources among the competing wants or demands will still arise—followed by the question of whose should carry the greatest weight? It is important to emphasise that the economists who are attracted to this position (in contrast to non-economists, for whom it is almost their natural home) do not seek to avoid or deny the allocation problem. What they are saying is that in allocating the health care budget it is necessary to use a maximand that trades off health gain against other aspects of health care service and provision that people value. For example, people *may* prefer easier access to a local facility which cannot aspire to the standards of care (in terms of health gain outcome) of a centre of excellence, to distant access to the latter. Or, since this example confounds the question of costs to patients with differential effectiveness, patients *may* be willing to trade off health gain against time spent being counselled, provided with information or being "given a sense of autonomy". The excellent surgeon may be able to increase the number of successful operations done by 10% if he or she spends 50% less time engaged in these functions with each patient. But the latter may be preferred. (It is important to see that even if health gain in terms of successfully operated on and recovered patients is correlated with information/counselling and feelings of autonomy—which will not necessarily be the case—this sort of trade-off may still be required.)

Proponents of this position have used the term "process utility" to emphasise that utility/value can arise from the way things are done, as well as from their results (which, in contrast, yield "outcome

utility”). But it would be preferable, if clumsier, to talk of “non-health gain outcome utility”. Why? Because if our aim is to evaluate *processes* in terms of the relationship between their outcomes and their costs (forgone benefits to others), it is not coherent to seek to evaluate processes *as processes*. And there is no need. We can simply recognise that non-health outcomes may be important and that identification and measurement of these non-health outcomes can and should be part of the evaluation. The task is then clear and coherent. Having established the trade-off between various health and non-health outcomes, we can ensure that resources are used in optimal fashion according to some mixed *health and non-health* maximand. The fact that some sources of “process utility” are regarded by many as “rights” cannot, says the analytical consequentialist, overcome the fact that fulfilling and respecting rights is usually a resource-consuming activity, one that will inevitably, in resource-constrained situations, have consequences for the fulfilment of the rights of others.

### Technical and allocational efficiency

Economists distinguish between technical efficiency and allocational efficiency. Technical efficiency is inherently partial. One can be technically efficient (or inefficient) only in relation to a specified task, such as reducing blood pressure or cancer mortality. Most economic evaluations are technical in the sense that they are seeking to establish whether intervention A is technically more efficient than intervention B in achieving some specified outcome (for example hip replacements, heart transplants, hours of geriatric care). But even if a society is operating at maximum *technical* efficiency in producing all the specific things it produces from the health care budget the question remains: is it producing the “right” amounts of these specific things. In other words is there *allocative* efficiency in the distribution of resources between the competing uses? It is entirely conceivable that a technically efficient procedure should be allocated no resources, because, despite being the best procedure for its task, it may generate such small benefits at such high cost that it would be allocationally inefficient to fund it.

Health care practitioners have traditionally avoided any concern with allocative efficiency, except in so far as they, very naturally, campaign for more resources for their area—either on the grounds of their technical efficiency (which should be exploited more) or on the grounds of their technical inefficiency (which requires resources if it is to be improved).

Tackling the allocational question *analytically* means we must be able to measure outcomes and resource usage of all procedures/services in a way that will enable cross comparisons to be made. If each area/service/procedure uses its own disease-specific or procedure-specific outcome measures we will simply end up with a long list which tells us the cost of a myocardial infarct or headache prevented, a hip or heart replaced, an episode of

depression avoided, and so on. Evaluations of *technical* efficiency using these condition-specific outcome measures will enable us to spend the budget set aside for each condition in a way that produces most benefits (given acceptance of the measure) but it will do nothing to answer the question as to how much should go to each in the first place.

Fortunately most health care procedures have the common final goal of *health*, something which is wider than the presence or absence of a particular disease, or the state of a particular disease indicator. Once it is accepted that “health” is the relevant goal we can proceed to devise an operational definition of health outcome. There is little disagreement that two sorts of health outcome are relevant—quantity and quality of life. Length of life is not problematic but for the quality component two sorts of health outcome measure, usually referred to as the psychometric and the economic, are on offer. Both are typically multidimensional, encompassing various aspects of some construct of health or health-related quality of life—such as ability to carry out normal activities, pain and discomfort, physical mobility, and depression. The key difference is that the psychometric measures (for example SF36) typically produce only a *descriptive* profile—a set of scores on various dimensions of health. The economic measures (for example the QALY) are designed to produce the precise inputs needed for economic evaluation, inputs which necessarily involve *relative valuation* of the different aspects of health and aggregation of the disparate aspects into a single *index* of health.

### Equity and the QALY

QALY-type measures have the unique advantage of allowing technical and allocative efficiency to be addressed simultaneously. How do they achieve this? Simply by saying that most health care services and procedures are producing the same thing, not different things. That thing is health, conceptualised as years of life weighted by the health-related quality of life experienced in these years.

Some of the most heated controversies surrounding the QALY approach are essentially methodological debates within that approach, but several have undoubted ethical significance. We can highlight only two.

First, in seeking the quality adjustments, *whom* should we ask? Should we seek the health state preferences of the general population, most of whom will have not experienced (yet) many of the health states that are being put before them? Or should we seek the preferences of those who have some knowledge and/or experience of them: patients who have experienced them—and adapted to them—on the one hand, or health practitioners (doctors, nurses) who have experienced them indirectly (but over a much greater range of cases than any individual patient) on the other? The question is clearly important only if the preferences of these groups diverge. Empirical studies confirm such variation exists, with patients, public and practitioners all returning different ratings from each



other, including nurses differing significantly from doctors. Effectiveness and cost-effectiveness can hinge on which set is used.

Second, having arrived at the quality adjustments, should we treat a QALY as a QALY as a QALY to whomsoever it accrues? Empirical work within the QALY approach has been dominated by what would be called correctly the *equally*-weighted QALY and incorrectly the *unweighted* QALY. In this a QALY accruing to any individual is regarded as of equal worth, whatever the personal characteristics of that individual.

There is, however, no arithmetic difficulty in adapting the concept to any other system of (unequal) weights, with QALYs accruing to different people weighted differently according to their characteristics, such as age, sex or race—or how many QALYs they have already enjoyed during their lifetime.

Within any unequal weighting approach, practical issues of coherence will arise because individuals have multiple characteristics. For example, some might desire to allocate resources in such a way as to favour young people as against old people (age-weighting), those with more severe conditions as against those with less severe ones (severity-weighting), and those with particularly dreaded conditions against those with less feared ones (dread-weighting). The practical difficulty of achieving a coherent overall allocation will increase exponentially with the number of characteristics on which such differential weighting is sought and in the end these practical difficulties may rule out anything other than equal weights.

### Personal characteristics

Leaving practicalities aside, what (if any) personal characteristics should be regarded as acceptable *ethical* bases for weighting? Age has been at the heart of recent public debate about QALYs and can be used to exemplify the issue. If QALYs are equally weighted it is a simple matter of arithmetic to show that the older someone is the less QALYs can be generated by any health care, other things (for example cure potential) being equal. Is this discrimination against the old? In a formal sense it may be, but it is nature, rather than QALYs, which is doing the discriminating. We could decide to weight QALYs inversely by age in such a way as to offset the age effect precisely, but formally this would only be replacing one form of discrimination by another. That is not an argument for not doing it, but it is an argument for not denying that discrimination (distinguishing between patient groups) is what is being sought when the equally weighted QALY is attacked as unethical because it is “ageist”. (By a similar argument equally weighted QALYs are “smokerist” but weights which attempt to offset this are discriminatory against non-smokers.)

Quality adjusted life years are here being used as measures of *capacity to benefit* from health care and thus, in equally weighted form, reflect all differences in that capacity among individuals. Logically,

any attempt to give greater weight to those with below average capacity to benefit will lead to less health gain in the population as a whole compared with that achieved when every individual is regarded as equally important. To the analytical consequentialist the QALY approach does not create these problems of equity—and issues of weighting—it simply makes them transparent and addressable.

### Conclusion: ethics as deliberative discourse

It would be nice to think that one could clearly separate out a set of specific ethical issues arising in outcome measurement, distinguishing them from the more general issues arising out of the design and operation of resource-constrained, publicly funded health care services. But if one thing emerges from this brief and highly selective discussion it is that it clearly is not possible.

For recent writers such as Seedhouse,<sup>4</sup> who take the Aristotelian line that, at its heart, ethics is a deliberative *process*, the key ethical question is how can that deliberative process be fair and just. In other words, how can what Habermas<sup>5</sup> conceptualises as the “ideal speech situation” be ensured?

This is an enormous question, but any acceptable answer must surely contain the requirement that one form of discourse is not privileged over another. In relation to a basic theme of the present paper it would seem to be unacceptable that intuition be privileged over analysis. Of course this raises a problem. If the analyst, on the basis of much thorough reflection and study, has arrived at concepts that are not part of the discursive equipment of others, is he or she to be disqualified from using them on this ground? This, it must be stressed, is not a question of asking for analysis to be privileged, only of asking that intuition not be privileged. Take QALYs. Most analyst consequentialists, whether in the end they end up being “for” or “against” QALYs as outcome measures, have spent considerable time getting to grips with the underlying theories and concepts, principles of measurement and calculation, and so on. The intuitionists tend to argue that the analytical case must be conveyed successfully *in intuitive terms* or else the analyst fails. But that involves allowing the intuitionists to be judge and jury in their own case. The analyst is surely entitled to “equal opportunities” in the face of such “analysisism”, the cognitive equivalent of “racism” and “sexism”.

A particular threat to equal opportunities arises from the operation of double standards. Explicit analytical approaches are ruthlessly examined and found to have many flaws and difficulties—as they undoubtedly do have. The conclusion is drawn that they are therefore “unacceptable”. However, the implicit intuitive approaches (whose weakness and flaws were usually the historical stimulus to the development of the more analytical alternative) are left alone, not being subjected to the same ruthless examination. In fact they are largely immune from it because they can be inspected only with great

difficulty. (Would allocation by equally weighted QALYs be more “ageist” than current practice, if we could establish the principles on which the latter operates?) Logically, the demonstration of weaknesses and limitations within one approach establishes *nothing* about its comparative worth compared with another. No evidence-based clinician would suggest that a new drug is unacceptable because it cures only x% of patients when he or she has no idea how many patients the existing alternative treatment cures. But such a combination of “double standards” and the “nirvana trap”—the use of perfection as a test of other’s approaches but not one’s own—is disturbingly prevalent in so-called ethical debates concerning health care evaluation and decision making.

*Jack Dowie is Professor of Health Impact Analysis at the London School of Hygiene and Tropical Medicine, London.*

### References

- 1 Scheff TJ. Decision rules, types of error, and their consequences in medical diagnosis. *Behavioral Science* 1963;8:97-107.
- 2 Hammond KR. *Human judgment and social policy: irreducible uncertainty, inevitable error, unavoidable injustice*. New York: Oxford University Press, 1996.
- 3 Mooney G. Beyond health outcomes: the benefits of health care. *Health Care Analysis* 1998;6:99-105.
- 4 Seedhouse D. *Ethics: the heart of health care*. Chichester: Wiley, 1998.
- 5 Habermas J. *Theory and practice*. London: Heinemann, 1971.

## 1st Asia Pacific Forum on Quality Improvement in Health Care

Three day conference

Wednesday 19 to Friday 21 September 2001

Sydney, Australia

We are delighted to announce this forthcoming conference in Sydney. Delegate enquiries are welcome.

The themes of the Forum are:

- Improving patient safety
- Leadership for improvement
- Consumers driving change
- Building capacity for change: measurement, education and human resources
- The context: incentives and barriers for change
- Improving health systems
- The evidence and scientific basis for quality improvement.

Presented to you by the BMJ Publishing Group (London, UK) and Institute for Healthcare Improvement (Boston, USA), with the support of the the Commonwealth Department of Health and Aged Care (Australia), Safety and Quality Council (Australia), NSW Health (Australia), and Ministry of Health (New Zealand).

For more information contact: [quality@bma.org.uk](mailto:quality@bma.org.uk) or fax +44 (0)20 7383 6869