On the Margins of Health Economics:
Searchers, Surveyors and the Monetary Value of a Qaly

March 13, 2014

Stephen Birch, D.Phil. [a,b]
Amiram Gafni, Ph.D. [b]

[a] Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, Canada
[b] Manchester Centre for Health Economics, University of Manchester, UK.

CHEPA WORKING PAPER SERIES

1. ABATRACT

Threshold searching and threshold surveying have been presented as separate non-arbitrary approaches for identifying the ICER threshold for determining the efficiency of a new health care intervention. More recently reconciliation between the approaches has been suggested by applying the societal willingness to pay for a QALY with routinely collected data on expenditures on health care interventions and health outcomes associated with those expenditures.

In this paper we show that both approaches confuse the concepts of the margin and the increment (as applied to costs and effects) and that this confusion represents a fundamental departure from economic theory. We illustrate how a practical method for determining the efficiency of investments in health care can be developed based directly on the theoretical principles of economics while also reflecting the real world setting facing decision-makers.

3

2. INTRODUCTION

The search continues in health economics for a simple solution to the complex problem of determining whether adopting a new health care intervention would increase the total health gains produced from the resources of a health care system and hence increase efficiency of that system. Two alternative *theoretical* solutions to the problem were presented by Weinstein and Zeckhauser (1973) based on the incremental cost effectiveness ratios (ICER) of health care programmes. The first solution (the **ICER league table** approach) requires that interventions be adopted in ascending order of ICER from among all possible interventions until available resources are exhausted. The alternative solution (the **ICER threshold** approach) requires that interventions with an ICER less than or equal to the shadow price of the budget (i.e., available resources) be adopted. However the authors were careful to emphasize the theoretical nature of these solutions and made no attempt to appraise the underlying assumptions required to support these solutions or their relevance for practical applications.   In particular the solutions require perfect divisibility and constant returns to scale in the production of all interventions, the willingness and ability of policy makers to provide an intervention to only a proportion of the patient group needing the service in addition to information on the effects and costs of all possible interventions.

Methods for dealing with the real world of decision-makers, where interventions are in many cases indivisible (either technically or ethically) and production involves non-constant returns to scale, have been presented in the literature (Birch and Donaldson 1987, Birch and Gafni 1992) and used in empirical applications (e.g., Tianawit et al. 2010, Cleary et al. 2010). More recently Weinstein (2012 p 513) acknowledged that in situations where there are technical or ethical limitations to 'partial implementation' of interventions (e.g., adopting 0.65 of a Magnetic Resonance Imaging machine or vaccinating a quarter of all children at risk of a condition) then "a generalised optimization framework would be necessary".

Despite this recognition of the practical limitations of the ICER approach to decision making, the economic evaluation literature has focussed on the ICER threshold approach: 'identifying' a threshold QALY value (or 'cost per QALY') to be used for accepting or rejecting possible new interventions by comparison with each intervention's ICER so relying implicitly on the Weinstein-Zeckhauser assumptions. Interventions with ICERs higher than the threshold are rejected while those with ICERs below the threshold are adopted. The rationale for this approached, as presented by Culyer et a. (2007) is that "the incorporation of any technology not in current use with an mhg (marginal health gain) above (the threshold) would represent an increase in (total) health gain (from available resources) as long as it displaces a technology with a lower mhg" (Culyer et al. 2007 p58).

Where full information on the incremental costs and effects of all possible programmes is available (i.e., an implicit assumption of Weinstein and Zeckhauser), the determination of the threshold ICER is straightforward and is given by the ICER of the last programme funded under the current budget. Without perfect information we cannot determine the threshold ICER because the ICERs that can be calculated for programmes for which perfect information is available cannot be compared with ICERs for programmes without perfect information (i.e., programmes with unknown ICERs). Alternative approaches must therefore be found to determine the threshold ICER. In some cases researchers simply adopt arbitrary values for the threshold ICER. For example when questioned about the adoption of a threshold of $20,000/QALY (Laupacis et al 1992) the lead author explained " we made it up" (Lorinc 2006). In other cases no attempt is made to justify the choice of the threshold value (Gafni and Birch 2006).

Two broad approaches have been suggested for determining the ICER threshold in non-arbitrary ways that correspond to two established methods of determining valuation in cost-benefit analysis; implied values from observed behaviour (the 'threshold searchers' approach presented by Culyer et al. (2007)) and contingent valuation through surveys (the 'threshold surveyors' approach presented by Baker et al. (2011)). Both approaches will be explained briefly in the next section. Baker et al (2011) also attempt to reconcile the two approaches by combining the values the public attach to different ways health gains can be generated (from the 'surveyors' approach) with routinely collected data on expenditures on health care interventions and health outcomes associated with those expenditures (from the 'searchers' approach).

In this paper we critically appraise the two approaches and the proposed reconciliation between the approaches as a basis for determining the efficient allocation of health care resources. We show that both approaches confuse the concepts of the margin and the increment (as applied to costs and effects) and that this confusion represents a fundamental departure from economic theory. As a result, only under very special assumptions, that generally do not hold in health care applications, would each method and the proposed reconciliation identify health care interventions that improve efficiency (i.e., increase health gains from available resources). An alternative approach is presented which illustrates how incremental costs and effects can be used to identify unambiguous improvements in efficiency without the need for special assumptions and in ways that reflect the real world setting in which decision-makers operate.

3. THRESHOLD SEARCHERS: THE IMPLIED THRESHOLD OR 'SUPPLY SIDE; APPROACH

The implied threshold approach (Culyer et al. 2007) involves plotting the *marginal* health gain from an additional unit of health care expenditure, **dH/dE**, against annual health care expenditure, **E** (in the cited paper the particular application is National Health Service (NHS) expenditure). Figure 1, taken from Culyer et al. (2007), presents the case of the marginal product (or marginal health gain) of current health care expenditure diminishing at a constant rate. Interventions are ordered in descending order of *marginal* health gain from additional expenditure (or what the authors call the 'impact on health per pound spent'), line **ha** on Figure 1.   Once available resources, **E***, have been exhausted (i.e., the total NHS budget has been spent) no more interventions can be funded.   The authors note that the marginal health gain at this point, a, represents the inverse of "the *marginal* cost-effectiveness of NHS expenditure" (Culyer et al. 2007 p57).[a]   However the authors interpret this to be the "threshold *incremental* cost-effectiveness ratio", i.e., the ratio to be used in comparison with ICERs of potential interventions for implementation, an interesting 'sleight of hand' (or at least 'sleight of words').   This implicitly (and some might argue conveniently) assumes that either the marginal health gain is equal to the incremental effect (or incremental health gain) *and* the marginal cost is equal to the incremental cost, or the ratio of the marginal health gain to marginal cost is equal to the ICER even though marginal health gains and costs are not equal to incremental health gains and costs.   Only under these assumptions (which follow from the Weinstein and Zeckhauser assumptions of perfect divisibility and constant return to scale) could we expect to have the downward sloping straight line for marginal health gains.   No attempt is made to identify or justify this assumption or to consider the consequences of the assumption for the analysis.

Figure 2 (also taken from Culyer et al. (2007)) introduces the marginal health gain for interventions not currently funded, **cf**, and the solution to allocating available expenditure **E*** between all possible interventions by aggregating expenditures on all interventions horizontally to give **hde**.   Optimal allocation of resources among all interventions is determined by adopting interventions in descending order of marginal health gains up to available NHS expenditure **E***. This requires **0E'** be allocated to existing interventions and **E'E*** (= **E*E''**) be allocated to new interventions. Funds have been reallocated from previously funded interventions with lower marginal health gains to new interventions with higher marginal health gains. The marginal cost-effectiveness of NHS expenditure has increased to **E*b** (>**E*a**), the area under the marginal health gain curve is greater than in Figure 1 and hence total health gains from NHS expenditure **E*** (and efficiency of NHS resource use) has increased. This presentation is substantively identical to the framework presented by Birch et al. (1996) and Birch and Gafni (2003) (see figure 3) when considering increases or reductions in the size of programmes under budget constraints where **0E*** on the

---

[a] If we swapped the axes, this would give us the expenditure required (or cost) for an additional unit of health gain, or the marginal health cost, dE/dH

horizontal axis is the resource constraint and the marginal health gain for the competing intervention slopes downwards from right to left from **c** to **f'**. In other words Culyer et al (2007) have simply swivelled the marginal health gain for new interventions, **cf**, around the vertical expenditure line at **E***). Under this approach marginal health gains of the current and new interventions for the same unit of heath care expenditure are compared directly and efficient allocation occurs at the point of intersection (equality of marginal health gain between current and new interventions), **h***, which by geometry can easily be seen to be at the point on the marginal health gain lines directly above **E'** and equal to the marginal health gain under the aggregated marginal health gain line at **b**.

However, by definition the *marginal* health gain of an intervention is given by comparing the health gain produced by an intervention with the health gain that would be produced with one less unit of expenditure on the same intervention (i.e., moving along the marginal health gain line) (Birch and Gafni 2003) while the *incremental* health gain is the difference between the health gain produced by the intervention and the health gain produced from whatever would be done for the same patient population in the absence of the intervention. Hence the assumption of equality between the marginal health gain of additional health expenditure and incremental health gain from adoption of a new programme requires that the incremental cost of the new programme equals one indivisible unit of health care expenditure (i.e., the margin), a very special case that doesn't seem to have any practical relevance when comparing among different health care interventions.

The ICER is given by comparing the difference between the health gains of the new intervention and the health gains produced by the current way of treating the same patient group (incremental health gain of the intervention), with the difference between the costs of providing the new intervention and the costs of the current way of treating the same patient group (incremental costs of the intervention), and taking a ratio of the incremental costs over the incremental effects, *no matter how large the size of these increments*. As a result, the accurate term for this ratio is the *average* cost per additional QALY (or ACAQ) where the QALYs is the measure of health gain, with its inverse being the average rate of return on (additional) investment, ARRI. When considering implementation of a new intervention, these increments should be based on the population to whom the intervention would be delivered (the 'policy population') as opposed to the subjects included in a trial (the 'study population'). The incremental costs and effects (and hence for example the opportunity costs) would be substantially different between policy and study populations even if the ICER was to be the same in both populations.

The use of averages (as in the ACAQ) as the basis for determining the marginal condition for efficiency suffers from at least two major problems; It assumes that intra incremental changes in programme size can be made (i.e., perfect divisibility) in order that the efficient reallocations *at the*

margin in the marginal analysis can be implemented; and it assumes that the ACAQ estimated from the incremental analysis is constant (i.e., constant returns to scale) so that the rate of return observed *on average* for the size of the incremental intervention is the same for intra incremental changes in intervention size. Note that these assumptions are not limited to the intervention(s) under evaluation but are also required for all possible interventions to be replaced in order to free up the additional resources required by the interventions under evaluation. Yet the authors already recognise that there is no reason to believe that marginal (and hence average) health gains are constant as "people with the best chances of being helped have already been helped" (p57). Assumptions of equality between marginal and incremental costs also imply constant return to scale and that there are no fixed costs associated with any interventions (Birch and Gafni, 1993). Readers unfamiliar with the problems that emerge where averages are used instead of margins are referred to the study of the 6th stool guaiac test (Neuhauser and Lewicki 1975).

It is worth noting that a key element of marginal analysis in economics is the ability it provides to compare alternative choices directly on the basis of benefits produced (i.e., compare like with like - the effects of spending one dollar more on one intervention with spending one dollar less on another). But the decisions that health care managers face are rarely of this type. Instead they involve investing large amounts of resources in interventions based on a mixture of fixed and variable costs with the consequence that the usual marginal conditions for efficiency are no longer relevant. Instead they involve lumpy investments that require different amounts of additional resources and hence cannot be compared directly or reduced to comparable size with other interventions.   As mentioned above, methods for evaluating lumpy investments have been developed and applied in the health economics literature

The main result of the search for a threshold in the presence of a fixed resource constraint is that new interventions that have a *marginal* health gain greater than that of the last intervention funded under the current constraint would represent an increase in health gain as long as it replaces a currently funded intervention with a lower *marginal* health gain. This is essentially a form of Programme Budget Marginal Analysis (PBMA) (Mooney, 2003, Mitton & Donaldson, 2004). However programmes are not introduced 'marginally' (i.e., should we spend an extra dollar on an MRI machine instead of the last dollar we spent on an endoscope?). Instead decision makers decide whether or not to invest several hundred thousand dollars in an MRI. The threshold searching approach and PBMA both use ICERs as indicators of changes in health gain and costs *at the margin* to compare the proposed new intervention (e.g., the MRI) with an intervention it might replace (e.g., endoscopy) even though the interventions involve different levels of resource investment. This suggests that PBMA might be more appropriately described as programme budgeting incremental analysis (PBIA). More recently Donaldson et al. (2010) mention the problem

of 'lumpy' investments in PBMA. They argue that efficiency would be improved by rational disinvestment under which expenditure is reduced on interventions identified as either not generating any health effects, or generating less health effects than can be produced by allocating the same expenditure to the new intervention. The authors continue to describe this as marginal analysis, possibly because they are considering changes in the size of existing interventions as opposed to replacing one intervention with another. But it is the *'lumpiness'* in the changes in interventions that gives rise to the problems with marginal analysis and the need to base analysis on increments, not the *nature of the programme comparisons*.

To summarize, the authors of the threshold searching approach accurately identify the well-established marginal conditions for the theoretical solution to the constrained maximisation problem. But this is no different to the Weinstein and Zeckhauser theoretical solution, which, although based on choices between programmes, involve *incremental* changes and assumes that these changes are divisible and simply an aggregate of constant *marginal* changes. As a result, the solutions of the threshold searching approach are based on non-marginal measures that depend on the same invalid assumptions of Weinstein and Zeckhauser.

4. THRESHOLD SURVEYORS: THE WILLINGNESS TO PAY OR 'DEMAND SIDE' APPROACH

The authors of the 'willingness to pay' (WTP) approach (the surveyors, Baker et al. 2011) seek to complement the input of the threshold searchers' approach with information on subjects' (individuals within the population) stated willingness to pay for health gains as a way of estimating a societal WTP. They claim to build on the approach of the threshold searchers (as described above) in order to provide a theoretical framework that enables reconciliation of the two approaches. Although the framework they present is fundamentally different from the threshold searchers' framework, it falls into the same conceptual trap of using incremental values to measure marginal changes. As a result, far from enhancing the threshold searchers' framework it compounds the errors of the approach.

The conceptual framework is presented in figure 4 and involves plotting total benefits, **B**, and total costs, **C**, (on the vertical axis) against 'quantity of health care', **Q**, on the horizontal axis. It is not clear how **Q** is to be measured, but it is something different from health care expenditure (or costs) because this is on the vertical axis. Benefits are presented as a **fixed multiple** of health care quantity, **B = VQ** where **V** is the fixed 'value per unit of health care', which they suggest is the value per QALY. But this confuses (or implicitly equates) inputs to health production (the quantity of health care, **Q**) with outcomes (QALYs). In other words, fixing **V** implies *both* a constant marginal

9

productivity of 'health care', **dQALY/dQ**, *and* a constant marginal valuation of QALYs, **dV/dQALY** (or the special case where any non constancies in one are always offset exactly by non constancies in the other)[a].

They recognise that diminishing marginal value of health care may occur but argue that this does not affect the essence of their argument implying the results are robust to relaxing the assumption (or more accurately both assumptions) although this is not shown in the analysis. **But** they then assume that the marginal cost of health care, **dC/dQ**, (which equals the marginal cost of QALY production under their implicit assumptions) increases, even under the modest variations in **Q** that they invoke to justify the constant marginal benefits assumption. The marginal condition for efficiency then occurs where the (constant) marginal benefit (i.e., the gradient of the benefit line) equals the (non-constant) marginal cost (i.e., the gradient to the total cost curve), i.e., **Q*** in Figure 4. If marginal costs were to be constant there would be a problem of either the Benefit and Cost curves being the same (in which case all points on each curve satisfy the efficiency condition), or the Benefit and Cost curves having different slopes at all points (in which case no point on either curve satisfies the efficiency condition). Because the opportunity cost of health care is the benefit forgone of not using health care (however measured) in other uses, the non-constant marginal costs (as implied in the total cost curve) imply non constant marginal benefits of other uses of health care resources. In other words it assumes that health care (and only health care) is characterised by constant marginal benefits and that the marginal cost of non health care expenditure, as measured by the benefits forgone from shifting expenditure from health care, is constant. Moreover, the total cost curve for health care assumes no fixed costs (or, in the absence of an origin on the figure, fixed costs may occur but are matched exactly by some fixed benefit of health care, although it is not clear what that would involve either conceptually or practically. Notwithstanding this somewhat 'designer' framework, it is argued that the optimal level of health care provision is at **Q*** where marginal cost of health care (the gradient of the total cost curve) equals the marginal benefit of health care, given by the marginal value of health care (assumed to be constant). But the marginal value of health care is interpreted to be identical to the value of the QALY (and hence is also assumed to be constant). The idea behind this condition is that to provide one more unit of health care would add more to cost (benefits forgone outside the health care system) than to benefits produced by the health care system and hence would not be an efficient use of resources. In effect, the threshold surveyors are 'determining' the (constant) marginal value of a QALY which is then used to determine the 'optimal' level of health care expenditure.

---

[a] This contradicts one of the authors' previous claims that "as a programme is extended, the marginal benefits are likely to fall" (Donaldson and Farrar 1993)

The problem identified by the surveyors is that the central health care authority "must have adequate information concerning the nature of the cost function . . ." without which "the constrained optimum (i.e., equality between marginal cost and marginal benefit) will simply be achievable only by chance" (p 440). But they make no mention of the availability of adequate information concerning the nature of the benefit function. Instead they simply assume, implicitly that benefits are a constant multiple of the quantity of health care. If this linear relationship between benefit and quantity is relaxed, indicating that the value of (or willingness to pay for) an additional unit of health care (or an additional QALY) varies with the quantity of health care (or quantity of health outcomes – the authors use these two concepts interchangeably), the optimal solution will be based on the traditional **MB=MC**.

Baker et al.(2011) suggest as an example that **Q** might be measured in QALYs. But the QALY is not a measure of the 'quantity of health care' as the authors describe Q to be, but the outcome of health care. Notwithstanding this issue of what Q represents, Baker et al.'s suggestion has several implications for their analysis. First their diagram does not generate an optimal amount of health care for decision makers but an optimal amount of QALYs based on equality between the marginal cost and marginal benefit of QALYs (not the marginal cost and marginal benefit of interventions). Second, there is no reason to believe that the marginal benefit of QALYs is constant and hence free of the usual notion of diminishing marginal valuation. The authors might argue that health is different from other commodities whose value at the margin diminishes as quantity increases (they don't make this argument, but no rationale is provided for why the marginal valuation of QALYs is constant). But the value of a commodity is established through what one is willing to sacrifice for the commodity. If the marginal value of QALYs is constant it means that the marginal opportunity cost (or value of what society is willing to sacrifice for a QALY) is also constant. Third, in the presence of indivisibilities the decision-maker does not face marginal decisions, but instead struggles to solve a budget allocation problem faced with a range of 'lumpy' investments of differing sizes.

The solution presented by Baker et al. (2011) is to provide decision makers with a "best estimate" of **MC** from the 'threshold searchers' (p441) (where **MC** is the expected additional cost of one more unit of health care, not the additional cost of replacing a current intervention with a new intervention) together with information on "the level of **V**" from the threshold surveyors (the "value of a QALY", not the additional value from replacing a current intervention with a new intervention). Interventions with **MC < V** would be adopted and those with **MC > V** would be rejected even though adoption would involve *increments* of costs (what is required for implementation) and health gains (what is produced by implementation), not *margins* (what is 'estimated' by the searchers and surveyors, respectively, albeit using invalid assumptions).

11

5. IF NOT THRESHOLDS, THEN WHAT?

The searchers and surveys base their approaches on the well established marginal conditions for efficiency even though decision makers do not face decisions about marginal changes in interventions (i.e., spending one more, or one less pound on the same intervention as opposed to increasing coverage of an intervention to a wider population) and do not have information about marginal cost effectiveness (the additional cost of producing one more QALY across different interventions). Instead they apply their respective approaches based on measures (or estimates) of incremental changes   replacing one programme by another (or changing the scope of an intervention from one population to another) which they use as a basis for estimating marginal changes by assuming marginal cost is constant (searchers) and marginal valuation is constant (surveyors) for all interventions or population sizes. But the assumption of a constant marginal cost, as adopted by the searchers, is incompatible with the surveyors' analysis while the assumption of a constant marginal valuation is incompatible with the searchers' analysis.   Finally, the searchers' use of the ICER as an analytical tool for determining efficiency implies constant marginal health gains even though they recognise diminishing marginal health gains in their presentation.

It is unclear why the researchers invoke these assumptions, and the ICER-based methods that they give rise to, given the longstanding recognition of the inability of cost-effectiveness methodology to inform decision-making about the efficiency of interventions. For example, standard mathematical programming models were presented to solve the optimization problem facing the decision-maker by Torrance et al. (1972). This involves switching the focus of evaluation away from an individual intervention, its effects and costs, towards the impact of introducing the intervention into the portfolio of interventions on total health gains. Similarly, Chen and Bush (1976) provided a framework for maximizing health care output subject to political and administrative constraints using mathematical programming techniques. Mathematical programming was identified as "the only universal approach to ranking (interventions) under a constraint" by Drummond (1980 p64).

Figure 5 provides a simple presentation of the mathematical programming solution using the same diagrammatic approach of Culyer et al. (2007) but also compatible with the questions being addressed by the surveyors, based on the data provided in Table 1. Consider the case of a decision maker working with a budget of $20 million (point **E*** in figure 5) that is currently spent on two interventions, **A** (incremental costs= $14million, incremental effects = 285 QALYs, ICER=49.1) and **B** (incremental costs = $6million, incremental effects = 55 QALYs, ICER= 109). Because we do not assume perfect divisibility, there is no marginal health gain relevant to the decision-maker. Instead we plot on the vertical axis the incremental health gain (**IHG**) (or incremental benefit from the surveyors' perspective) expressed as the average rate of return on investment (ARRI) (QALYs per

$1million expenditure), which is the inverse of the ICER, in descending order.   The ARRI of interventions **A** and **B** are 20.36 and 9.17 respectively.   The area under the IHG curve **AA'BB'** represents the total health gain from the budget of $20million and is given by the sum of the IHG for both interventions (or the ARRI times by the expenditure on the intervention summed over both interventions). So total health gain for **A** and **B** is 340 QALYs[a].

The right hand 'panel' on the diagram shows the corresponding incremental health gains (or benefits) for interventions **C** (incremental costs = 18million, incremental effects= 346 QALYs, ICER=52) and **D** (incremental costs = $2million, incremental effects= 17QALYs, ICER=117.6) not currently funded from the budget (e.g., new technologies). Note that following Culyer et al. (2007), the ICER for current intervention **B** is larger than the ICER for new intervention **C**. However the resources currently spent on **B** are insufficient to support **C**, so the decision maker cannot simply replace **B** by **C**. Moreover, The ICER for new intervention **C** is greater than the ICER for the other current intervention, **A**, so it would not be appropriate to replace **A** by **C** (and the resources used to support **A** are insufficient to support **C** anyway). So although the ICER threshold approach says **C** should replace **B**, there is no way of doing this without also cancelling **A**. But the ICER threshold approach says **A** is preferred to **C** anyway (it has a lower ICER than **C**).

However, the area under the IHG curve, **CC'DD'** in the right hand panel represents the total health gain if the budget was to be spent on the new interventions **C** and **D**. This is 363 QALYs which exceeds the total health gain of the current interventions from above. This can be seen more clearly by rotating the new interventions IHG curve **CC'DD'** around the vertical line at **E*** to give **CFGJ** in the current interventions panel. In other words, although the ICER for current intervention **A** is less than the ICER for new intervention **C** *and* the ICER for current intervention **B** is less than the ICER for new intervention **D**, moving the entire budget from **A** and **B** and spending it on **C** and **D** would increase total health gain (and hence increase the efficiency of use of available health care resources, **E***).   Use of the ICER threshold rule and implicit assumptions about incremental changes being simple aggregates of constant marginal changes leads the decision-maker to inefficient decisions (i.e., choose the allocation of resources that produces less health gains).

Comparing the areas under the IHG curves essentially represents a graphical form of the mathematical programming procedure and provides an unambiguous approach to determining whether introducing a new intervention represents an improvement in efficiency. Algebraic forms of the mathematical programming approach have been provided elsewhere (e.g., Birch and

---

[a] The marginal valuation of health gains need not be constant. Because we are only interested in health gains then an allocation of available resources that generates more health gain than an alternative allocation of the same resources must produce more social value provided the marginal valuation of health gain remains positive.

Donalsdon,1987) together with restricted (second best) solutions that deal with circumstances where full information on all possible programmes are not available but produce solutions that generate unambiguous improvements in efficiency (e.g. Birch and Gafni 1992).

6. DISCUSSION

"Despite over 20 years of formalised economic evaluation the impact   of   economic evaluation on efficiency of healthcare resource allocation   is hard to ascertain" (Drummond 2012).

We have argued that despite the non-marginal nature of the problems faced by health care decision makers having been recognised by researchers for over 25 years (e.g., Birch and Donaldson 1987), the search for marginal solutions through the identification of ICER thresholds or the social value of a QALY is still pursued in the economic evaluation literature. Ironically, this search adopts incremental data related to differences between interventions as proxies for marginal measures (differences in both costs and benefits between the current quantity of an output and one more unit of output) when it is the incremental data that reflects the choices faced by decision makers. It has been argued that provided the incremental cost of an intervention represents only a very small proportion of a decision-maker's budget the problems associated with indivisibilies of interventions may not be of practical significance (McCabe 2007). But even where the incremental cost is small (relative to the budget), being able to take advantage of the higher average rate of return of the new intervention will depend on interventions with small incremental costs (that also have lower average rates of return) being available for cancellation. If no such interventions are available, it would mean cancelling programmes (or a portion of these programmes, assuming delivering services to only a portion of those in need is ethically acceptable) with large incremental costs which, although having lower average rates of return than the new intervention have much larger total health gains. We cannot simply assume that these can be replaced by adopting the new intervention in a larger size – that would require constant marginal benefit of the new intervention, i.e., it assumes there are always more people in the population with the same level and severity of the condition treated using the new intervention and at constant marginal cost. The problems introduced by the use of incremental costs and health gains to proxy marginal conditions stem from the assumption of constant marginal health gains and cost for all possible interventions, not just the intervention under consideration.

Methodologies were developed many years ago for dealing with the challenges faced by maximisation under a resource constraint in the presence of indivisibilities and non-constant returns to scale. These methodologies have been adapted for health care applications. While both

'searchers' and 'surveyors' remain devoted to the ICER threshold approach, the empirical evidence of the use of this approach remains clear – far from increasing (let alone maximising) the health gain generated from available resources, the approach has been associated with increasing total costs without evidence of any increase in total health gains (Birch and Gafni 2010). In the interests of academic accountability it is perhaps timely to ask the 'searchers' and 'surveyors' to justify further use of the threshold approach in terms of the benefits to society and achieving the social goal of maximizing health gains from available resources.

15

Figure 1. Marginal health gain from current interventions[1]

*Marginal
health gain*
dH/dE

h

a

0

E*

Current interventions

NHS expenditures per year, E
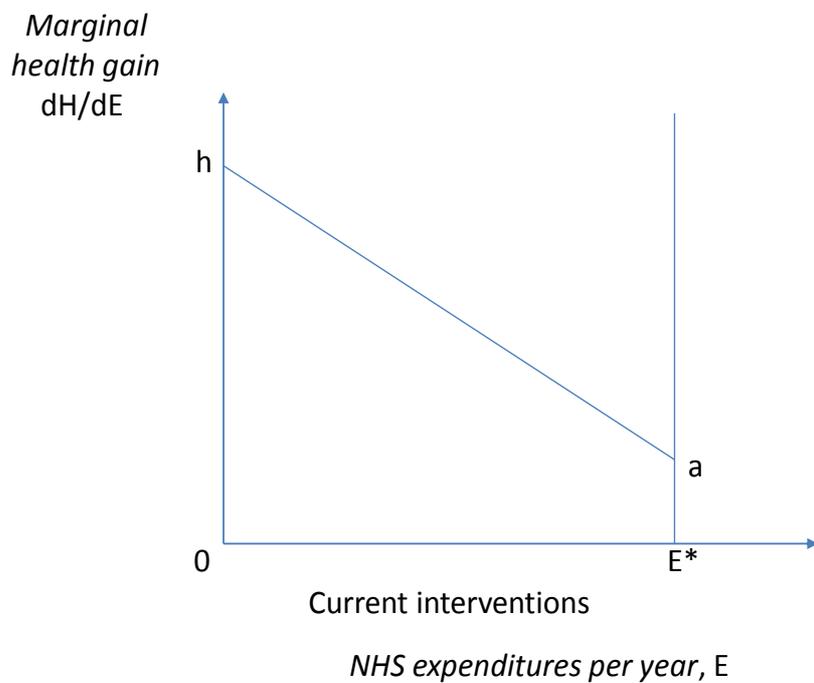
1. Based on Culyer et al. (2007)

Figure 2. Allocating NHS expenditures with new interventions[2]
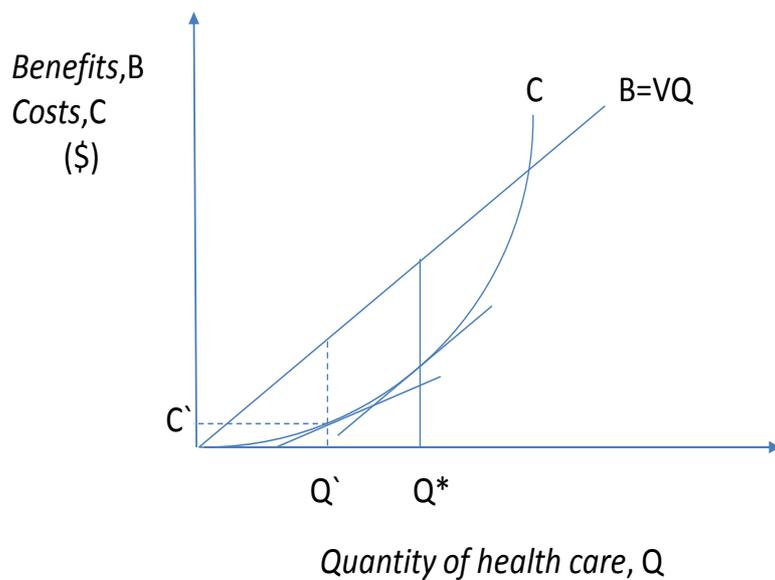

NHS expenditures per year, E
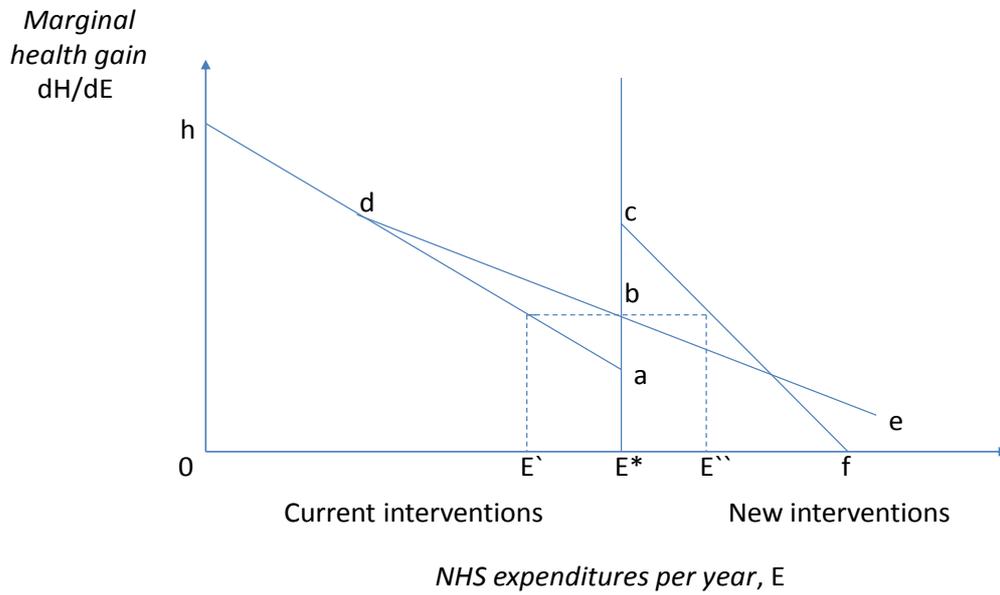
2. Based on Culyer et al. (2007)

Figure 3. Allocating NHS expenditures between interventions of variable size[3]
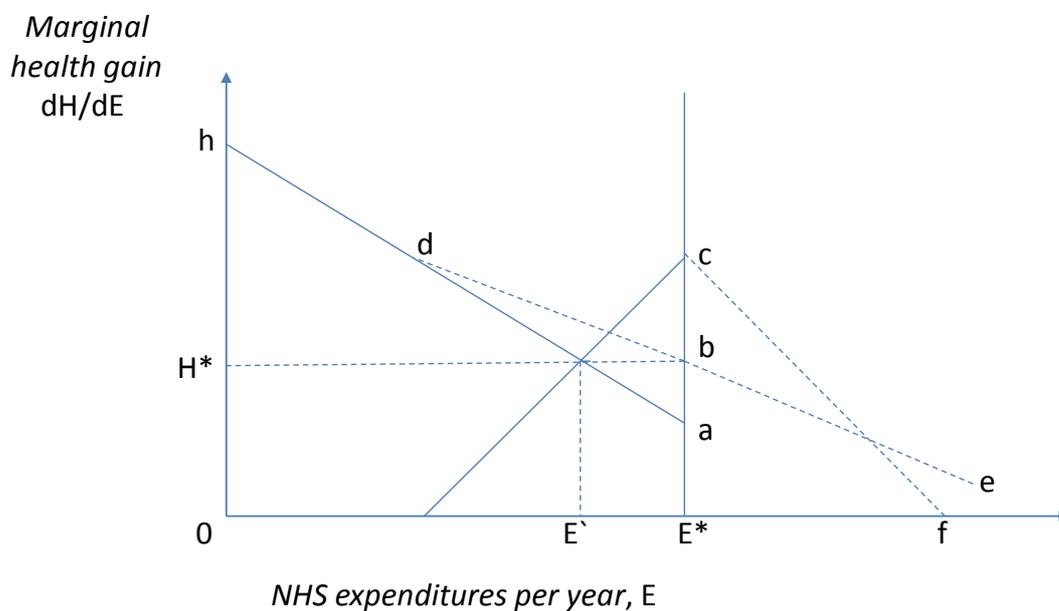
3. Adapted from Birch and Gafni (2003)

Figure 4. Finding a value of a QALY[4]

4. Based on Baker et al. (2011)

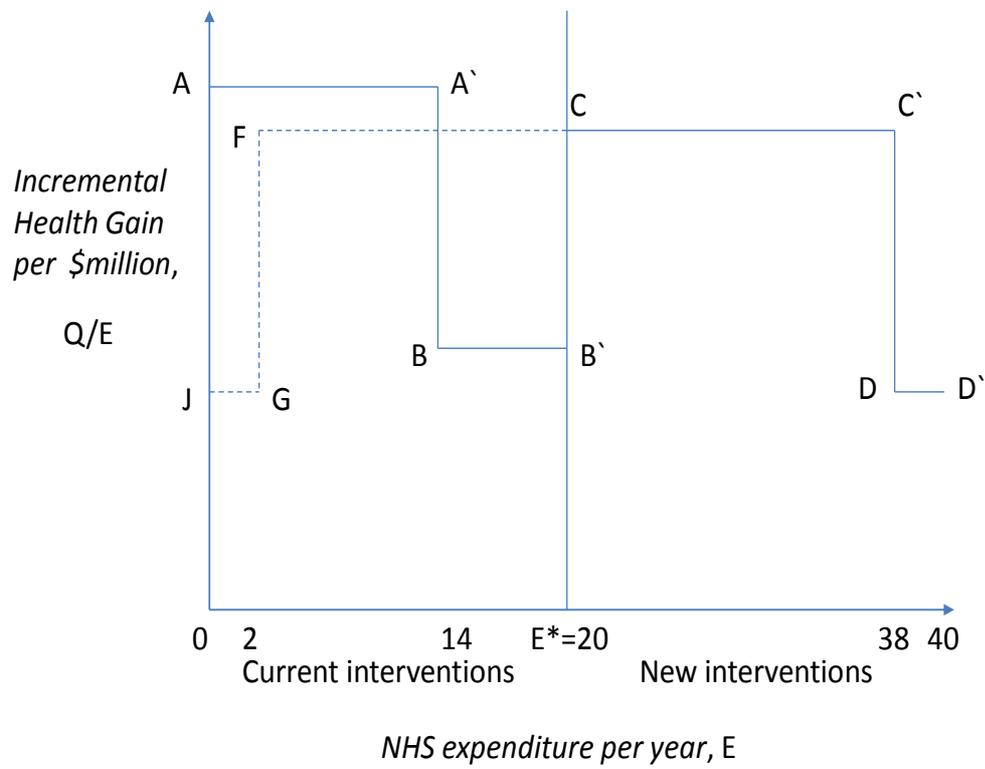**Figure 5. Maximising efficiency under indivisibilities and non-constant returns to scale**

Table 1. Incremental effects and incremental costs of interventions

Intervention	Incremental Effect (QALYs)	Incremental Cost ($mill)	ICER (Cost/QALY)	ARRI (QALY/Cost)
A	285	14	49.1	20.36
B	55	6	109.0	9.17
C	346	18	52.0	19.22
D	17	2	117.6	8.50

References

Baker R, Chilton S, Donaldson C, Jones-Lee M, Lancsar E, Mason H, Metcalf H, Pennington M, Wildman J. (2011) Searchers vs surveyors in estimating the monetary value of a QALY: resolving a nasty dilemma for NICE. *Health Economics, Policy and Law*, 6: 435-447.

Birch S, Donaldson C. (1987) Applications of cost-benefit analysis to health care: Departures from welfare economic theory. *Journal of Health Economics*, 6: 211-225.

Birch S, Gafni A. (1992) Cost-effectiveness/utility analyses: Do current decision rules lead us to where we want to be? *Journal of Health Economics*, 11: 279-296.

Birch S, Gafni A. (2003) Economics and the evaluation of health care programmes: Generalisability of methods and implications for generalisability of results. *Health Policy,* 64: 207-219.

Birch S, Gafni A. (2007) Economists' dream or nightmare? Maximising health gains from available resources using the NICE guidelines. *Health Economics, Policy and Law*; 2; 193-202.

Birch S, Gafni A. (2010) The inconvenient truth of economic evaluation: Benefits forgone as a determinant of efficiency and implications for decision–making in Rosen B, Israeli A and Shortell S (eds.) *Improving health and health care: Who is responsible? Who is accountable? Israel National Institute for Health Policy Research, Jerusalem, Israel.*

Birch S, Leake J, Lewis D. (1996) Economic issues in the development and use of practice guidelines: an application to resource allocation in dentistry. *Community Dental Health,* 13: 70-75.

Chen M., Bush J. (1976) Maximising health system output with political and administrative constraints using mathematical programming. *Inquiry, 13:* 215-227.

Cleary S, Mooney G, McIntyre D (2010) Equity ands efficiency in HIV treatment in South Africa: the contribution of mathematical programming to priority setting. *Health Economics,* 19: 1166-1180.

Culyer A, McCabe C, Briggs A, Claxton K, Buxton M, Akehurst R, Sculpher M, Brazier J. (2007) Searching for a threshold, not setting one; the role of the National Institute for Health and Clinical Excellence*. Journal of Health Services Research and Policy* , 12: 56-58.

Donaldson C, Farrar S. (1993) Needs assessment: Developing an economic approach. *Health Policy*, 25: 95-108.

Donaldson C, Bate A, Mitton C, Dionne F, Ruta D. (2010) Rational disinvestment. *Quarterly Journal of Medicine*,

Drummond M. (1980) *Principles of economic appraisal in health care*. Oxford, Oxford University Press.

Drummond M. (2012) Twenty years of using economic evaluations for reimbursement decisions: what have we achieved? *Centre for Health Economics Research Paper 75*, University of York, York, UK.

Gafni A, Birch S (2006) Incremental cost-effectiveness ratios (ICERs): The silence of the lambda. Social Science and Medicine, 62: 2091-2100.

Laupacis A, Feeny D, Detsky A, Tugwell P (1992) How attractive does a

new technology have to be to warrant adoption and utilization?
Tentative guidelines for using clinical and economic evaluations. Canadian Medical Association
Journal 1992, 146: 473-481.

Lorinc J. (2006) How much are drugs worth? A fledgling medical science attempts an answer.
University of Toronto Magazine 37. [online]. Available from URL:
http://www.magazine.utoronto.ca/06Spring/drugs.asp

McCabe C (2007) The cost effectiveness threshold, Briefing paper for the methods working party,
The National Institute for Health and Clinical Excellence, 2007.

Mitton C, Donaldson C. (2004). *Priority setting toolkit: A guide to the use
 of economics in health care decision-making*. London: BMJ Books.
Mooney G. (2003). *Economics, medicine and health care* (3rd ed.)*. Harlow,
UK: Pearson Education Limited.

Neuhauser D, Lewicki A. (1975) What do we gain from the sixth stool guaiac? *New England Journal
of Medicine,* 293: 226-228.

Tianawit S, Chongsuvivatwong V, Birch S. (2009) Optimizing the mix of basic dental services for
Southern Thai schoolchildren based on resource consumption, service needs and parental
preference. *Community Dentistry and Oral Epidemiology*, 37: 372-380.

Torrance G, Thomas W, Sackett D. (1972) A utility maximization model for evaluation of health care
programmes. *Health Services Research*, 7, 118-133.

Weinstein M. (2012) Decision rules for incremental cost effectiveness analysis, in Jones A. (Ed.)
The Elgar Companion to Health Economics, Second edition. Cheltenham, Edward Elgar.

Weinstein M, Zeckhauser R. (1973) Critical ratios and efficient allocation, *Journal of Public
Economics*, 2:147-157.