Musings on the meaning of incremental opportunity cost

Ken Buckingham

Summary

- Uncontroversially, if a treatment provides health benefits at a resource cost, the opportunity cost is the value of resources used (as the loss to society of alternative uses of those resources).
- More controversially, if a treatment saves resources (thus providing the potential benefits elsewhere) at the expense of a loss in health, that loss in health is the opportunity cost of the benefits achieved elsewhere.
- It is a mistake to define the resources as opportunity cost; sometimes they are, sometimes they are not.
- While resource use can be negative, opportunity cost cannot, since the definition of opportunity cost implies an opportunity forgone, ie a loss.
- A ratio statistic of resource consequences and health effects using a consistent definition of opportunity cost is suggested.
- Both gains (whether health improvements or cost savings) and opportunity costs (whether as a loss in health or a loss of resources) are denominated using a common numeraire.
- The ratio of gains to losses is calculated and is referred to as the ‘payback ratio’.
- This procedure is analogous to net benefit calculations in which the losses are effectively subtracted from the gains.

Section 1 Introduction

This paper discusses two important concepts and their joint interpretation in health care decision-making: the concept of opportunity cost; and the concept of incremental analysis[1]. It is structured as follows: Section 2 defines the terminology. Section 3 considers the application of opportunity cost. Section 4 discusses and raises some related issues. Section 5 concludes.

While the concepts of opportunity cost and incremental analysis are fundamental to health economics, it is difficult to find papers that consider their joint interpretation, ie “incremental opportunity costs”. The definition of this term may appear trivial, this is not the case, by failing to apply a rigorous definition we face difficulties in the interpretation of findings where savings are achieved at the expense of a benefit forgone.

A Medline search on the terms “incremental” AND “opportunity” AND “cost” revealed just 8 citations. Two raised issues relating to costing methodology[2 3]. None considered the definition of “incremental opportunity cost”. The conventional view is that cost relates to resources, and effectiveness relates to health. Under this view, savings are simply ‘negative costs’, and health reductions are ‘negative health gains’. The idea that resource savings may be obtained at the expense of a reduction in health, and hence that this reduction in health is the true opportunity cost, does not fit easily within this approach.

However, several authors[4-8] have noted conceptual problems relating to Incremental Cost Effectiveness Ratios (ICERs). Heitjan et al[6] note that,

> ‘The ICER is an ill-defined parameter because it identifies pairs of cost and effectiveness differences that can have altogether different interpretations.’

They demonstrate the construction of confidence limits for an ICER,

> ‘.. that lies in a particular quadrant, thus avoiding the problems of mixing ratios that lie on opposite sides of the vertical axis.’

Laska et al[7] note that,

> ‘Thus, knowledge of the value of the magnitude of the cost-effectiveness ratio alone is insufficient to ascertain which treatment is preferred. To make that determination, the sign of ε (the incremental measure of effect) or of γ (the incremental measure of cost), must also be known and φ (the incremental cost effectiveness ratio) must be compared with φ_1 (the cost effectiveness ratio of the treatment indexed 1) …….’

A similar point is made by Stinnett et al[4],

> ‘Thus, R hat (the estimator for the cost effectiveness ratio) has no meaningful interpretation unless it is presented in the context of the quadrant of the ΔE - ΔC plane to which it corresponds.’

Section 2 Defining the terms

Health effects are assumed to be measured in Quality Adjusted Life Years (QALYs) and valued at the maximum price society is willing to pay for them. This societal willingness to pay for a QALY is denoted by ω, to avoid confusion with the Lagrangian multiplier (λ) used in constrained optimization definitions of shadow price. The distinction between societal willingness to pay and ‘shadow price’ originating from constrained optimization problems is considered in Section 4.2.3.

Health care decision analysis usually adopts an incremental approach which considers the change that results in implementing one treatment in preference to another. Incremental analysis is adopted because decision-making is about making choices between alternatives either doing something rather than doing nothing, or doing something in place of something else.

According to standard economic definitions (eg Pearce[9]) opportunity cost is,

> “Perhaps the most fundamental concept in economics.”

It is defined as,

> “ … the value of the foregone alternative action.”

Given the fundamental importance of opportunity cost, we need to consider how it is applied in the context of incremental health care decision-making.

While arithmetic and algebra cope well with negative quantities, confusion can arise in text. For this reason 'savings' is used to describe negative resource consequences and 'dis-benefits' is used to describe negative health effects. We use 'gain' to denote either a positive health effect or a negative resource consequence. This is contrasted with 'opportunity cost', which is used to denote either a positive resource consequence (the forgone benefits that those resources might have obtained) or a negative health effect (as when we forgo a health effect to achieve a saving). The use of the phrase 'opportunity cost' in this context is justified below. The principal difference from the usual health economic terminology is the use of opportunity cost in its classical economic sense as a benefit forgone, thus including a health benefit forgone to achieve a saving.

Denote treatments by T_i (subscripts denote different treatments, with subscript 's' denoting the standard treatment). Denote cost by C_i, opportunity cost by O_i, health effect by E_i, the monetary value of the effects by $\omega \times E_i$, and the 'gain' associated with a decision by G_i: Where we wish to emphasise that the source of an opportunity cost is a reduction in health we denote this (dis-benefit) by D_i ($-E_i = D_i$). Similarly, where we wish to emphasise that the source of a gain is a resource saving we denote a negative cost (saving) by S_i ($-C_i = S_i$). Incremental changes are denoted by Δ. **Table 1** summarises this terminology.

Table 1 Terminology

Term	**Meaning**	**Comments on its use**	**Notation**
Resource consequences	The resource consequences of a decision	Can be either positive or negative	C
Costs	The resources required to implement a decision	Only used in the text to relate to positive costs. Where emphasis is required the phrase 'resource cost' is sometimes used.	+C
Saving	A reduction in the use of resources resulting from implementing a decision	Used in preference to negative costs.	S
Health Effects	The health consequence of a decision	Unlike 'costs' the word 'effect' has neither positive nor negative connotations. Positive effects are referred to as 'health benefits' and negative effects are referred to as 'dis-benefits'.	E
Health Benefit	A positive health consequence of a decision	Used where we wish to emphasise a positive health effect.	+E

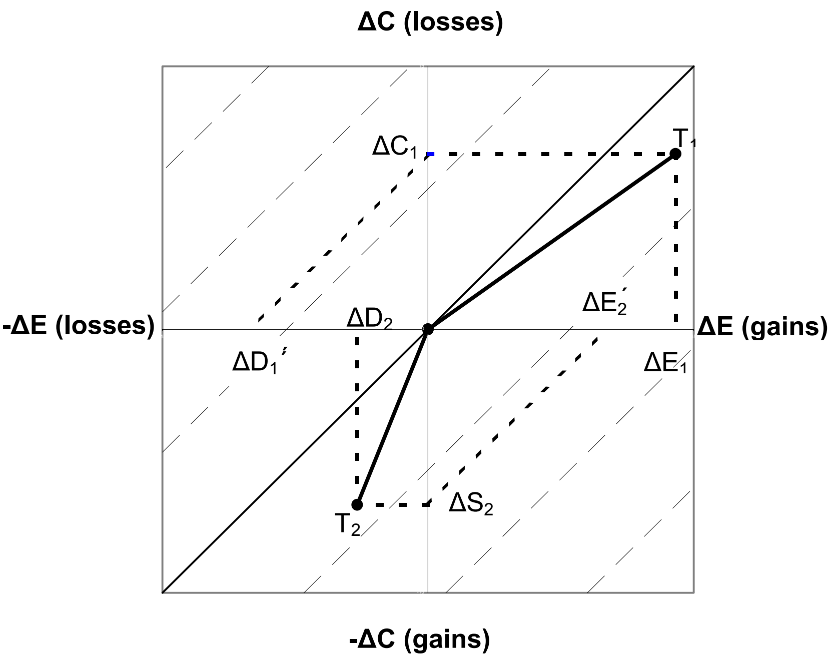
Term	**Meaning**	**Comments on its use**	**Notation**
Dis-benefit	A negative health consequences of a decision	Used in preference to negative effects.	D
Opportunity cost	The value of the forgone alternative action	As defined in general economic texts and not simply the resources used to achieve a health gain.	O
Loss	The value of the alternative action	Used as an alternative to opportunity cost as possibly less contentious.	O
Gain	The advantage for which the opportunity cost is forgone	Hopefully a reasonably intuitive concept.	G
Societal value of a health gain	The social willingness to pay for a health gain	Compared and contrasted with the shadow price from constrained optimization.	ω
Shadow price	The addition to the objective function that results from a slackening of a constraint.	Compared and contrasted with societal willingness to pay for a health gain.	λ

Section 2.1 Incremental opportunity cost

Consider the two treatments, T_1 and T_2 in relation to a standard treatment T_s. T_1 results in an increase in health in comparison with T_s, and an increase in resource use. T_2 results in a reduction in health in comparison with Ts and a saving in resources.

Incremental analysis involves 'bundling' the effects of a decision. The decision to implement T_1 (in place of T_s) results in a health benefit E_1 and the loss of the health benefit $-E_s$ from the discontinued treatment T_s to give an incremental ('bundled') health benefit equal to ΔE_1. The decision also results in the resource cost C_1 for the implementation of T_1 against which we can set the saving in resources C_s from the discontinued treatment T_s to give an incremental ('bundled') value of the resources used equal to ΔC_1. We can describe ΔE_1 as the net gain from the decision. The opportunity cost is ΔC_1. We choose the net gain and forgo the net resource costs. We can represent this in incremental cost effectiveness space as shown in **Figure 1**. The four quadrants in this figure are referred to by their compass orientation.

Figure 1 Incremental cost effectiveness for two illustrative treatments



Now consider the decision to implement T_2. Again this results in a gain in health (E_2) and the loss of the health E_s. However, this time E_s exceeds E_2. We now have an incremental ('bundled') loss of health equal to dis-benefit ΔD_2. Again, the decision to implement T_2 results in the resource cost (+C_2) against which we can set the saving in resources +C_s if C_s is no longer implemented. For this treatment in the South West quadrant, the C_s exceeds C_2 resulting in a saving ΔS_2. Now it is the saving that constitutes the gain from the decision. While the opportunity cost is the health dis-benefit.

Section 3 Applying the concept of opportunity cost

There are both ratio and arithmetic approaches to examining the resource consequences and the health effects of a treatment. The commonly used ratio approach is the incremental cost effectiveness ratio in which the numeric value of the incremental resource consequences is divided by the numeric value of the incremental health effects. This paper presents an alternative ratio approach in which the incremental value of the gains is divided by the value of their incremental opportunity cost. This is shown to be closely analogous to commonly used arithmetic approaches since they both utilise the concept of opportunity cost in its classical economic interpretation, as the value of the forgone alternative. The paper develops this comparison between the commonly accepted arithmetic approach and the newly described ratio approach by providing examples of each.

Arithmetic approaches[8 10] have been applied in various ways, sometimes as net benefit, sometimes as net cost. Sometimes values are expressed in QALYs sometimes they are expressed in currency. The approach here is to express net benefits in QALYs. For positive values of resource consequences the health equivalent is $+\Delta C / \omega$, ie a health dis-benefit, while for negative values of the resource consequence (ie savings), the equivalent health effect is positive and is calculated as $-\Delta C / \omega$ (ie $\Delta S / \omega$).

Section 3.1 Net benefit calculations

For T_1 implemented in preference to T_s to achieve health benefits, we must subtract the losses (the opportunity cost) from the gains (in this case the health benefits). We calculate the health equivalent of the resource consequence as $\Delta C_1 / \omega$, which we denote by $\Delta D_1'$. Thus:

Equation 1 Net benefit = $\Delta E_1 - \Delta C_1 / \omega = \Delta E_1 - \Delta D_1' = \Delta G_1 - \Delta O_1$

Ie the net benefit is equal to the value of the additional health, less the value of the additional resources used in obtaining it. Those resources representing an opportunity cost of lost health that could have been achieved had they been used elsewhere.

For T_2 implemented in preference to T_s to achieve savings, the health equivalent of those savings is equal to $\Delta S_2 / \omega$, denoted as $\Delta E_2'$, against which we must set the opportunity cost, which is the health dis-benefit ΔD_2:

Equation 2 Net benefit = $\Delta S_2 / \omega - \Delta D_2 = \Delta E_2' - \Delta D_2 = \Delta G_2 - \Delta O_2$

Ie the net benefit is equal to the health equivalent value of the savings less the value of the health effects that are forgone.

Figure 1 represents this situation. Resource consequences are projected onto the health effect axis to show their health equivalents. (The diagonal lines represent social indifference curves the slopes of which indicate the price that society is prepared to pay for a health gain, ie the slopes are equal to ω.) Transformed in this way, we see the equivalence between resource savings and health benefits. They are both gains and appear as either a health benefit or as the equivalent of a health benefit.

Section 3.2 The 'payback ratio', a ratio analogue of net benefit

Incremental gain to opportunity cost ratios use the same information as net benefit calculations, but express that information in ratio rather than difference format. For brevity the term 'payback ratio' is used to describe the incremental gain to opportunity cost ratio. Replacing the minus operator used in net benefit calculations with the division operator, we have

Equation 3 Payback ratio of $T_1 = \Delta E_1 / (\Delta C_1 / \omega) = \Delta E_1 / \Delta D_1' = \Delta G_1 / \Delta O_1$

Equation 4 Payback ratio of $T_2 = (\Delta S_2 / \omega) / \Delta D_2 = \Delta E_2' / \Delta D_2 = \Delta G_2 / \Delta O_2$

Payback ratios are not defined in the North West or South East quadrants where there are respectively no incremental gains and no opportunity costs.

Where payback ratios are defined, decreasing resource use is monotonically related to both an increasing net benefit and an increasing payback ratio (remembering that decreased resource use in the South West quadrant increases the quantity of savings). Similarly increasing effectiveness is monotonically related to both an increasing net benefit and an increasing payback ratio.

Note that identical results would have been obtained had we chosen to represent all values in their financial rather than their QALY equivalents. The ratio of values is independent of the choice of numeraire.

Section 3.3 Incremental resource to health outcome ratios

By ignoring the concept of opportunity cost, and choosing not to convert health effects into their value equivalents, we can construct incremental resource to health outcome ratios as follows:

Equation 5 Resource to health outcome ratio of $T_1 = \Delta C_1 / \Delta E_1$

Equation 6 Resource to health outcome ratio of $T_2 = \Delta S_2 / \Delta D_2$ or $-\Delta C_2 / -\Delta E_2$

It is hard to know what joint interpretation to put on these ratios, since their numerical values are not meaningfully comparable. High values of T_1 in the North East quadrant represent high positive costs for small health outcomes, while high values of T_2 in the South West quadrant represent large savings for small health losses.

Incremental resource to health outcome ratios have acquired the soubriquet 'incremental cost effectiveness ratios' through long use. In fact the term 'cost' is a misnomer in the South West quadrant where it does not have an economic cost interpretation (ie it is not an opportunity cost).

Section 3.4 Deterministic use of net benefits

Consider comparison of the treatments T_1 and T_2 above. Suppose T_1 produces 1 QALY per person at a cost of £10,000 and that T_2 saves £100,000 per person with a loss of 1 QALY per person. Which treatment is preferable? We need to know what else we might have done with the £10,000 cost of T_1 and what can we do with the £100,000 savings of T_2. To evaluate these options we need to know the value of the measure of health benefit ie, the value of ω.

For T_1 and assuming the value society that society attributes to a QALY (ω) is £50,000, we have from Equation 1:

Net benefit = $\Delta E_1 - \Delta C_1 / \omega$ = 1 - £10,000 / £50,000 = 0.8 of a QALY per person treated

For T_2 we have from Equation 2:

Net benefit = $\Delta S_2 / \omega - \Delta D_2$ = £100,000 / £50,000 – 1 = 1.0 QALY per person treated

Treatment 2 is just preferable to treatment 1.

Now assume that the value society attributes to a QALY (ω) is only £20,000. In this case we have from Equation 1:

Net benefit = $\Delta E_1 - \Delta C_1 / \omega$ = 1 – (£10,000 / £20,000) = 0.5 of a QALY per person

And from Equation 2:

Net benefit = $\Delta S_2 / \omega - \Delta D_2$ = £100,000 / £20,000 – 1 = 4.0 QALY per person

As a result of altering the value that society attributes to a QALY, we alter the relative values of treatment 1 and treatment 2. With a lower value per QALY, the health gains provided by treatment 1 decrease, while the savings from treatment 2 can buy more QALYs. There is nothing new or controversial in these results. They are provided for comparison with the payback ratio calculations using the same figures.

Note however that it is insufficient to quote only the net benefit per person. The decision maker would very probably also wish to know whether the net benefit is a measure of the resources released from a treatment, or of the gains from a treatment that requires additional funding. Note also that the measure of net benefit is expressed as the number of QALYs per person.

Section 3.5 Deterministic use of payback ratios

Consider again treatment T_1 in the example above where £10,000 is spent per person to achieve 1 QALY gain. If we assume that the societal value of a QALY is £50,000 per QALY, then the £10,000 spent per person is equivalent to 1/5 of a QALY spent per person. This outlay of 1/5th of a QALY per person results in an increase in health of one QALY per person, or a ratio of 5 QALYs gained per QALY forgone.

Considering treatment T_2 (£100,000 saved for the loss of 1 QALY), and again assuming that the value of a QALY to be £50,000, then the saving of £100,000 would be worth 2 QALYs to society. Thus implementing T_2 enables us to obtain the value of 2 QALYs in exchange for 1 QALY forgone.

Using Equation 3 we have:

Payback ratio of $T_1 = \Delta E_1 / (\Delta C_1 / \omega) = 1 / (£10{,}000 / £50{,}000) = 5{:}1$

And from Equation 4:

Payback ratio of $T_2 = (\Delta S_2 / \omega) / \Delta D_2 = (£100{,}000 / £50{,}000) / 1 = 2{:}1$

By using the societal willingness to pay, and expressing each option in terms of its payback ratio, we are able to compare the projects and observe that the QALY gains from T_1 are achieved at a lower QALY cost (ie a higher return). To emphasise the health impact of these choices, I have expressed all options in terms of their equivalent health gains and losses. As mentioned previously, the results are exactly the same if we convert all gains and losses into their financial equivalents.

If the value society places upon a QALY is only £20,000, ie ω is lower, then the value that society places on the gains from T_1 is less, while the loss of health from T_2 is of less concern to society. In this example we see that the preference between the two projects is reversed. Again from Equation 3:

Payback ratio of $T_1 = \Delta E_1 / (\Delta C_1 / \omega) = 1 / (£10{,}000 / £20{,}000) = 2{:}1$

And from Equation 4:

Payback ratio of $T_2 = (\Delta S_2 / \omega) / \Delta D_2 = (£100{,}000 / £20{,}000) / 1 = 5{:}1$

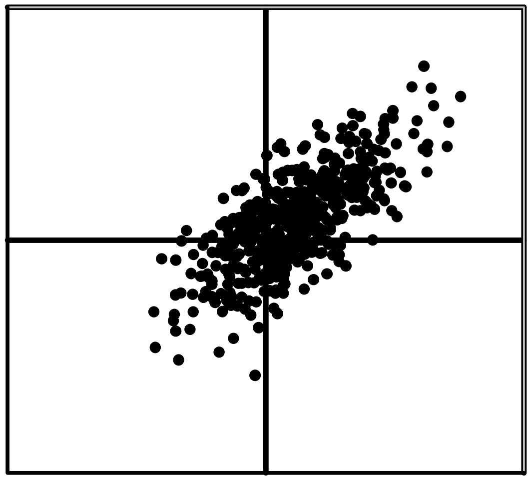
Payback ratios (like net benefits but unlike conventional cost effectiveness ratios) allow us to make a direct comparison between treatments that release savings and treatments that improve health at a cost. Again, as in the case of net benefits, the decision makers will need to know which of these is likely to occur, so that they can manage their budget accordingly.

Section 3.5.1 QALY league tables

For some years QALY league tables have been used to compare treatments[11]. There are acknowledged difficulties in using them, for example: when different health valuation or costing methodologies have been applied; when equity problems must be addressed; or when there are difficulties deciding on an appropriate comparator treatment. However comparative judgements are nevertheless made between treatments on the basis cost per QALY information. QALY league tables have hitherto been used for interventions in which health improvements are acquired at a resource cost (ie in the North East quadrant). As a result, the difficulty in making comparisons between quadrants has not arisen. In these circumstances the cost per QALY provides an adequate decision rule. However, if we wished to prioritise interventions across quadrants a cost per QALY analysis would be inadequate. Fortunately it would be computationally straightforward to convert the existing reported cost per QALYs into a payback ratios (invert the ratios and multiply by ω) and to extend those league tables by reporting on payback ratios for cost saving interventions. The use of a payback ratio would then enable direct comparisons to be made between health improving and cost saving interventions.

Section 3.6 Stochastic use of net benefit

Figure 2 Scatterplot of incremental cost effectiveness estimates



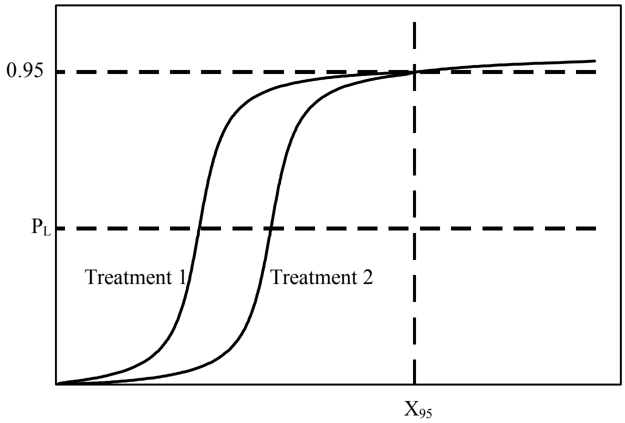
Stochastic analyses arise in the context of simulation modelling or data analysis, where 'bootstrapping' provides a useful approach. This latter is considered here. Imagine that we have information about the costs and the effects of both the standard treatment and the new treatment for a set of subjects in a study. A single bootstrap estimate of incremental cost effectiveness could be determined by drawing samples of matched pairs of resource drawing matched samples of resource consequences and health effects from the data. For bootstrapping, 'bootstrap' samples are drawn that are equal in size to the original sample data. They are drawn with replacement. The process is repeated to produce a large number of cost effectiveness estimates. These estimates can arise in any of the four quadrants of the incremental cost effectiveness plane as shown in Figure 2.

Stinnet et al[4] have shown how to produce a cumulative distribution function in net benefits. That part of their analysis is summarised here for comparison with payback ratios. They argue that cumulative distribution functions are useful because they enable distributions to be examined for stochastic dominance and that

> 'Stochastic dominance is a powerful analytic tool because it allows one to identify cases in which a decision maker should unambiguously prefer one alternative over another despite the presence of uncertainty, with only very general assumptions required regarding the decision maker's utility function.'

Figure 3 shows cumulative distribution functions relating to two treatments. The horizontal axis represents net benefit scaled such that worse options (lower net benefit) are to the right. The cumulative distribution function shows the probability (on the vertical axis) of obtaining net benefits at least as good as the values represented on the horizontal axis. The further the right we are on the horizontal axis, the lower is the threshold value of net benefits that we apply, hence the more easy it is to satisfy that threshold, and hence the higher the probability that it might be achieved. The hatched line at X_{95} illustrates a 95% probability of achieving the criterion. At the 95% confidence level each treatment is equally acceptable. At lower probability levels (for example at P_L) Treatment 1 is preferred to (to the left of) Treatment 2. At all probability levels Treatment 1 is either better than, or as good as treatment 2. Treatment 1 is therefore said to be stochastically dominant.

Figure 3 Cumulative distribution function for net benefits



Stinnett et al go on to state,

'In contrast the concept of stochastic dominance cannot be applied in analyses based on incremental CE ratios, because preferences are not monotonic in R (the cost effectiveness ratio) and monotonicity is a necessary condition for stochastic dominance.'

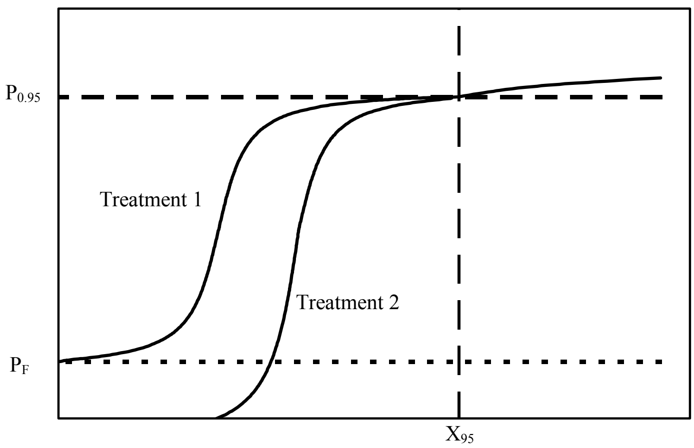
Thus Cost Effectiveness Acceptability Curves[12] are not cumulative distribution functions and cannot be used to examine stochastic dominance.

Essentially the difficulty is that higher (or lower) cost effectiveness ratios are dis-preferred where they represent health gains bought at a resource cost, but are preferred where they represent savings achieved at the expense of a reduction in health. Where replications materialise in both the North East and the South West quadrants, a higher cumulative distribution function for one treatment could result in either high probabilities of large savings or high probabilities of large costs. The former would make the treatment preferred, the latter dis-preferred.

Section 3.7 Stochastic use of payback ratios

This difficulty is in part resolved when we apply a more rigorous definition of incremental opportunity costs as suggested above. When payback ratios are defined, preferences for those payback ratios are monotonic in acceptability. Higher payback ratios are preferred to lower payback ratios whether the treatments they refer to materialise in the North East or the South West quadrant.

Figure 4 Cumulative distribution functions in payback ratios



This is illustrated in Figure 4. Again both treatments share the probability of achieving payback ratios beyond a certain point, while Treatment 1 is preferred at all lower probabilities.

Treatment 1 is shown as having a 'floor to its payback ratio. P_F shows the probability that the treatment will result in both health benefits and resource savings. Note that ceiling payback ratios are also possible, since payback ratios are not defined if there are no gains (in the North West quadrant).

Section 4 Discussion

Section 4.1 Reclaiming the definition of opportunity cost

It is commonplace for texts to identify opportunity costs with resources used[2 13 14]. For example, Gold et al note that,

> "A common and tractable method useful in calculating the societal **opportunity cost** of a health treatment in a cost-effectiveness analysis thus locates and assigns a price to each of the resources **consumed or saved** by the intervention." (emphasis by KB).

Under this definition, resources saved can constitute an opportunity cost. This focus on resources is entirely understandable. Health economists do not usually evaluate treatments in isolation, but compare one treatment with another. They usually employ an incremental analysis. The first stage of such an analysis is to collect separate evidence on the effects and resource use for both the existing and the alternative technology. Both of these individually will (usually) require the expenditure of resources, and those resources will be their opportunity cost. It is when we take that information and determine the incremental costs and incremental benefits that the incremental costs of implementing the treatment can become negative and hence represent a gain rather than an opportunity cost.

Moreover our usual concern has been with the implementation of treatments that provide improvements in health care, rather than those that are designed to reduce the use of resources. Ie, we are more usually concerned with the North East than the South West quadrant.

Since health is the focus of cost effectiveness analysis, it is useful to draw a sharp distinction between health effects and resource effects. However, in identifying opportunity costs with resource effects, we lose the normally accepted definition of opportunity cost as the value of the alternative forgone. Rather than contrast health effects with opportunity costs, it might be more accurate to contrast health effects with resource effects; and to contrast gains (of whichever kind) with opportunity cost (of whichever kind).

Section 4.2 The various interpretations of 'shadow price'

This section considers alternative views about the appropriate interpretation of the concept of shadow price. A full examination is beyond the scope of this paper, however the issue needs to be addressed, if only briefly.

Section 4.2.1 Prices as ratios

Prices are ratios. They are the ratio of what you pay to what you get. A trader needs to know two distinct sets of prices; the prices of what they sell (the selling price) and the prices of what they buy (the buying price). Prices can themselves be expressed as ratios of one another, as when traders examine the ratio of selling price to buying price (they usually subtract 1 from this ratio and describe it as the 'percentage mark-up'). This informs them about the profitability of transactions. Traders need prices.

The distinction between these two sets of prices is central to the arguments that follow and informs the distinction between societal willingness to pay and shadow price.

Section 4.2.2 Payback ratios are ratios of prices ie they are ratios of ratios

I have suggested that an appropriate definition of opportunity cost leads to a re-specification of cost effectiveness ratios as payback ratios. These are ratios of health benefits to costs or ratios of savings to dis-benefits and are standardised using the value that society assigns to a unit of the health benefit. They can be re-stated as a ratio of prices. This ratio of 'price' ratios can be regarded in a similar way to the trader's mark-up.

In North East quadrant the payback ratio is:

Equation 7 $E / (C \div \omega) = \omega / (C \div E)$

C / E is the unit cost of producing the health gain. If we denote this by κ we have:

Equation 8 Payback ratio = ω / κ

ie, the payback ratio is the societal value of a QALY (compare with a selling price) divided by the cost of obtaining the QALY (compare with a buying price). We have a high (advantageous) payback ratio if the cost of obtaining a QALY is small.

Now consider the South West quadrant. The payback ratio is:

Equation 9 $(S \div \omega) / D = (S \div D) / \omega$

If we denote (S÷D), the ratio of the financial savings to the reduction in health, by σ:

Equation 10 Payback ratio = σ / ω

Ie the money saved per QALY forgone (compare with a selling price) divided by the value of a QALY forgone (compare with a buying price). We have a high (advantageous) payback ratio if the saving per QALY forgone is large.

The aim here has been to emphasise that a payback ratio might be regarded as a ratio of price ratios, otherwise there is little said in this section that has not been said previously.

Section 4.2.3 Decision-making under constrained optimization

Now consider decision making under constrained optimization. In an industrial context, where a firm cannot expand some of its productive capacity up to the point where the marginal revenue obtained by increasing that factor of production is equal to the marginal cost of that factor, the factor acquires a shadow price in excess of the cost of acquiring it. This shadow price is denoted by λ, and the scarce resource is said to place a 'binding constraint' on output. Suppose a factory is facing delays getting new equipment. It might like to add more lathes, but has to wait for them. In constrained optimization problems, the shadow price is the amount by which the objective function would be increased when a constraint is relaxed by one unit. In this case the objective function would be to maximise profits, and the amount that this could be increased, by having an additional lathe, would be the shadow price of the lathe. The shadow price of other pieces of equipment could be determined in the same way.

Is it reasonable to apply a similar analysis to the health care system? Consider a funding shortage. We might wish to know by how much the output could be increased by having more resources available. If the value of the output were measured in QALYs, then by analogy to the factory situation, the shadow price would be the number of additional QALYs that could be produced if the resource constraint were relaxed by (say) an incremental £100,000. Suppose we knew that we could produce 10 QALYs from the marginal treatment with an additional £100,000. Under these assumptions, the shadow price of the £100,000 would be 10 QALYs.

It would still be meaningful to ask what value the public would place upon those QALYs, ie to know the societal willingness to pay for a QALY. Suppose the societal willingness to pay for a QALY to be £50,000, then the value obtained by relaxing the constraint by £100,000 would be 10 x £50,000 or £500,000. If we denote the shadow price denominated in so many QALYs per £ of additional resources as λ_Q, we can determine the shadow price denoted in so many £s per £ of additional resources as $\lambda_£$, such that:

$\lambda_£ = \omega \times \lambda_Q$.

The shadow price of cash resources in an overly cash constrained health care system might thus be £500,000 per £100,000. Each additional £1 added to the health care budget would produce health benefits that would be valued at £5.

Note that the shadow price of the marginal project (the number of QALYs obtained per additional £ spent) is the reciprocal of the cost per QALY (the number of £ spent to produce an additional QALY) from the marginal project.

The shadow price $\lambda_£$ is the payback ratio for the marginal investment. Thus if we denote the cost per unit of health gain of the marginal project as κ_m, we have:

$\lambda_£ = \omega \times \lambda_Q = \omega / \kappa_m$.

If we are resource constrained $\lambda_£ > 1$ and payback ratio >1

Note that λ, ω and κ, although all ratios in QALYs and resource costs, nevertheless describe different, distinct concepts.

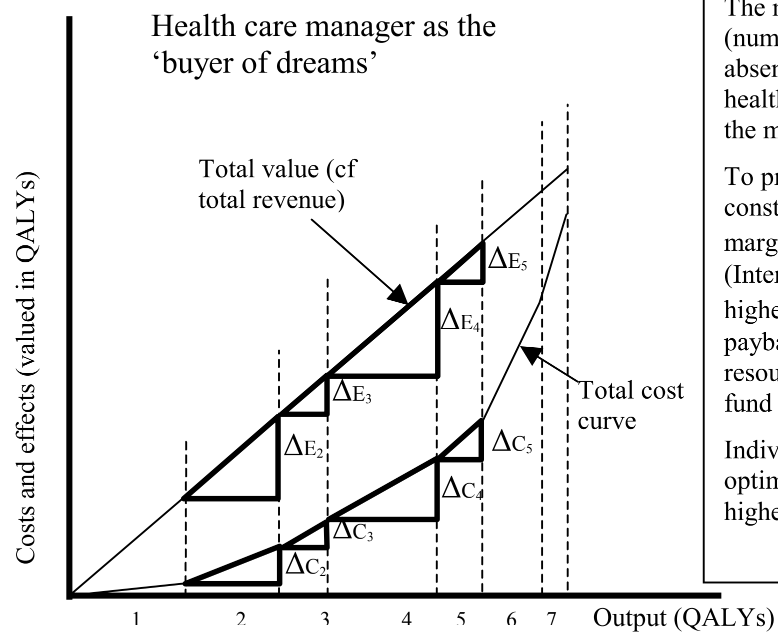
It remains to ask whether this calculation is feasible and also whether it is appropriate in the context of health care. Unfortunately it is not feasible, primarily because it requires information about all the treatments that are available. Without this information, we do not know which treatment is the marginal treatment, nor do we know whether the resource constraint is binding. There may be many treatments that would not be undertaken if their health effects and cost consequences had been evaluated.

The question of whether the approach is appropriate depends upon whether the public would wish to see a binding resource constraint, or whether they would prefer to set the societal value of a QALY and fund appropriately. It might be considered irrational for the public policy makers to knowingly set a resource constraint that results in a shadow price that differs from the value assigned to a QALY, ie where the objective function and the constraint are in conflict with one another. The constraint denies them gains that they have acknowledged in the value they assign to health gains. At best we might consider this to be a short term expediency.

Section 4.2.4 The health care manager as the buyer and seller of dreams

When the health care manager pays for a treatment, he or she is buying the dreams of improved health for the people who will benefit from it. When the manager is thinking of discontinuing a treatment, he or she is selling or liquidating the dreams of those who would benefit. If he or she were a trader they would be looking for the highest rate of return on the deals that they make. Health care managers may be doing something similar. High payback ratios on the patients' dreams that they buy are achieved when the health gain is high in relation to the cost of buying that care. High payback ratios on the patients' dreams that they sell are achieved when the savings they get are high in relation to the loss of health.

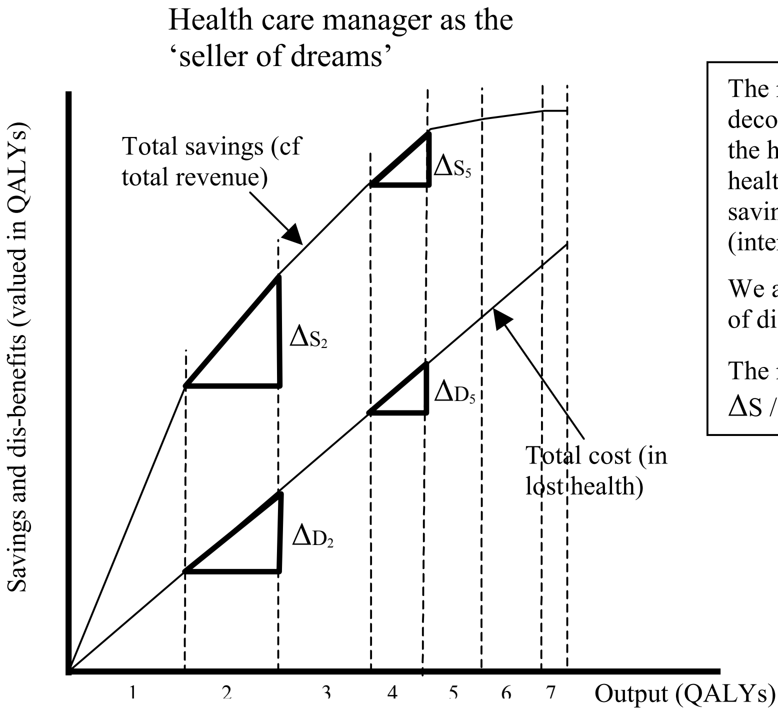
Figure 5 Marginal analysis of health care interventions.



The manager has options on 7 interventions (numbered along the horizontal axis). In the absence of resource constraints, the manager buys health care interventions until the marginal cost = the marginal value (intervention 5).

To prioritise interventions under a resource constraint, they are ranked in order of the ratio of marginal value to marginal cost ie $\Delta E/\Delta C$. (Intervention 3 is preferred to 4 although 4 has a higher net benefit ($\Delta E - \Delta C$).) $\Delta E/\Delta C$ is the payback ratio. This exceeds a value of 1 where resources are constrained. (For example, we can fund only the first two interventions.)

Indivisibilities and mutual exclusivity can affect optimisation decisions. In these circumstances the highest payback ratio is not necessarily preferred



The manager has options to continue or decommission 7 interventions as numbered along the horizontal axis. The manager decommissions health care interventions until the marginal savings = the marginal dis-benefit (intervention 4).

We assume that there is no constraint on quantity of dis-benefits.

The net benefits = $\Delta S - \Delta D$, the payback ratio = $\Delta S / \Delta D$.

These points are illustrated in Figure 5, which demonstrates that the representation of dis-benefits as the cost of an intervention is consistent with a conventional profit maximisation representation analysis. The actions of commissioning and decommissioning health care interventions are displayed separately and correspond to the North East and South West quadrants (respectively) in cost effectiveness space.

The upper half of the figure differs from a normal cost effectiveness representation only in expressing the value of both costs and effects in a common unit (in this case in QALYs). In the absence of a resource constraint, this provides a decision rule whereby interventions are selected up to the point at which the marginal cost equals the marginal benefit. The lower half of the figure demonstrates the reversal of roles between resource consequences and health effects. The health effects now comprise the cost and savings take the place of marginal benefits. Again the appropriate decision rule is to implement interventions up to the point where the marginal cost equals the marginal benefit.

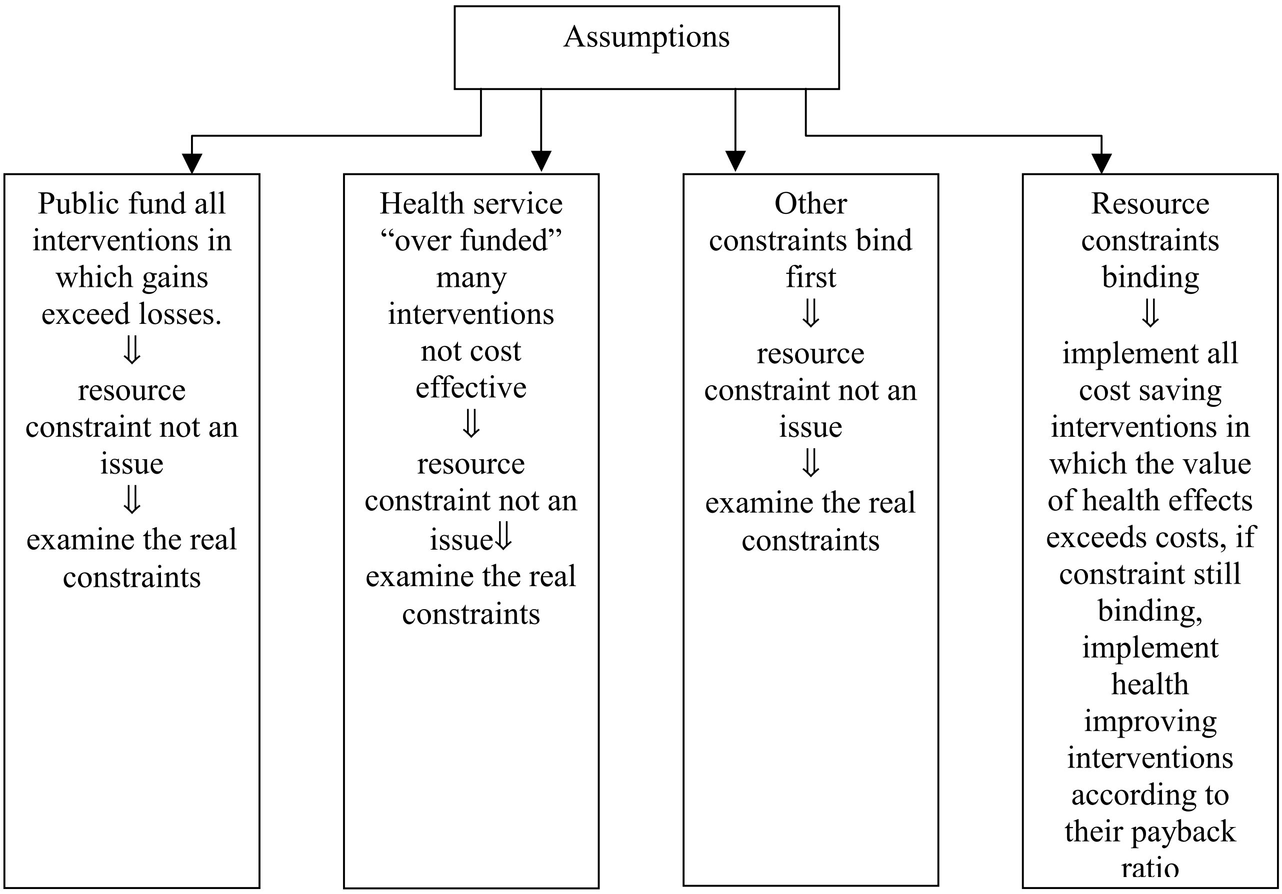
Section 4.3 Introducing other constraints

We have seen that any intervention that provides a positive net benefit, or equivalently a payback ratio greater than unity should be implemented. To understand why they are not, we must consider the nature of the constraints that prevent their implementation. If the constraint were on the available resources, then total benefits would be maximised by obtaining the greatest return on those resources. Resource constraints only present a problem for decision makers in the North East quadrant. (No projects should be implemented in the North West quadrant, while interventions in the South West and South East quadrants release resources and hence are not constrained by limits on them.) Resource constraints may not bind under any of three conditions: either if the public will fund any programme that satisfies the minimum criteria above (net benefit greater than zero or equivalently payback ratio greater than 1); or if the health service is already over resourced (ie it is using some of its resources wastefully); or if another constraint is binding.

Such another constraint might be the resource used in decision making itself. To consider this question, we need to distinguish between health care costs and 'deliberation costs'. Health care costs include the set up costs, running costs and the capital equipment that are required once the decision has been made to implement a intervention. Deliberation costs are the 'managerial' resources required in deciding whether to implement an intervention. They include the staff involved in committee meetings, board meetings, preliminary discussions with employee representatives, clinical effectiveness groups, pharmaceutical advisors, public consultations, press conferences, answering public enquiries, discussions with Members of Parliament, meetings of and with special interest groups etc. These are the 'stuff' of management in health care. They all require the use of 'management' time (where management is broadly defined to include, paid managers as well as elected representatives or governmental appointees etc). At the outset, these costs are unknown, but managers will make an assessment of their probable significance at an early stage in the consideration of a health care intervention.

The possible importance of this constraint might be illustrated by a hypothetical example. Suppose a Health Authority could improve patient care and save money by replacing three cottage hospitals with a better equipped modern hospital. The cottage hospitals might have been built in the nineteenth century to provide easy access to care before the coming of the car. Detailed discussions would take place with the staff working in the hospitals, with the management staff responsible for them, with the clinicians who provide care in the hospitals, with the local communities, with the relevant Members of Parliament, with representatives of the Department of Health, possibly with the Minister of Health. Planning staff would attempt to assess the impact on: patient and visitor travelling; staffing costs; administrative overheads. Health Authority members would attend special meetings and so on. In this example, it is clear that health care costs would not result in a binding constraint in providing improvements in care, if costs could be saved rather than incurred, nevertheless, there may be a constraint on the amount of managerial resources available to implement the intervention.

Figure 6 Incremental opportunity costs under various assumptions about constraints



Obtaining sufficient managerial capacity may not simply be a problem of finding the financial resources to pay it. The 'managerial' constraint may be binding for a variety of reasons. For example, there might be: administrative restrictions on the number of 'grey suits'; shortages of management with relevant expertise or credibility; or limits on the availability of the time of Health Authority members. A shortage of 'managerial' capacity to implement change, may mean that the opportunity cost of using that time is very high, and that managerial capacity acquires an inflated shadow price. If it is the case that managerial capacity is the real binding constraint within health care decision making, that requires further economic analysis. It is beyond the scope of this paper to fully address the issue here. The conditions under which incremental opportunity costs are influenced by resource and managerial constraints are summarised in Figure 6.

Section 5 Conclusion

I have argued that the phrase 'opportunity cost' should be applied to a health gain forgone. Given the well-established use of the term to describe resource utilisation, this is likely to be controversial. A less challenging alternative might be to refer to 'losses' and 'gains'. This would simply be 'a rose by another name'. This ratio provides a way of comparing gains to losses in a way that is commensurable for both health gains bought at a cost in resources and resource savings bought at a cost in health.

Given all the problems and uncertainties that will continue to exist in the analysis of cost effectiveness data, is it preferable to operate within a logically consistent definition of opportunity costs and gains, or to operate with one that derives from an accounting definition of costs?

References

1. Anderson JP, Bush JW, Chen M, Dolenc D. Policy space areas and properties of benefit-cost/utility analysis. *Jama.* 1986;255(6):794-5.
2. McIntosh E, Donaldson C, Ryan M. Recent advances in the methods of cost-benefit analysis in healthcare. Matching the art to the science. *Pharmacoeconomics.* 1999;15(4):357-67.
3. Mugford M, Hutton G, Fox-Rushby J. Methods for economic evaluation alongside a multicentre trial in developing countries: a case study from the WHO Antenatal Care Randomised Controlled Trial. *Paediatric and Perinatal Epidemiology.* 1998;12(Suppl 2):75-97.
4. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis.[comment]. *Medical Decision Making.* 1998;18(2 Suppl):S68-80.
5. O'Brien BJ, Briggs A. Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods. *Statistical Methods in Medical Research* 2002;11:455-468.
6. Heitjan DF, Moskowitz AJ, Whang W. Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Economics.* 1999;8(3):191-201.
7. Laska EM, Meisner M, Siegel C. Statistical inference for cost-effectiveness ratios.[comment]. *Health Economics.* 1997;6(3):229-42.
8. Zethraeus N, Johannesson M, Jonsson B, Lothgren M, Tambour M. Advantages of using the net-benefit approach for analysing uncertainty in economic evaluation studies. *Pharmacoeconomics.* 2003;21(1):39-48.
9. Pearce DW. *Macmillan dictionary of modern economics*. 4 ed. Basingstoke and London: Macmillan Press Ltd, 1992.
10. O'Hagan A, Stevens JW. Net health benefits: a new framework for cost-effectiveness analysis from clinical trial data. *Health Economics.* 2001;10(4):303-15.

11. Wasem J, Hessel F, Kerim-Sade C. Methods of comparative economic evaluations of therapies and for rational allocation of resources across sectors of health care systems - introduction, advantages, risks. *Psychiatrische Praxis.* 2001;28(Suppl 1):S12-20.
12. van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. *Health Economics.* 1994;3(5):309-19.
13. Donaldson C, Gerard K. *The economics of health care financing*. London: Macmillan Press Ltd, 1993.
14. Gold MR, Siegel JE, Russell LB, Weinstein CM. *Cost-effectiveness in health and medicine*. New York, Oxford: Oxford University Press, 1996.