Discounting and decision making in the economic evaluation of health care technologies

Karl Claxton[1,2]*, Mike Paulden*[1]*, Hugh Gravelle*[1]*, Werner Bouwer*[3] *and*
Anthony J Culyer[2,4]

[1]*Centre for Health Economics,* [2]*Department of Economics and Related Studies, University of York, UK,*
[3]*Institute for Medical Technology Assessment, Erasmus University,* [4]*Department of Health Policy, Management and Evaluation, University of Toronto*

SUMMARY

Reimbursement authorities around the world differ in their guidance on discounting costs and health benefits in cost-effectiveness studies of health care technologies. The purpose of this paper is to establish appropriate discounting rules, and demonstrate that they depend critically on whether the budget for health care may be regarded as fixed and whether the social objective is to maximise the present consumption value of health or health itself. Discounting costs and health benefits in cost-effectiveness analysis has been the subject of recent debate - some suggesting a common rate for both and others suggesting a lower rate for health. We derive optimal discounting for decision rules based on incremental cost-effectiveness ratios (ICERs) and maximising net health and consumption benefit for a variety of plausible possible circumstances. Using simple numerical examples we show how each of the views found in the literature on discounting turn on key judgments of fact and value; most importantly, on whether the social objective is to maximise discounted health outcomes or the present consumption value of health; on whether the budget for health care is fixed; on the expected growth in the cost-effectiveness threshold; and on the expected growth in the consumption value of health. We demonstrate that if the budget for health care is fixed, growth in the cost-effectiveness threshold may justify a higher discount rate on costs than on health when decisions are based on ICERs. Growth in the consumption value of health implies that the discount rates on health and costs should be less than the consumption interest rate. If the budget for health care is not fixed or is set optimally (i.e. all costs fall on consumption) and maximum consumption value is the social objective then the discount rate on health should be less than on costs. Whether one believes that the objective should be the maximization of the present value of health or the present consumption value of health, the conventional wisdom of discounting both health and costs at the consumption interest rate is valid only under dubious assumptions about values and facts.

INTRODUCTION

Maximising lifetime welfare subject to intertemporal constraints requires that the rate of return to savings should be equal to the rate of pure time preference plus the elasticity of the marginal utility of consumption multiplied by the growth rate of per capita consumption (Ramsey, 1928). These principles have been extended to social choices, requiring the derivation of a social time preference rate (STPR) (Zhuang et al. 2007) and have informed all subsequent debate. In 2003, the UK Treasury recommended in its Green Book that the STPR be adopted as the "standard real discount rate" and employed the classic Ramsey formula to estimate this as 3.5% per annum (HM Treasury 2003) based on three elements: a pure time preference rate of 0.5%, a catastrophic risk premium of 1%, and a further 2% to represent the combined effect of the elasticity of the marginal utility of consumption and the growth in per capita consumption. In 2004, NICE amended its discounting guidance, requiring that cost-effectiveness evaluations discount both costs and health effects at this 3.5 per cent rate (NICE 2004). This change, which has been maintained in the recent revision to guidance in 2008, brought NICE into approximate line with assessment bodies in other countries and other commentators (Lipscomb et al 1996), including the Washington Panel (Gold et al. 1996), which argued that "the case for such global adjustments in CEA [to reflect a belief that the real value of health is increasing over time] for the societal perspective has yet to be fully made" and recommended common discounting at a rate of 3 per cent per annum.

Previously, NICE had required that costs and health effects be discounted at a real annual rate of 6 per cent and 1.5 per cent respectively (NICE 2001), reflecting guidance from the UK Department of Health (1996). This guidance was reissued in 2004 and was closely based on Gravelle & Smith (2000). It recommended that, "If health effects are measured in quantities, e.g. quality adjusted life-years, and the value of health effects is increasing over time, discounting the volume of health effects at a lower rate than costs is a valid method of taking account of the increase in the future value of health effects", suggesting rates of 3.5% for costs and 1.5% for health effects.

Over recent years there has been much debate over the justification for differential discounting of costs and effects. Brouwer et al. (2005) questioned the need for common discounting, describing the 2004 NICE guidance as "an unexpected step back in time", likely to "worsen the cost-effectiveness ratio, especially for preventative interventions". Claxton et al. (2006) described the differential discounting of costs and health effects as illogical, since exogenous budget constraints for health care in each period implies that future costs are equivalent to future health benefits forgone. In response, Gravelle et al. (2007) demonstrated that a policy of differential discounting would be a practical, although not the only, way of accounting for changes in the consumption value of health effects over time and for exogenous health care budgets.

In view of these conflicting recommendations it is not surprising that practitioners and policy makers remain confused about the appropriate discount rates to be applied. In this paper we show that some of these conflicts and contradictions are more apparent than real and arise from different judgements about values and positive empirical questions: i) whether the social objective of the type of health care decision makers

CEA claims to inform is to maximise welfare or health itself; ii) the fixity of the budget for health care and whether it is appropriately set; iii) any expected changes in the cost-effectiveness threshold and the consumption value of health over time; and iv) the social time preference for health and for consumption. We identify the key questions which those taking decisions in the health care sector need to resolve before the appropriate discount rates to apply to costs and health effects can be established.

There are quite profoundly different but quite reasonably held views about the role economic analysis ought to play in social choice (Ng; Sen), particularly in health (Culyer and Brouwer), therefore the paper deals with two broad characterisations how economic evaluation might inform decisions about health care and is organised as follows. Firstly, appropriate discounting when CEA is regarded as informing those responsible for maximising an exogenous objective (the present value of the time stream of health) subject to exogenous budget constraints in each period is examined. The appropriate discount rates are established when decisions are based on either net health benefits or a comparison of the ICER to the current value of the cost-effectiveness threshold. Secondly, how appropriate discounting changes when a wider social objective of maximising welfare (the present consumption value of health) is adopted with and without exogenous budget constraints is then examined. Appropriate discount rates are established when decisions are based on either net present consumption value or a comparison of the ICER to either current values of the threshold or the social consumption value of health. In all these circumstances the discount rates applied to unadjusted health gains, d_h, and to unadjusted additional costs d_c have a number of components such as a social time preference for health and consumption and adjustments for expected growth in the cost-effectiveness threshold and the consumption value of health. Finally, the paper concludes with a discussion of the implications for bodies with responsibility for making health care decisions. The notation used through the following sections is described in Table 1 for reference.

MAXIMISING THE PRESENT VALUE OF HEALTH

Cost-effectiveness analysis is commonly seen as a means of maximising an exogenous objective (the present value of health itself) subject to an exogenous budget constraint (Gold et al, Drummond et al). Decisions can be based on either net health benefit (Phelps and Mushlin, Stinnett and Mulhay) or by comparing the ICER to a cost-effectiveness threshold which represents the cost-effectiveness of the health care which will be displaced (Johansson and Weinstein). Assuming divisibility and constant returns (Birch and Gafni), this is equivalent to a mathematical programming approach where the cost-effectiveness threshold is the reciprocal of the shadow price(s) of a single budget constraint (Stinnett and Palteil) or a series of budget constraints over multiple periods (Epstein et al). In these circumstances CEA cannot be used to make claims about social welfare or the optimality or otherwise of the budget for health care. Its role is more modest, claiming to inform social decisions in health rather than prescribing social choice. It is this role that CEA has tended to play in policy and it fits well with the view (Claxton et al) that bodies such as NICE in the UK can be appropriately treated as the agents of a socially legitimate higher authority which is unable to express an explicit and coherent social welfare function. In these circumstances the agent can be regarded as a delegated authority but one that cannot be asked to improve social welfare, since it cannot be specified. Rather, explicit resources are allocated by the authority and are accompanied by a set of explicit and specific objectives (e.g., to improve health) for the agent to employ. The implications of this process (i.e., the shadow prices of the constraints imposed by the higher authority and what they imply for social time preference for health) are a partial social expression of some unknown underlying latent welfare function.

The allocation problem

The implications for decision rules and discounting can be illustrated using a simple allocation problem where a new technology is compared to a single alternative (e.g., current clinical practice). A social decision maker has estimates of a time stream of both the technology's expected incremental health benefits, Δh_t, and its incremental costs, Δc_t, in each period, t. The social decision maker faces an exogenous budget in each period and has an estimate of the cost-effectiveness threshold, k_t, which is expected to grow at a rate g_k, such that $k_{t+1} = k_t.(1+g_k)$. Assuming that all costs fall on the budget constraint, the expected health forgone in each period due to the additional costs of adopting the technology is $\frac{\Delta c_t}{k_t}$. Since all costs fall on the health care sector (no costs fall on future consumption) all future costs are future health forgone, so discounting incremental costs is simply a proxy for discounting health forgone. The present value of health is derived by discounting future health gained and forgone at a rate r_h [1] which reflects the social decision maker's willingness to trade current for future health, plus any generic or catastrophic risk r_r not otherwise accounted for in the estimates of expected cost and effect.[2]

Decision rules

The present value of health gained and forgone in each of two periods by accepting the technology is given in the table below.

Time Period	*1*	*2*
Present Value of Health Gained	Δh_1	$\frac{\Delta h_2}{(1+r_h)}$
Present Value of Health Forgone	$\frac{\Delta c_1}{k_1}$	$\frac{\Delta c_2}{k_2.(1+r_h)}$

---

[1] How a social time preference rate for health might be established in these circumstances and how this might differ when CEA is used to maximise an explicit social welfare function is discussed later.

[2] These risks tend to be small, although 1% is used by the UK Treasury. The following sections disregard a separate discussion of r_r but it can be regarded as component of r_h or r_c. Whether there is a particular role for generic risks in health care is discussed further in the discussion section.

A decision based on discounted net health benefit would be to accept the technology if:

$$\Delta h_1 + \frac{\Delta h_2}{(1+r_h)} > \frac{\Delta c_1}{k_1} + \frac{\Delta c_2}{k_2.(1+r_h)} \qquad (1)$$

The technology is not accepted unless the health expected to be gained exceeds the health expected to be forgone due to the additional costs of the technology. By calculating health gained and forgone in each period any growth in the threshold has been fully reflected in the future health forgone. Costs are simply a proxy for health forgone, so both costs and health benefits are discounted at the same rate r_h.

Alternatively and equivalently the decision rule can be expressed as a comparison of the ICER to the current period threshold, k_1. However, the discount rates will need to be adjusted to reflect any growth in k_t. The net health benefits in (1) can be rearranged to provide an ICER which should be less than k_1 if the technology is to be accepted:

Since $k_1 > 0$,
$$k_1.\left[\frac{\Delta c_1}{k_1} + \frac{\Delta c_2}{k_2.(1+r_h)}\right] < k_1.\left[\Delta h_1 + \frac{\Delta h_2}{(1+r_h)}\right]$$

$$\Delta c_1 + \frac{k_1}{k_2}\frac{\Delta c_2}{(1+r_h)} < k_1.\left[\Delta h_1 + \frac{\Delta h_2}{(1+r_h)}\right]$$

Since $\Delta h_1 + \frac{\Delta h_2}{(1+r_h)} > 0$,
$$\frac{\Delta c_1 + \frac{k_1}{k_2}\frac{\Delta c_2}{(1+r_h)}}{\Delta h_1 + \frac{\Delta h_2}{(1+r_h)}} < k_1$$

Since $k_2 = k_1.(1+g_k)$, $\frac{k_1}{k_2} = \frac{1}{(1+g_k)}$,
$$\frac{\Delta c_1 + \frac{\Delta c_2}{(1+r_h).(1+g_k)}}{\Delta h_1 + \frac{\Delta h_2}{(1+r_h)}} < k_1 \qquad (2)$$

It is plausible to a assume that r_h and g_k are small so that $(1+g_k)(1+r_h) = (1+r_h+g_k+r_h.g_k) \approx (1+r_h+g_k)$ and (2) approximates to:

$$\frac{\Delta c_1 + \frac{\Delta c_2}{(1+r_h+g_k)}}{\Delta h_1 + \frac{\Delta h_2}{(1+r_h)}} < k_1 \qquad (3)$$

Comparing the ICER to the current period threshold will be appropriate if a discount rate of

$$d_h = r_h \tag{4}$$

is applied to health gained and

$$d_c \approx r_h + g_k = d_h + g_k \tag{5}$$

is applied to unadjusted costs. That is, the discount rate for health gained is the decision maker's rate of time preference for health, r_h, whilst the discount rate for costs is approximately equal to r_h, plus the growth rate of the cost-effectiveness threshold, g_k. Thus differential discounting of costs and health effects is justified when $g_k > 0$ and requires a higher discount rate for costs. The intuition is that if the threshold is expected to grow then future costs are less important because they lead to less health forgone. Therefore less weight ought to be placed on future costs. One way to achieve this is by discounting them at a higher rate.

There are many different ways in which such adjustments could in principle be made. For example, costs could be initially adjusted using g_k, and then be discounted at the same rate, r_h, as health benefits. However, applying a discount rate ($d_c \approx r_h + g_k$) to the unadjusted stream of costs in (3) seems most intuitive, straightforward and consistent with the way discounting is commonly done in CEA. The key point, however, is not how one allows for $g_k \neq 0$ (differential discounting of unadjusted costs or common discounting of adjusted costs) but whether one does so.

Table 2 illustrates this point with a numerical example which is used through out the following sections. In the example, a constant threshold ($g_k = 0$) implies that the technology should be rejected. Decisions based on net health benefit or ICER are equivalent when using a common discount rate of r_h. However, if the threshold is expected to grow then future additional costs displace less health and the technology should be accepted. This is fully captured in the health forgone each period when calculating net health benefits. However, some adjustment must be included in d_c when the decision is based on a comparison of the ICER with k_1. The example demonstrates that decisions based on either net health benefit or ICERs will be equivalent only as long as an adjustment is made for any expected growth in the threshold.

MAXIMISING THE PRESENT CONSUMPTION VALUE OF HEALTH

The role of economic analysis in informing social choice traditionally takes a wider view of social objectives than simple sector-specific outcome (Mishan, others?). Such a role for CEA is more ambitious; providing a means of making statements about social welfare but also requiring the specification of an explicit social welfare function which will have more than mere health as its arguments. This view of CEA is well represented in the literature. The definition of social welfare can be based on individual preferences, expressed or revealed (a welfarist view) or modified by other social arguments (an extra welfarist view) (Garber and Phelps, Meltzer and Culyer and Brouwer). Importantly, it enables prescriptions for social choice to be made rather than simply informing social decision makers within the confines of the health care system. It allows claims to be made about the optimality or otherwise of any exogenous budget constraints and can also inform decisions when the health care system does not face fixed budgets imposed by some higher authority. This use of CEA is less well represented in policy, partly due to the difficulty of identifying any welfare function carrying some broad consensus or social legitimacy (Arrow, Sen and others). Nevertheless, it is important to examine the implications as health must inevitably be traded with other welfare arguments, most notably consumption, by social decision makers: whether this is done implicitly by a higher authority setting the budget constraints or more explicitly using some explicit welfare function.

An exogenous budget constraint

The wider social objective of maximising social welfare is consistent with maximizing the present consumption value of health if the social welfare function is assumed to depend only on the time streams of health and consumption. If the health care budget each period is regarded as fixed, and assuming that all costs fall on the health care sector, there will be no effects on consumption. Maximising social welfare is thus equivalent to maximising the present consumption value of health, even when the welfare function is not separable in health and consumption (Gravelle *et al.* 2007). The consumption value of health in period t is v_t (the amount of consumption in period *t* that is equivalent to 1 unit of health in period *t*) and is expected to grow over time at a rate g_v, such that $v_{t+1} = v_t.(1+g_v)$. Since all costs fall on the health care sector (no costs fall on consumption), all future costs are future health forgone which are also valued at v_t in the same way as health gained. Therefore, discounting incremental costs is simply a means of discounting the consumption value health forgone. The net present consumption value of health is derived by discounting future health gained and forgone at a rate r_c (the proportionate increase in consumption in period t+1 which would be equivalent to a 1 unit increase in period t). The budget in each period can be regarded as set optimally with respect to this social welfare function if $v_t = k_t$.

Decision rules

The present consumption value of health gained and forgone in each of two periods by accepting the technology is given in the table below.

Time Period	*1*	*2*
Present Consumption Value of Health Gained	$v_1.\Delta h_1$	$\frac{v_2.\Delta h_2}{(1+r_c)}$
Present Consumption Value of Health Forgone	$\frac{v_1.\Delta c_1}{k_1}$	$\frac{v_2.\Delta c_2}{k_2.(1+r_c)}$

A decision based on the net present consumption value of health would be to accept the technology if:

$$v_1.\Delta h_1 + \frac{v_2.\Delta h_2}{(1+r_c)} > \frac{v_1.\Delta c_1}{k_1} + \frac{v_2.\Delta c_2}{k_2.(1+r_c)} \qquad (6)$$

The technology is accepted only if the present consumption value of health gained exceeds the present consumption value of the health forgone. Any growth in the threshold is fully reflected in future health forgone and any growth in the consumption value of health is also fully reflected in both the value of health gained and forgone, so both costs and health benefits are discounted at a common rate of r_c.

Equivalently this decision can be expressed as a comparison of the ICER to the current period threshold but discount rates will need to be adjusted to reflect any growth in v_t as well as k_t. The net present consumption value in (6) can be rearranged into an ICER by multiplying both sides of (6) by $k_1 / v_1[\Delta h_1 + (v_2 / v_1)\Delta h_2 / (1+r_c)]$ and using the definition of the growth rates of the consumption value of health ($v_2 / v_1 = 1+g_v$) and the threshold ($k_2 / k_1 = 1+g_k$) to get:

$$\frac{\Delta c_1 + \frac{k_1}{k_2}\frac{v_2}{v_1}\frac{\Delta c_2}{(1+r_c)}}{\Delta h_1 + \frac{v_2}{v_1}\frac{\Delta h_2}{(1+r_c)}} = \frac{\Delta c_1 + \frac{(1+g_v).\Delta c_2}{(1+r_c).(1+g_k)}}{\Delta h_1 + \frac{(1+g_v).\Delta h_2}{(1+r_c)}} < k_1 \qquad (7)$$

Using the plausible assumption that r_c, g_v and g_k are small so that $(1+r_c)/(1+g_v) \approx (1+r_c-g_v)$ and $(1+r_c)(1+g_k)/(1+g_v) \approx (1+r_c+g_k-g_v)$ means that (7) approximates to:

$$\frac{\Delta c_1 + \frac{\Delta c_2}{(1+r_c+g_k-g_v)}}{\Delta h_1 + \frac{\Delta h_2}{(1+r_c-g_v)}} < k_1 \qquad (8)$$

Comparing the ICER to the current period threshold will be appropriate if a discount rate of

$$d_h \approx r_c - g_v \tag{9}$$

is applied to heath gained and

$$d_c \approx r_c + g_k - g_v = d_h + g_k \tag{10}$$

is applied to unadjusted costs. That is, the discount rate for health gained is approximately equal to the social rate of time preference for consumption, r_c, *minus* the growth rate in the consumption value of health, g_v, reflecting the more valuable future health in terms of consumption when $g_v > 0$. The discount rate for incremental costs is the social rate of time preference for consumption, r_c, *minus* g_v (because future health forgone will also be more valuable), *plus* the growth rate of the cost-effectiveness threshold, g_k, adjusting for the fact that future costs will displace less future health if $g_k > 0$.

This suggests that any growth in the consumption value of health should lead to a lower discount rate for *both* costs and health benefits. The intuition seems clear: future costs are simply future health forgone so if future health gains are valued more highly then future costs must also be valued more highly as well. One way to reflect these higher future values is to discount both costs and health gains at a lower rate than consumption.[3] As previously differential discounting of costs and health effects is appropriate only when $g_k > 0$ and requires a higher discount rate applied to costs. Simply observing $g_v > 0$ does not itself justify differential discounting. An additional assumption about the nature of social choice and the way budgets are set by the higher authority is required. For example, if budgets are judged to have been set optimally with respect to the consumption value of health and the social welfare function it presupposes, then $v_t = k_t$ and $g_k = g_v > 0$, i.e., the recommendation of a 1.5% rate for health and 3.5% for costs made by the UK Department of Health would be appropriate if $r_c = 0.035$, $g_v = 0.02$, and NHS budgets fully reflected this growth in consumption value so that $g_k = g_v$[4].

This is illustrated in Table 3 using the same numerical example. Even when $v_t > k_t$, so budgets are suboptimal (with respect to the welfare function) and the health care system is underfunded. decisions based on the present value of consumption or health will differ only insofar as r_c is not equal to r_h, i.e., a consumption value of health

---

[3] As discussed in section 2 there are many different ways in which adjustments for g_v and g_k could be made. However applying d_h and d_c to the unadjusted time streams of costs and health benefits in (6) seem intuitive and is consistent with the way discounting tends to be applied in CEA.

[4] Of course budgets do not necessarily need be optimal with respect to the welfare function, all that is required is $g_k = g_v$ even if $v_t \neq v_t$.

greater than the cost effectiveness threshold simply ‘scales up’ both health gained and forgone. In this example the higher rate of $r_c = 0.035$ rather than $r_h = 0.02$ in Table 2 makes the technology less cost-effective because there is more net health gain in period 2 and the same decision to reject the technology should be made even though $v_t > k_t$ and ICER< v_t .

Any growth in the consumption value of health leads to lower discount rate for *both* costs and health benefits. In this example the lower common rate of $r_c - g_v = 0.015$ is slightly lower then $r_h = 0.02$ in Table 2 so the technology appears marginally more cost-effective but the correct decision remains to reject the technology. Of course there maybe many other circumstances when failure to use a lower common discount rate to reflect growth in consumption value of health will lead to suboptimal decisions.

No exogenous constraint

Previous sections have examined a health care system that resembles the NHS in some key respects: one that is collectively funded with budget constraints fixed by a central and socially legitimate authority. However, other health care systems do not resemble the NHS in this respect - they do not have exogenous budget constraints imposed by some higher authority. Rather, the resources devoted to health care maybe determined by adopting those health technologies for which the individual (or some social) consumption value of the health gained exceeds the consumption costs. In such a health care system, all the incremental costs of accepting technologies fall not on health but on wider consumption. In these circumstances, the notion of a threshold representing the shadow price of a binding budget constraint is redundant and a decision based on the net present consumption value of health would be to accept the technology if:

$$v_1.\Delta h_1 + \frac{v_2.\Delta h_2}{(1+r_c)} > \Delta c_1 + \frac{\Delta c_2}{(1+r_c)} \qquad (11)$$

The technology is accepted only if the present consumption value of health gained exceeds the present value of consumption losses – since there is no fixed budget constraint no health is forgone. Any growth in the consumption value of health gained is fully reflected in each period’s net consumption benefit so both costs and health benefits are discounted at common rate r_c .

Equivalently this can also be expressed using an ICER which must be compared to the current period consumption value of health, v_1, since k_1 is now redundant. As before discount rates must be adjusted to reflect any growth in v_t . The net present consumption value in (11) can be rearranged into an ICER which should be less then v_1 if the technology is to be accepted:

$$\frac{\Delta c_1 + \dfrac{\Delta c_2}{(1+r_c)}}{\Delta h_1 + \dfrac{(1+g_v).\Delta h_2}{(1+r_c)}} < v_1 \qquad (12)$$

Which, assuming r_c and g_v are small approximates to:

$$\frac{\Delta c_1 + \dfrac{\Delta c_2}{(1+r_c)}}{\Delta h_1 + \dfrac{\Delta h_2}{(1+r_c - g_v)}} < v_1 \qquad (13)$$

Comparing the ICER to the current period consumption value of health will be appropriate if a discount rate of

$$d_h \approx r_c - g_v \qquad (14)$$

is applied to heath gained and

$$d_c = r_c \qquad (15)$$

is applied to unadjusted costs. That is, the discount rate for health gained is approximately equal to the social time preference rate for consumption, r_c, minus the growth rate in the consumption value of health g_v, reflecting the greater value of future health in terms of consumption when $g_v > 0$. However, the discount rate for incremental costs is simply r_c, since all cost are consumption losses; no health is forgone. Here, differential discounting of costs and health effects is justified by observing $g_v > 0$ alone and a lower rate for health is implied rather than a higher rate for costs. These discount rates will be identical to those reported in (9) and (10) when exogenous budget constraints are imposed as long at they are optimally set with respect to the social welfare function ($k_t = v_t$ so $g_k = g_v$), i.e., the fixity or otherwise of the constraint is irrelevant for discounting if budgets are regarded as optimal since same discount rates apply albeit for different reasons which require different assessments to be made by policy makers. Both require an assessment of v_t and g_v, but if budgets are in fact fixed then an assessment of k_t and g_k is also required. In addition, a sustainable claim as to why budgets are necessarily set optimally with respect to the social welfare function is also needed.

The general issues are illustrated in Table 4 using the same simple numerical example. Unlike previously where opportunity costs fell on health, now with no binding constraints they fall on consumption and the technology should be accepted since $v_t > ICER$. Any growth in the consumption value of health ($g_v = 0.02$) affects only the valuation of the benefits. This can be reflected in discounting health gains at a lower rate than the consumption costs, reducing the ICER and making the technology more cost-effective.

DISCUSSION

In this paper we have shown that the appropriate discount rates to apply depend on different judgements about social values and positive empirical questions. These include: i) whether the social objective is to maximise welfare (the consumption value of health) or health itself; ii) the fixity of the budget for health care and whether it is appropriately set; iii) any expected changes in the cost-effectiveness threshold and the consumption value of health over time; and iv) the social time preference for health and for consumption. The discount rate to apply also depends on how cost-effectiveness is represented. For example, if cost-effectiveness is expressed as net health benefit or net present consumption value then any growth in k_t and v_t will have been fully accounted for. If, as is commonly the case, cost-effectiveness is expressed as an ICER then some adjustment must be made for any expected growth in k_t and v_t and one way to do so is to make some adjustment to the discount rates applied to costs and health effects.

The almost universal prescription by government bodies and regulatory institutions is that health effects and costs should be discounted at the same rate. Most recommend a rate which, implicitly at least, represents the social time preference rate for consumption, r_c. In the case of NICE this is made explicit where the common rate of 3.5% is justified with reference to the UK Treasury estimate of r_c But, whether a social objective of the present value of health or the present consumption value of health is adopted, a common rate which is equal to r_c is correct only if one is willing to make value or factual judgements that seem to us to be dubious.

Irrespective of social objectives, the common discounting of health and costs to form an ICER rests on the implicit assumption of $g_k = 0$. This is in principle a positive empirical question which rests on the expected growth in the budget for health care and the growth in the productivity of health care technologies. If the budget increases over time but the technology is fixed then the marginal project in a period becomes less productive of health i.e., $g_k > 0$. On the other hand, the faster is technological progress in health care, or the costs of existing technologies are reduced, for example due to patent expiry and generic entry, the more likely it is that the marginal cost of producing health falls over time and $g_k < 0$. Therefore, the conventional wisdom of a common rate rests on the constancy of the cost-effectiveness threshold over the type of time horizons relevant in CEA.

However, which rate ought to be applied does critically depend on the objectives of social choice in health care. If the social objective is taken to be maximising the present value of health then this implies that the discount rate on health effects should be $d_h = r_h$ and the discount rate on costs should be $d_c \approx r_h + g_k$ where $1 + r_h$ is the rate at which society is willing to trade future for current health. Thus, the conventional wisdom is correct only if $g_k = 0$ and $r_h = r_c$, i.e., if the threshold remains constant and the rate at which a social decision maker is willing to trade current and future health is equal to the rate they are willing to trade current and future consumption. There seems no reason to suppose that $r_h = r_c$, indeed there are good reasons to believe that it should not, particularly if one accepts that the type of explicit social welfare function that implies r_c can not adequately describe social choice. Rather, the resources devoted to health care over time and the resulting health in each period represents society's willingness to trade current and future health, i.e., the choices of the higher authority in setting budgets implies a value for r_h: a partial social expression of some unknown, latent welfare function that may include many conflicting social arguments, for example inter-temporal health equity concerns among others.

If the social objective is taken to be maximising the present consumption value of health, the discount rate on health effects should be $d_h \approx r_c - g_v$ and the discount rate on costs should be $d_c \approx r_c + g_k - g_v$. Hence from this viewpoint the conventional wisdom is correct only if, as before, the cost-effectives threshold is constant ($g_k = 0$) and the consumption value of health is also constant over time ($g_v = 0$). There are two arguments which suggest that an assumption of $g_v = 0$ is dubious. The first is based on the conventional welfare theoretic approach which is often used in deriving the consumption interest rate to be used in public projects, exemplified by the UK Treasury Green Book. The social rate of time preference for consumption r_c is determined by the ratio of the welfare gain from an increase in future consumption to the welfare gain from an increase in current consumption. Since the health of the population is expected to grow more slowly than its consumption of goods and services, as time passes, the welfare gain from better health will increase relative to the welfare gain from increases in consumption: the social value of health in terms of consumption will increase so $g_v > 0$. Secondly, if one is willing to assume that individual decisions when faced with tradeoffs of health or risks against income are a valid source of welfare information then there is also empirical evidence which suggests that $g_v > 0$. Many studies suggest that the value of a statistical life increases with income and hence increases over time as income is expected to grow (Costa and Kahn (2004), Hammitt et al (2006), Hall and Jones (2007), Viscusi and Aldey (2003)).

Irrespective of questions of social value the conventional wisdom of a common discount rate representing the social rate of time preference for consumption rests on implicit assumptions that are difficult to support. The question of whether a common rate is justified rests on a factual empirical question of whether the cost-effectiveness threshold is expected to be constant or grow over time horizons relevant to CEA. Even when a common rate is justified, which rate ought to be applied to costs and health effects does depend critically on questions of social value. However, there is little justification for $r_h = r_c$ when health is the maximand or $g_v = 0$ when consumption value is the social objective. Therefore a common rate of r_c seems difficult to sustain irrespective of social objectives in health care.

REFERENCES

Arrow KJ. 1950. A Difficulty in the Concept of Social Welfare. Journal of Political Economy 58(4): 328–346.

Birch S, Gafni A. 1993. Changing the problem to fit the solution: Johannesson and Weinstein's (mis)application of economics to real world problems. Journal of Health Economics 12, 469-476.

Brouwer WBF, Niessen LW, Postma MJ, Rutten FFH. 2005. Need for differential discounting of costs and health effects in cost effectiveness analyses. British Medical Journal 331: 446-448.

Brouwer W, Culyer AJ, Job N, van Exel A, Rutten FH. 2008. Welfarism vs. extra-welfarism. Journal of Health Economics.

Brouwer W, Koopmanschap M. 2000. On the economic foundations of CEA. Ladies and gentlemen, take your positions! Journal of Health Economics 19: 439-459.

Claxton K, Sculpher MJ, Culyer AJ. 2007. Mark versus Luke? Appropriate methods for the evaluation of public health interventions. Centre for Health Economics, University of York. CHE Research Paper 31.

Claxton K, Sculpher M, Culyer A et al. 2006. Discounting and cost-effectiveness in NICE - stepping back to sort out a confusion. Health Economics 15(1): 1-4.

Costa DL, Kahn ME. Changes in the value of life, 1940–1980. Journal of Risk and Uncertainty, XXIX (2004), 159–180.

Culyer AJ, McCabe C, Briggs A, Claxton K, Buxton M, Akehurst R, Sculpher M, Brazier J. Searching for a threshold, not setting one: the role of the National Institute of Health and Clinical Excellence. Journal of Health Services Research and Policy, 2007, 12(1); 56-58.

Department of Health. 1996. Policy Appraisal and Health: A Guide from the Department of Health. Department of Health: London.

Department of Health. 2004. Policy Appraisal and Health: A Guide from the Department of Health (Reissued). Department of Health: London.

Drummond MF, Sculpher MJ, O'Brien B, Stoddard GL, Torrance GW. 2005. Methods for the Economic Evaluation of Health Care Programmes (3rd edn). Oxford University Press: Oxford.

Epstein D, Chalabi Z, Claxton K, Sculpher M. 2007. Efficiency, equity and budgetary policies: informing decisions using mathematical programming. Medical Decision Making 27: 128-37.

Garber AM, Phelps CE. 1997. Economic foundations of cost-effectiveness analysis. Journal of Health Economics 16: 1-31.

Gold MR, Siegel JE, Russell LB, Weinstein MC (eds). 1996. Cost-effectiveness in Health and Medicine. Oxford University Press: Oxford.

Gravelle H, Brouwer WBF, Niessen LW, Postma MJ, Rutten FFH. 2007. Discounting in economic evaluations: stepping forward towards optimal decision rules. Health Economics 16(3): 307-317.

Gravelle H, Smith D. 2001. Discounting for health effects in cost-benefit and cost-effectiveness analysis. Health Economics 10: 587-599.

Garber AM, Phelps CE. 1997. Economic foundations of cost-effectiveness analysis. Journal of Health Economics 16: 1-31.

Hall RE, Jones CI. The value of life and the rise in health spending. Quarterly Journal of Economics, February 2007, Vol. 122 (1), pp. 39-72.

Hammitt JK, Liu J-T, Liu J-L. Survival is a luxury good: the increasing value of a statistical life. Mimeo. 2006. http://unjobs.org/authors/jin-tan-liu Accessed 19 February 2008.

HM Treasury. 2003. Appraisal and Evaluation in Central Government. HM Treasury: London.

Johannesson M, Weinstein MC. 1993. On the decision rules of cost-effectiveness analysis. Journal of Health Economics 12: 459–467.

Meltzer D. 1997. Accounting for future costs in medical cost-effectiveness analysis. Journal of Health Economics 16: 33-64.

Mishan E. 1967. Criteria for Public Investment: Some Simplifying Suggestions. Journal of Political Economy 75: 139-46.

Ng YK. 1983. Welfare economics: introduction and development of basic concepts. MacMillan.

NICE. 2001. Technical Guidance for Manufacturers and Sponsors on Making Submissions for a Technology Appraisal. NICE: London, March 2001.

NICE. 2004. Guide to the Methods of Technology Appraisal. NICE: London, April 2004.

Phelps CE, Mushlin AI. 1991. On the near equivalence of cost-effectiveness analysis. International Journal of Technology Assessment in Health Care 17: 12-21.

Ramsey FP. 1928. A Mathematical Theory of Saving. The Economic Journal 38: 543-559.

Sen A. 1970. The Impossibility of a Paretian Liberal. Journal of Political Economy 78(1): 152-57.

Stinnett AA, Mullahy J. 1998. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical Decision Making 18(suppl): S68-S80.

Stinnett AA, Paltiel D. 1996. Mathematical programming for the efficient allocation of health care resources. Journal of Health Economics 15: 641-653.

Viscusi WK, Aldy JE. The value of a statistical life: a critical review of market estimates throughout the world," Journal of Risk and Uncertainty, XXVII (2003), 5–76.

Weinstein MC, Manning WG. 1997. Theoretical issues in cost-effectiveness analysis. *Journal of Health Economics* 16: 121-128.

Zhuang J, Liang Z, Lin T, De Guzman F. 2007. Theory and Practice in the Choice of Social Discount Rate for Cost-Benefit Analysis: A Survey. ERD Working Paper Series No. 94. Asian Development Bank: Philippines

TABLES

Table 1: Notation used throughout in the order in which it appears

Symbol	Definition
d_h	Discount rate applied to unadjusted health gains
d_c	Discount rate applied to unadjusted costs
Δh_t	Incremental health benefits in period t
Δc_t	Incremental costs in period t
k_t	Cost-effectiveness threshold in period t
g_k	Growth rate of the threshold $((k_{t+1}/k_t) - 1)$
r_h	Social time preference rate for health (the proportional increase in health in period *t*+1 that a social decision maker regards as equivalent to 1 unit increase in period *t* health)
v_t	The consumption value of health in period t (the amount consumption in period t regarded as equivalent to 1 unit of health in period t)
g_v	Growth rate of consumption value of health $((v_{t+1}/v_t) - 1)$
r_c	Social time preference rate for consumption (the proportional increase in consumption in period *t*+1 that is equivalent to a 1 unit increase in consumption in period *t*)

Table 2: Simple numerical example (k_1 = £20,000, r_h = 2%)

t	1	2
Δc_t	£150,000	£150,000
Δh_t	5	10

		$g_k = 0\%$ (k_2 = £20,000)	$g_k = 5\%$ (k_2 = £21,000)
Present value of health gained	$\Delta h_1 + \frac{\Delta h_2}{(1+r_h)}$	$5 + \frac{10}{1.02} = 14.804$	$5 + \frac{10}{1.02} = 14.804$
Present value of health forgone	$\frac{\Delta c_1}{k_1} + \frac{\Delta c_2}{k_2.(1+r_h)}$	$\frac{£150k}{£20k} + \frac{£150k}{£20k * 1.02} = 14.853$	$\frac{£150k}{£20k} + \frac{£150k}{£21k * 1.02} = 14.503$
Net health benefit (NHB)		$14.804 - 14.853 = -0.049$	$14.804 - 14.503 = 0.301$
Decision	*Accept if* $NHB > 0$	Reject	Accept
Present value of health gained	$\Delta h_1 + \frac{\Delta h_2}{(1+r_h)}$	$5 + \frac{10}{1.02} = 14.804$	$5 + \frac{10}{1.02} = 14.804$
Present value of incremental cost	$\Delta c_1 + \frac{\Delta c_2}{(1+r_h+g_k)}$	$£150k + \frac{£150k}{1.02} = £297{,}059$	$£150k + \frac{£150k}{1.07} = £290{,}187$
ICER		$\frac{£297{,}059}{14.804} = £20{,}066$	$\frac{£290{,}187}{14.804} = £19{,}602$
Decision	*Accept if* $ICER < k_1$	Reject	Accept

Table 3: Simple numerical example ($k_1 = £20,000$, $v_1 = £50,000$, $r_c = 3.5\%$)

		$g_k = 0\%$, $g_v = 0\%$ ($k_2 = £20,000$, $v_2 = £50,000$)	$g_k = 0\%$, $g_v = 2\%$ ($k_2 = £20,000$, $v_2 = £51,000$)	$g_k = 5\%$, $g_v = 2\%$ ($k_2 = £21,000$, $v_2 = £51,000$)
Present (consumption) value of health gained	$v_1.\Delta h_1 + \frac{v_2.\Delta h_2}{(1+r_c)}$	$£50k*5+\frac{£50k*10}{1.035} = £733,092$	$£50k*5+\frac{£51k*10}{1.035} = £742,754$	$£50k*5+\frac{£51k*10}{1.035} = £742,754$
Present (consumption) value of health forgone	$\frac{v_1.\Delta c_1}{k_1} + \frac{v_2.\Delta c_2}{k_2.(1+r_c)}$	$\frac{£50k*£150k}{£20k} + \frac{£50k*£150k}{£20k*1.035} = £737,319$	$\frac{£50k*£150k}{£20k} + \frac{£51k*£150k}{£20k*1.035} = £744,565$	$\frac{£50k*£150k}{£20k} + \frac{£51k*£150k}{£21k*1.035} = £726,967$
Net consumption benefit (NCB)		$£733,092 - £737,319 = -£4227$	$£742,754 - £744,565 = -£1,811$	$£742,754 - £726,967 = £15,787$
Decision	*Accept if* $NB > 0$	Reject	Still reject	Accept
Present value of health gained	$\Delta h_1 + \frac{\Delta h_2}{(1+r_c-g_v)}$	$5+\frac{10}{1.035}=14.662$	$5+\frac{10}{1.015}=14.852$	$5+\frac{10}{1.015}=14.852$
Present value of incremental cost	$\Delta c_1 + \frac{\Delta c_2}{(1+r_c+g_k-g_v)}$	$£150k+\frac{£150k}{1.035} = £294,928$	$£150k+\frac{£150k}{1.015} = £297,783$	$£150k+\frac{£150k}{1.065} = £290,845$
ICER		$\frac{£294,928}{14.662} = £20,115$	$\frac{£297,783}{14.852} = £20,050$	$\frac{£290,845}{14.852} = £19,583$
Decision	*Accept if* $ICER < k_1$	Reject	Still reject	Accept

Table 4: Simple numerical example ($v_1 = £50{,}000$, $r_c = 3.5\%$)

		$g_v = 0\%$ ($v_2 = £50{,}000$)	$g_v = 2\%$ ($v_2 = £51{,}000$)
Present (consumption) value of health gained	$v_1.\Delta h_1 + \frac{v_2.\Delta h_2}{(1+r_c)}$	$£50k * 5 + \frac{£50k * 10}{1.035} = £733{,}092$	$£50k * 5 + \frac{£51k * 10}{1.035} = £742{,}754$
Present value of consumption forgone	$\Delta c_1 + \frac{\Delta c_2}{(1+r_c)}$	$£150k + \frac{£150k}{1.035} = £294{,}928$	$£150k + \frac{£150k}{1.035} = £294{,}928$
Net consumption benefit (NCB)		$£733{,}092 - £294{,}928 = £438{,}164$	$£742{,}754 - £294{,}928 = £447{,}826$
Decision	*Accept if* $NCB > 0$	Accept	Still accept
Present value of health gained	$\Delta h_1 + \frac{\Delta h_2}{(1+r_c-g_v)}$	$5 + \frac{10}{1.035} = 14.662$	$5 + \frac{10}{1.015} = 14.852$
Present value of incremental cost	$\Delta c_1 + \frac{\Delta c_2}{(1+r_c)}$	$£150k + \frac{£150k}{1.035} = £294{,}928$	$£150k + \frac{£150k}{1.035} = £294{,}928$
ICER		$\frac{£294{,}928}{14.662} = £20{,}115$	$\frac{£294{,}928}{14.852} = £19{,}858$
Decision	*Accept if* $ICER < v_1$	Accept	Still accept