THE UNIVERSITY of York

CHE
Centre For Health Economics

Budget Allocation and the Revealed Social Rate of Time Preference for Health

CHE Research Paper 53

Budget Allocation and the Revealed Social Rate of Time Preference for Health

Mike Paulden[1,2]
Karl Claxton[1,3]

[1]Centre for Health Economics, University of York, UK
[2]Toronto Health Economics and Technology Assessment (THETA) Collaborative, University of Toronto, Canada
[3]Department of Economics and Related Studies, University of York, UK

October 2009

Background

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The new CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Disclaimer

Papers published in the CHE Research Paper (RP) series are intended as a contribution to current research. Work and ideas reported in RPs may not always represent the final position and as such may sometimes need to be treated as work in progress. The material and views expressed in RPs are solely those of the authors and should not be interpreted as representing the collective views of CHE research staff or their research funders.

Further copies

Copies of this paper are freely available to download from the CHE website www.york.ac.uk/inst/che/pubs. Access to downloaded material is provided on the understanding that it is intended for personal use. Copies of downloaded papers may be distributed to third-parties subject to the proviso that the CHE publication source is properly acknowledged and that such distribution is not subject to any payment.

Printed copies are available on request at a charge of £5.00 per copy. Please contact the CHE Publications Office, email che-pub@york.ac.uk, telephone 01904 321458 for further details.

Centre for Health Economics
Alcuin College
University of York
York, UK
www.york.ac.uk/inst/che

Abstract

Appropriate decisions based on cost-effectiveness evaluations of health care technologies depend upon the cost-effectiveness threshold and its rate of growth as well as some social rate of time preference for health. The concept of the cost-effectiveness threshold, social rate of time preference for consumption and social opportunity cost of capital are briefly explored before the question of how a social rate of time preference for health might be established is addressed. A more traditional approach to this problem is outlined before a social decision making approach is developed which demonstrates that social time preference for health is revealed through the budget allocations made by a socially legitimate higher authority. The relationship between the social time preference rate for health, the growth rate of the cost-effectiveness threshold and the rate at which the higher authority can borrow or invest is then examined. We establish that the social time preference rate for health is implied by the budget allocation and the health production functions in each period. As such, the social time preference rate for health depends not on the social time preference rate for consumption or growth in the consumption value of health but on growth in the cost-effectiveness threshold and the rate at which the higher authority can save or borrow between periods. The implications for discounting and the policies of bodies such as NICE are then discussed.

JEL Classification: I18, H43

Keywords: Economic evaluation. Discounting. Cost-effectiveness analysis

Introduction

Economic evaluation of health technologies, specifically cost-effectiveness analysis (CEA), is commonly seen as a means of satisfying an explicit social objective subject to an exogenous budget constraint (Gold *et al.,* 1996; Drummond *et al.,* 2005). Within this ‘social decision making perspective’, 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 within the health care sector rather than prescribing social choice in general. It is this role that CEA has tended to play in policy and it fits well with the view (Claxton *et al.,* 2007; Claxton *et al.*, forthcoming) that decision making bodies such as NICE in the UK can be seen as the agent of a socially legitimate higher authority which is unable to express an explicit, complete and coherent social welfare function. These circumstances can be regarded as a hierarchy (Mookherjee, 2006) where the agent (NICE) acts as a delegated authority but one that cannot be asked to improve social welfare, since it cannot be specified by the principal. Rather, the principal allocates resources and gives the agent a responsibility to pursue explicit and specific objectives (such as maximising the present value of health). The implications of this process reveal a partial but socially legitimate expression of some unknown underlying latent social welfare function. For example, the budget allocated by the higher authority implies a particular cost-effectiveness threshold which is a revealed expression of how much society wishes to pay for improvements in health generated by collectively funded health care.

In these circumstances, appropriate decisions made by the agent based on cost-effectiveness evaluations of health care technologies depend upon the cost-effectiveness threshold, k_t, in each period t, the growth rate of the cost-effectiveness threshold, g_k, (such that $k_{t+1} = k_t \; 1+g_k$), and some social rate of time preference for health, r_h (Claxton *et al.*, in submission).[1] For a technology with costs and effects in two time periods, a decision rule expressed in terms of discounted net health benefit is to accept the technology if:

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

where Δh_t and Δc_t represent the incremental health effects and incremental costs in period t of accepting the technology, both discounted at a common rate of r_h. Alternatively, the decision rule can be rearranged and expressed as a comparison of the technology’s incremental cost effectiveness ratio (ICER) to the current period cost-effectiveness threshold, k_1. The technology should be accepted if (see Claxton *et al.*, in submission)[2]:

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

The discount rate applied to Δh_t remains:

$$d_h = r_h\,, \qquad (3)$$

[1] Such decisions also necessarily assume divisibility and constant returns (Birch and Gafni, 1993).

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

but the discount rate applied to Δc_t becomes:

$$d_c \approx r_h + g_k, \quad (4)$$

because some adjustment of the discount rate is needed to reflect any growth in k_t, e.g. if $g_k > 0$ ($g_k < 0$) then future costs are less (more) important because they displace less (more) health and this can be reflected in a higher (lower) discount rate. Therefore, a common discount rate of r_h is required if the cost-effectiveness threshold is believed to be constant over the period where costs differ. However, if $g_k \neq 0$ and decisions are based on an ICER decision rule, differential discounting of costs and health effects is required (Claxton *et al*., in submission).

Irrespective of whether the decision maker adopts a decision rule based on ICERs or on discounted net health benefit, appropriate estimates of Δh_t, Δc_t, r_h and k_t are required to ensure that decisions are made in line with objective delegated to the agent by the higher authority (which for the purposes of this paper is taken to be to maximise the present value of health). While the modelling of health care technologies to obtain estimates of Δh_t and Δc_t is now well-established, there is far less understanding of: i) what the cost-effectiveness threshold, k_t, represents and how this threshold might change over time; ii) what rate of time preference for health, r_h, ought to be used by the agent in making decisions within the health sector; iii) how r_h might be related to the budget allocations made by the higher authority and the resulting values of k_t; and iv) how r_h is related to the social rate of time preference for consumption, r_c, and the social opportunity cost of capital, r_s.

This paper is presented as follows. The concepts of the cost-effectiveness threshold, social rate of time preference for consumption and social opportunity cost of capital are briefly explored before a discussion of how a rate of time preference for health might be established. A more traditional approach to this problem is outlined before a social decision making approach is developed which demonstrates that social time preference for health is revealed through the budget allocations made by a socially legitimate higher authority. The relationships between r_h, k_t, g_k and r_c or r_s are then examined. The implications for discounting and the policies of bodies such as NICE are then discussed.

The cost-effectiveness threshold

The cost effectiveness threshold represents an estimate of health forgone as other health care activities are displaced within a budget constrained health care system to accommodate any additional costs of a technology approved by the agent (McCabe *et al.*, 2008). A national body, such as NICE, needs an estimate of what is expected to be forgone across the health care system in each period. This will change as circumstances are expected to change, tending to rise with increases in budget and health care costs but tending to fall with increases in productivity (Culyer *et al.*, 2007).

In each period the state of technology and the efficiency of the health care system as well as the prevailing prices of inputs can be regarded as fixed. In these circumstances it is reasonable to assume that health output, H_t, is a positive function of the health system budget in that period, B_t, and that this function is concave so that incremental increases in the budget result in diminishing marginal increases in health output:

$$H_t = F_t(B_t), \text{ where } \frac{dF_t(B_t)}{dB_t} > 0 \text{ and } \frac{d^2F_t(B_t)}{dB_t^{\,2}} < 0.$$

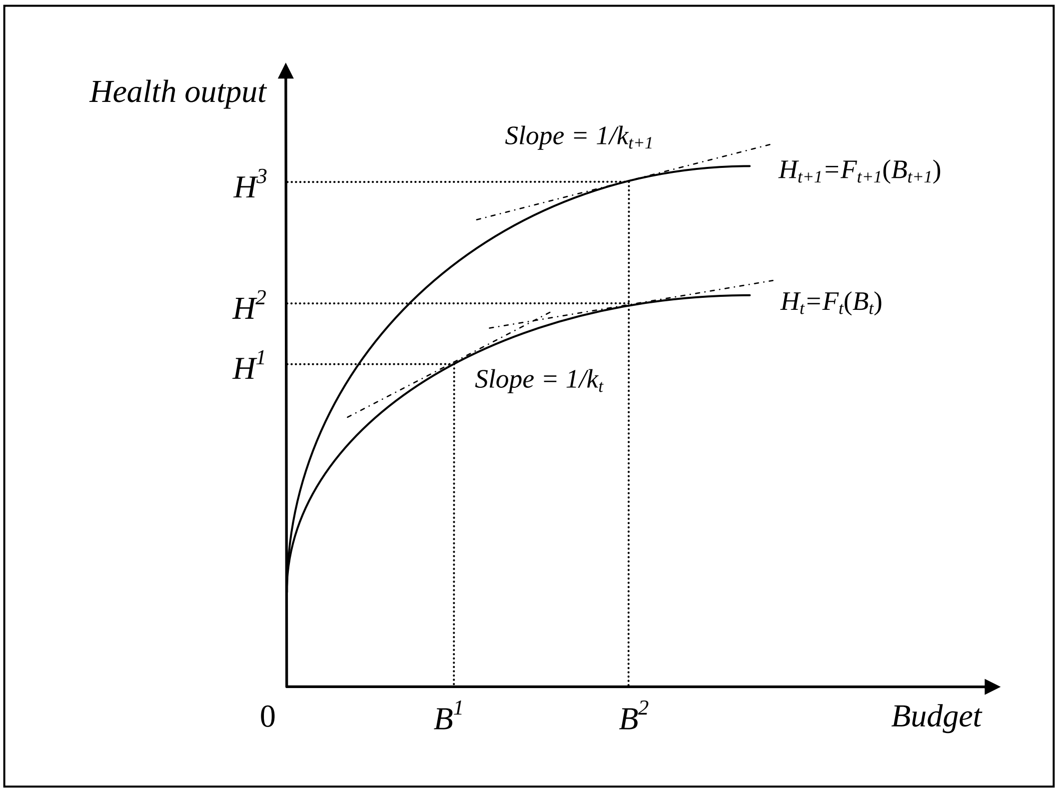

Figure 1. Health production and the threshold

This relationship between budget and health output in two periods is illustrated in Figure 1. The threshold, k_t, is the reciprocal of the slope at the point on the production function corresponding to the budget.[3] Since the slope of the production function represents the 'shadow price' of the budget (health gained for a marginal increase in budget), the threshold is the reciprocal of this shadow price. In period t the prices of inputs and technology are fixed so k_t will unambiguously increase with the budget, i.e. k_t is greater at *B*^{2} than at *B*^{1}. However, whether or not k_t increases over time will depend on changes in prices, technology as well as the budget. In Figure 1, $H_{t+1} = F_{t+1}\ B_{t+1}$ is drawn to represent a health care system that is more productive in the next period, possibly due to innovation in health technology and medicine, greater efficiency in the delivery of health care services and/or a fall in the prices of particular inputs, e.g. due to the entry of generic drugs. In these circumstances, with a constant budget of *B*^{2}, the threshold will fall ($k_{t+1} < k_t$). As such, the threshold will not necessarily grow with increases in the budget and only if the effect of budget growth is greater than that of improvements in productivity. Figure 1 illustrates this situation where the effect on the threshold of an increase in budget from *B*^{1} in t to *B*^{2} in $t+1$ more than offsets that of the growth in productivity (so $k_{t+1} > k_t$); nevertheless, g_k is much smaller than if prices and technology had remained unchanged.

In making an assessment of whether k_t is likely to increase with the budget it is also necessary to consider whether the agent has full and equal discretion over all types of health care expenditure. For

[3] Or, equivalently, k_t is the derivative of the budget with respect to health output: $k_t = \dfrac{dB_t}{dF_t(B_t)} = \dfrac{dB_t}{dH_t}$.

The growth rate of the threshold is given by $g_k = \dfrac{k_{t+1}}{k_t} - 1 = \dfrac{dF_t(B_t)}{dF_{t+1}(B_{t+1})}\dfrac{dB_{t+1}}{dB_t} - 1$.

example, if any growth in the overall budget is spent on national initiatives or other activities that cannot or cannot easily be displaced by the agent's decisions, then any additional costs of approved technologies must be accommodated by displacing other more effective activities elsewhere. If none of the additional budget is spent on activities that can be displaced then the threshold will remain constant (or fall with increased productivity). Therefore, it is growth in expenditure on more 'discretionary' parts of the budget and changes in the productivity and input prices of those health care activities which more likely to be displaced which are most relevant.

The social time preference rate

The social rate of time preference for consumption, r_c, often referred to simply as the social time preference rate (STPR), represents the proportionate increase in consumption required in period $t+1$ for society in period t to be indifferent between period t and period $t+1$ consumption. It is subject to an extensive literature (Ramsey, 1928; HM Treasury, 2003; Zhuang *et al.,* 2007), and is generally regarded as comprising of three elements: a catastrophic risk premium; a rate of pure time preference; and a third element which accounts for the diminishing marginal utility of future consumption when per capita consumption is expected to increase over time. Therefore, STPR represents social preferences over current and future consumption but the particular value of r_c is also linked to production possibilities and the observed rates of return to capital.

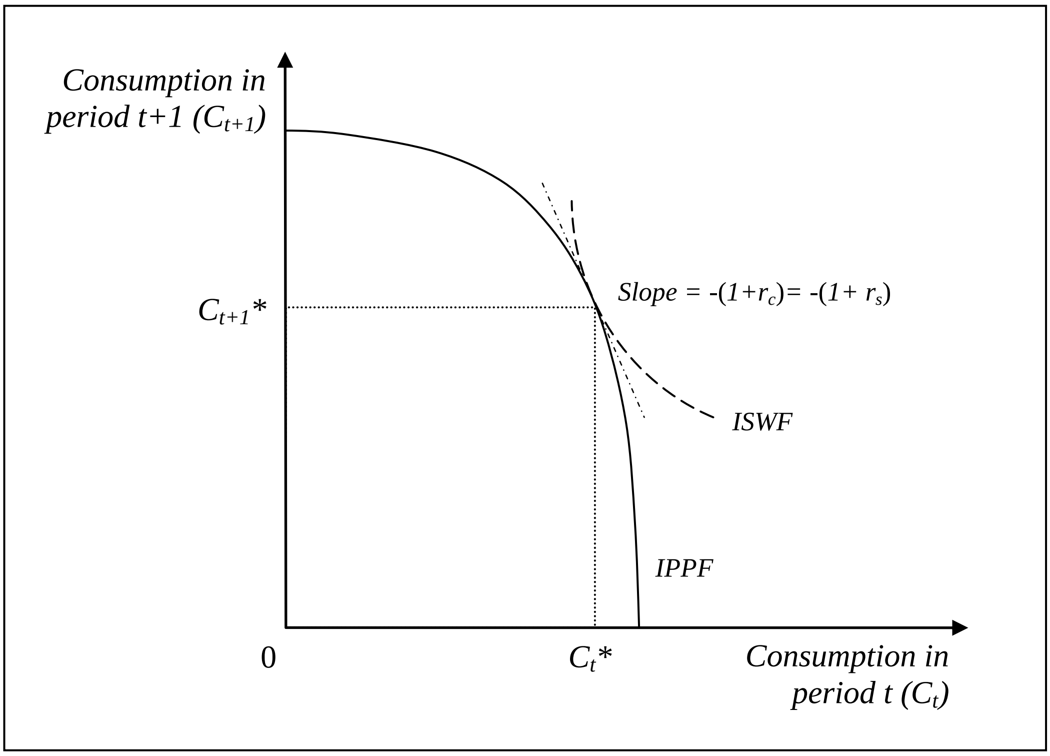

Figure 2. Intertemporal choice and production

The social time preference for consumption and the possibilities of transforming current consumption into future consumption through investment returns are simply illustrated over two periods in Figure 2. Social preferences for consumption between the two periods can be described by the dashed indifference curve which represents an intertemporal social welfare function (ISWF). The slope at any point on the ISWF is given by $-\ 1+r_c$ and describes the rate at which society is willing to trade current for future consumption. However, which point on the ISWF society will chose to locate depends on the intertemporal production possibilities, i.e. how current consumption can be transformed into future consumption by forgoing current consumption which can be invested at a rate of return to provide future consumption opportunities. These production possibilities are described by the intertemporal production possibility frontier (IPPF). The shape of the IPPF is determined by the marginal productivity of capital, which in this case exhibits markedly diminishing returns. The slope at any point on the IPPF is given by $-\ 1+r_s$, where r_s represents the marginal rate of return on capital. In the absence of distortions it also represents the social opportunity cost of capital, i.e. the social value of what is forgone elsewhere by government using capital (borrowing) for other purposes (Pearce and Nash, 1981; Gold *et al.,* 1996).

If society chooses to consume at point C_t*, C_{t+1}* this presumably is a point on the highest attainable ISWF which must be tangent to the IPPF. At this point the slope of the ISWF and IPPF are equal and STPR must be equal to the marginal rate of return (so $r_c = r_s$). Therefore, STPR can be estimated directly from the sum of its elements or inferred from observed rates of return (Gold *et al.,* 1996). For example, the UK Treasury has stipulated 3.5 per cent as the "standard real discount rate" for public sector investment appraisal (HM Treasury, 2003) based on the sum of the three elements: a pure time preference rate of 0.5%, a catastrophic risk premium of 1%, and 2% to represent the combined effect of the elasticity of the marginal utility of consumption and the growth in per capita consumption. Alternatively, STPR could be inferred by observing r_s, the rate at which the higher authority (government) can save or borrow. If markets are complete and undistorted, and the higher authority that allocates budget is regarded as a safe investment, then $r_s = r_c$, which would be equivalent to the real yield on a bond issued by the higher authority in period t which matures in period $t+1$. However, where distortions exist, for example uncorrected negative externalities or corporate taxation, then observed market rates will tend to be higher than r_s (Pearce and Nash, 1981). Similarly, where there are effects on future generations capital markets may be incomplete and observed rates will tend to be higher than the social opportunity cost of capital (Dasgupta *et al.*, 1999). For the purposes of the remaining discussion we denote r_s as the rate at which the higher authority that allocates resources is able to borrow or invest and initially assume that markets are complete and undistorted so that r_s also represents the social opportunity cost of capital ($r_s = r_c$).

The rate of time preference for health

A welfarist prescription

Traditionally economic analysis is more ambitious than the social decision making perspective described earlier, claiming to make statements about social welfare and prescribing social choice rather than simply informing decisions made by an agent with devolved and narrowly defined responsibilities (Mishan, 1967; Boadway and Bruce, 1984). This requires a particular view of social welfare, commonly resting on individual preferences revealed through choices that individuals make (especially in markets) or modified by specification of an explicit social welfare function which will have consumption as well as health as its arguments. Since both consumption and health enter the social welfare function it provides a clear link between the social time preference rate for consumption, r_c, and the social time preference rate for health, r_h (Gravelle and Smith, 2001).

If consumption and health are either the only social arguments or they are separable from all other arguments of social value then decisions which maximise the consumption value of health will also maximise social welfare (Gravelle *et al.*, 2007). A decision based on net health benefits in (1) can be extended to express the consumption value of net health gains. The technology should be accepted if:

$$v_1\Delta h_1 + \frac{v_2\Delta h_2}{(1+r_c)} > \frac{v_1\Delta c_1}{k_1} + \frac{v_2\Delta c_2}{k_2(1+r_c)}\ , \qquad (5)$$

where the consumption value of health in period t (v_t) is the amount of consumption in period *t* that is equivalent to 1 unit of health in period *t* and Δh_t and Δc_t are both discounted at a common rate of r_c.

This can also be expressed as a comparison of the ICER to the current cost effectiveness threshold (Claxton *et al.*, in submission). The technology should be accepted if[4]:

$$\frac{\Delta c_1 + \dfrac{\Delta c_2}{(1+r_c+g_k-g_v)}}{\Delta h_1 + \dfrac{\Delta h_2}{(1+r_c-g_v)}} < k_1, \qquad (6)$$

where g_v is the growth rate in v_t such that $v_{t+1} = v_t(1+g_v)$. The discount rate applied to Δh_t becomes:

$$d_h \approx r_c - g_v, \qquad (7)$$

because some adjustment of the discount rate for health is needed to reflect growth in v_t, e.g. if $g_v > 0$ then future health is more valuable in terms of consumption and this can be reflected in a lower discount rate. The discount rate applied to Δc_t becomes:

$$d_c \approx r_c - g_v + g_k = d_h + g_k, \qquad (8)$$

reflecting both growth in the consumption value of future health forgone and any changes in the rate at which future health will be forgone (g_k). Therefore, the traditional approach to economic analysis provides a clear prescription for the social rate of time preference for health, and its relationship to the rate of time preference for consumption:

$$d_h = r_h \approx r_c - g_v. \qquad (9)$$

However, the clear prescription that $r_h \approx r_c - g_v$ comes at some considerable cost. Firstly, it requires specification of v_t and g_v. Even if a particular consumption value of health could be agreed, the social welfare function it presupposes is unlikely to be complete and capture everything of social value. Nor is it likely to be carrying some broad consensus or social legitimacy, particularly when considering decisions with direct health impacts. In this case it implies that health and consumption are the only arguments of social value, or that they are separable from other arguments (e.g. education, equity, social solidarity, etc.) in some more complete description of social welfare. Adopting an incomplete description will lead to prescriptions which conflict with other legitimate objectives of social policy and other social arguments which – although difficult to formalise – may be important (e.g. widely held notions of social justice, intertemporal health equity, etc.). For example, observing $g_v > 0$ would suggest that greater social weight should be given to health gains for future patients, even though they are expected to enjoy more health, for no other reason than they are expected to have higher incomes and consume more than current patients. This might be regarded as a socially unacceptable conflict with equity concerns and widely held notions that access to health care should not be based on income but on capacity to benefit.

Furthermore, under this prescription, the budget allocation decisions made by a socially legitimate higher authority have no normative significance whatsoever. Indeed, it is v_t alone which expresses social value. The budget constraint and the implied value of k_t is not a legitimate expression of social value. Rather, the higher authority's budget allocation is simply a nuisance, an inefficiency which prevents the maximisation of (this particular definition of) social welfare - unless by chance

[4] Using the plausible assumptions 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 $(1+r_c).(1+g_k)/(1+g_v) \approx (1+r_c+g_k-g_v)$

$v_t = k_t$ and the budget is regarded as 'optimal' with respect to the presupposed welfare function. In short, it requires the imposition of a particular social welfare function, which will not be universally acceptable or even likely to carry some broad social consensus (Arrow, 1950; Sen, 1970), but nevertheless can stand above legitimate social democratic processes.

A social decision making approach

In contrast, the more modest social decision making perspective described earlier regards the value k_t as a revealed partial expression of the value society places on health generated by collectively funded health care. Similarly, a partial but legitimate expression of the social time preference rate for health (generated by collectively funded health care) is also revealed by the allocation of budgets over time. A simple two period model is used to graphically illustrate this in Figure 3. It demonstrates that r_h is implied by the budget allocation and the health production functions in each period. As such, r_h depends not on r_c and g_v, but on growth in the cost-effectiveness threshold, g_k, and the rate at which the higher authority can save or borrow between periods, r_s, which will be equal to r_c if markets are complete and undistorted.

Budget allocation

The higher authority's choice in allocating total resources between two periods is illustrated in bold in the south west quadrant of Figure 3. The total resources of B^* available at the start of period *t* must be fully allocated between the two periods, so for each possible choice of budget in period *t*, B_t, there is a unique budget in period $t+1$, $B_{t+1} = (B^* - B_t).(1+r_s)$. Therefore, the budget constraint depends on the rate of return on those resources saved in period *t* to provide budget in $t+1$, i.e. r_s is equal to $B^{**} - B^* \big/ B^*$ in Figure 3.

Production functions for health

The relationship between budget allocated in each period and health output in each period is represented by the health production functions in the south east and north west quadrants. In each period the state of technology and input prices can be regarded as fixed, however, just as in Figure 1, H_{t+1} is drawn to represent a health care system that is more productive in the next period, e.g. due to innovation in medicine. H_{t+1} is also drawn conditional on budget allocated in period *t*, i.e. health care provided in period *t* with costs that fall on B_t may also generate health improvements in the next period which is reflected in the positive intercept of H_{t+1}. As in Figure 1, the slope of each production function at the corresponding budget represents the shadow price of the budget constraint or the reciprocal of the cost-effectiveness threshold, k_t or k_{t+1}. Therefore, in making budget allocation choices between periods, the higher authority determines health output in both periods and in turn the respective thresholds.

Intertemporal health production possibility frontier

Each unique budget allocation described in the south west quadrant is, through the production functions described in the south east and north west quadrants, associated with a combination of health outputs in each period in the north east quadrant. The combinations of H_t and H_{t+1} which result from the possible choices of B_t and B_{t+1} describe the intertemporal health production possibilities frontier (IHPPF). The slope of the IHPPF is given by $\dfrac{dH_{t+1}}{dH_t} = -(1+r_h)$, which represents the rate at which current health output can be transformed into future health through budget reallocation. The health production possibilities are determined by the productivity of the health care system in each period. The choice for the higher authority is where to locate on the IHPPF by allocating health budgets over each period.

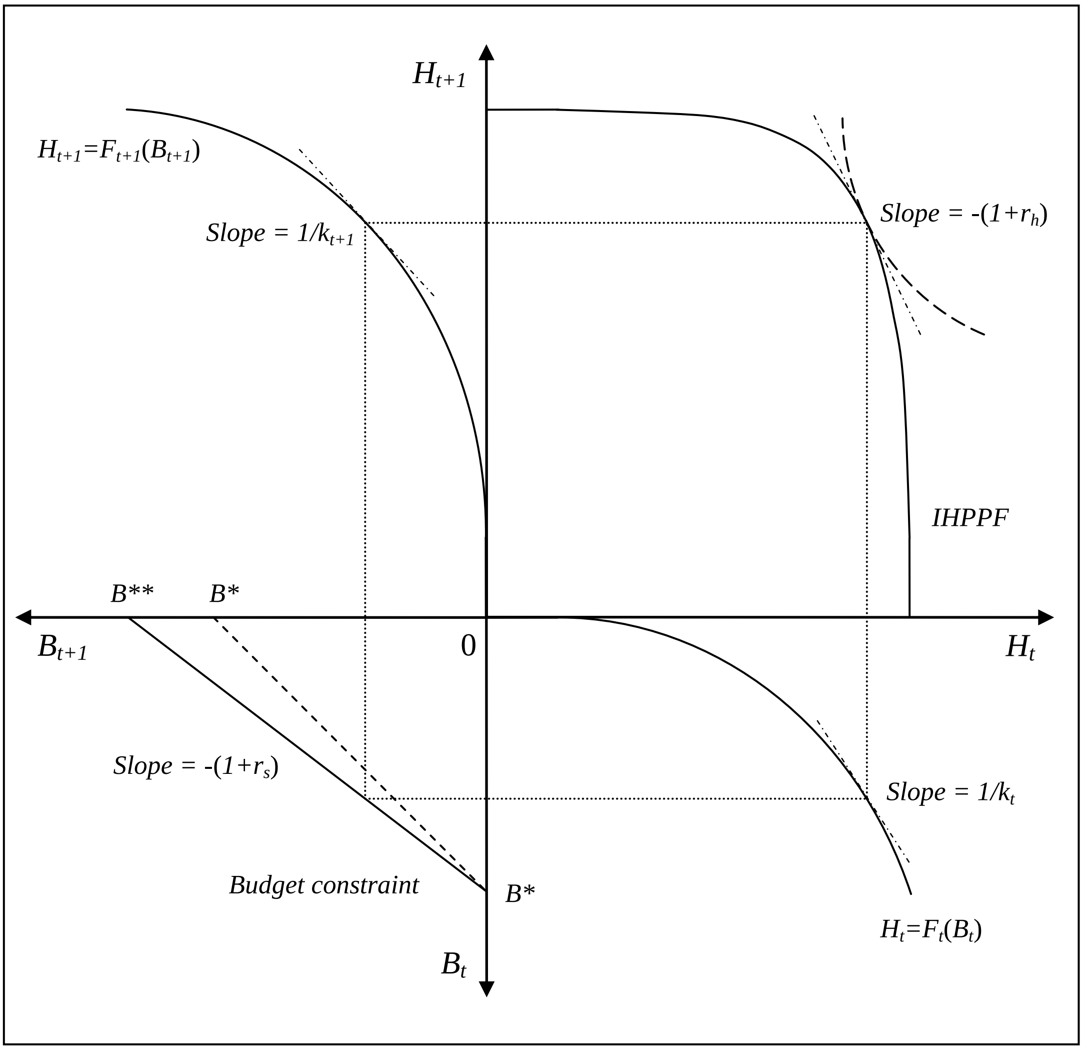

Figure 3. Budget allocation and time preference for health

Revealed social time preference for health

By allocating budget between periods the higher authority in effect chooses a point on the IHPPF. It is reasonable to suppose that this point is regarded as socially preferred to other possible points that could have been chosen by allocating the budget in a different way. Therefore, the slope of the IHPPF at this point not only represents how current health can be transformed into future health but also reveals the rate at which the higher authority is willing to trade current for future health, i.e. an expression of social time preference for health. This expression is only partial in the sense that a complete intertemporal social welfare function cannot be specified and there will be many other arguments of social value at play when budget allocations are made. Therefore, although the dotted line in the north east quadrant suggests an indifference curve with common properties, it remains unknown. The things that can be reasonably known are B_t, B_{t+1}, the associated estimates of k_t and k_{t+1} and therefore r_h at the chosen point on the IHPPF.

Deriving r_h from r_s and g_k

The relationship between the revealed social rate of time preference for health, the rate at which the higher authority can borrow and invest and the growth in the cost effectiveness threshold can be established. For example, consider an increase in the period t budget of ΔB_t. This will result in an increase in health output in period t of $\frac{\Delta B_t}{k_t}$. Given that the total budget for the two periods is constrained, this increase in the period t budget is financed by borrowing against the period $t+1$ budget at an interest rate of r_s, resulting in a decrease in the period $t+1$ budget of $\Delta B_t.(1+r_s)$. This in turn results in a decrease in period $t+1$ health output of $\frac{\Delta B_t.(1+r_s)}{k_{t+1}}$, which can be discounted to represent the decrease in terms of period t health as $\frac{\Delta B_t.(1+r_s)}{k_{t+1}.(1+r_h)}$.

For the budgets chosen by the higher authority to be consistent with its objective of maximising the present value of health, there must be no discounted net health benefit to choosing an alternative pair of budgets. Where ΔB_t is marginal, the resulting increase in period t health output is therefore equivalent to the resulting decrease in period $t+1$ health output represented in terms of period t health: $\frac{\Delta B_t}{k_t} = \frac{\Delta B_t.(1+r_s)}{k_{t+1}.(1+r_h)}$.

This rearranges to $\frac{k_{t+1}}{k_t} = \frac{1+r_s}{1+r_h}$ and, since $\frac{k_{t+1}}{k_t} = 1+g_k$, this rearranges to:

$$r_h = \frac{1+r_s}{1+g_k} - 1. \tag{10}$$

Where r_s and g_k are small, this simplifies to:

$$r_h \approx r_s - g_k \tag{11}$$

If markets are undistorted, so $r_s = r_c$, then $r_h \approx r_c - g_k$, which is very similar to the traditional welfarist prescription in (9). The important difference is that it is growth in the social value of health revealed by budget allocation that matters rather than growth in the consumption value of health derived from some external social welfare function.

Thus the higher authority's rate of time preference for health, r_h, may be derived from the interest rate the higher authority faces, r_s, and the growth rate of the cost-effectiveness threshold, g_k. Since the threshold in each period (and, in turn, the rate of growth of the threshold, g_k) is determined by the budgets set by the higher authority, the setting of these budgets implies the higher authority's preferred value of r_h. This derivation of r_h is intuitive. Where $g_k = 0$, r_h is simply the rate at which the higher authority can borrow or save, r_s. If r_s increases, it becomes relatively more desirable to decrease B_t and increase B_{t+1} so as to take advantage of the higher interest rate; for a particular set of budgets to remain consistent with the higher authority's objective therefore requires a higher rate of time preference for health, r_h. Meanwhile, any increase in g_k makes it relatively more desirable to increase B_t at the expense of B_{t+1}, since relatively less health is forgone due to a marginal decrease in the budget in period $t+1$ (due to the relatively higher threshold) than is gained due to the marginal increase in the budget in period t. As such, for a particular set of budgets to remain consistent with the higher authority's objective requires a lower rate of time preference for health, r_h.

Appropriate discount rates in CEA

The discount rates to apply when cost-effectiveness is expressed as an ICER can be based on information which is already assumed to be available. The discount rate applied to Δh_t from (3) becomes:

$$d_h \approx r_s - g_k \,, \qquad (12)$$

and the discount rate applied to Δc_t from (4) becomes[5]:

$$d_c = r_s \,. \qquad (13)$$

It should be noted that where $g_k = 0$, $d_c = d_h$, so differential discounting of costs and health effects is only appropriate when there is expected to be growth in the cost-effectiveness threshold. If the UK Treasury rate of 3.5% is regarded as an appropriate estimate of the STPR and if markets are regarded as complete and undistorted (so $r_s = r_c$) then the current NICE policy of discounting costs and health effects at 3.5% will be appropriate if the cost-effectiveness threshold is expected to be constant over the period where there are incremental costs and health benefits associated with the technology in question.

Discussion

A social decision making perspective regards bodies such as NICE as the agents of a socially legitimate higher authority that allocates budget to the health care system. Estimates of Δh_t, Δc_t, and k_t are required by the agent to ensure that decisions are made in line with the objective delegated by the higher authority. In addition, an instruction from the higher authority as to the rate at which it can borrow or invest when allocating resources, r_s, is also needed. However, the social rate of time preference for health is not a choice that the agent can or should make. Rather, just like the cost-effectiveness threshold, it is revealed through budget allocation decisions. Therefore, not only is k_t a revealed (albeit partial) expression of social value, but $r_h \approx r_s - g_k$ is also a revealed but partial expression of the social rate of time preference for health generated by collectively funded health care. The relationship between r_h derived in this way and r_c depends on whether there are no distortions in capital and product markets and on whether the cost-effectiveness threshold is constant. If both these conditions hold then $r_s = r_c$, $g_k = 0$ and the current NICE policy of discounting costs and health effects at 3.5% would be correct if the current UK Treasury estimate of the STPR is deemed appropriate. However, there are good reasons to suppose that this rate is too high. Aside from the difficulty of basing STPR on dubious estimates of each of its 3 elements, a real rate of return of 3.5% far exceeds observed marginal rates of return - even before any adjustments are made for distortions (e.g. environmental externalities or intergenerational effects), which would be required to estimate a social opportunity cost of capital. Importantly, from a social decision making perspective, a real rate of 3.5% is far in excess of current real yields on UK government bonds - as of 30 September

---

[5] Since $r_h = \dfrac{1+r_s}{1+g_k} - 1$, substituting this into the formula for d_c from Claxton *et al.* (in submission):

$$d_c = r_h + g_k + r_h . g_k = \left[\frac{1+r_s}{1+g_k} - 1\right] + g_k + \left[\frac{1+r_s}{1+g_k} - 1\right] . g_k = \left[\frac{1+r_s}{1+g_k} - 1\right] . (1+g_k) + g_k$$

$$= \left[\frac{1+r_s - 1 - g_k}{1+g_k}\right] . (1+g_k) + g_k = 1 + r_s - 1 - g_k + g_k = r_s .$$

2009, real rates of 1.24%, 1.31% and 0.19% are expected at 5, 10 and 20 years respectively (Bank of England, 2009). A lower common rate of discount might therefore be a more appropriate policy for NICE to adopt if the threshold is not expected to grow over relevant time horizons.

If the threshold is expected to grow at the same rate as some consumption value of health based on individual preference or derived from an external social welfare function ($g_k = g_v$) then, whether a welfarist prescription or social decision making perspective is adopted, the same discounting policy should be used as long as $r_s = r_c$. If it is assumed that the higher authority always allocates budget to maximise the welfare function that any chosen definition of v_t presupposes then budgets would necessarily be 'optimal' with respect to this definition of social welfare, and $v_t = k_t$ and $g_v = g_k$. However, this assumes that an explicit and complete welfare function can be specified which both describes social choice and how it is mediated through the social democratic process, i.e. precisely what the social decision making approach assumes is not possible. Alternatively, it is possible to simply define $g_v = g_k$ by assuming that the only expression of the value of health is the cost-effectiveness threshold – but now v_t and g_v are redundant because it is k_t and g_k that express social value. In these circumstances it is not possible to appeal to evidence of $g_v > 0$ based on individual preferences or the analysis of particular welfare functions to justify $g_k > 0$. There appear to be only two coherent positions to take: either assume budgets will match a chosen welfare function, in which case evidence of v_t and g_v from the chosen welfare function would be relevant; or assume that budgets are an expression of some unknown, latent welfare function, in which case it is evidence of k_t and g_k that is relevant.

Whether or not the threshold has been growing or is expected to grow over relevant time horizons is, in principle at least, an empirical question (Martin *et al.*, 2008). Simply observing real growth in the budget for health care, however, is not sufficient evidence for $g_k > 0$. It will also depend on improvements in the productivity of health care and whether any increased expenditure is discretionary and 'displaceable' by NICE guidance. Over recent years much of the real budget growth in the UK NHS has been devoted to national initiatives that are not easily displaced, e.g. new contracts for General Practitioners and consultants, national waiting time targets, information technology initiatives, etc. It also includes the guidance issued by NICE itself, which is mandatory. Therefore, any real growth in what remains will have been much more modest and more likely to be offset by growth in the productivity of displaceable activities, e.g. drugs, devices, procedures and other services. Similarly, although there has been a general rise in input prices for the UK NHS, much of this inflation has been driven by staff as well as capital and overhead costs, a great deal of which cannot easily be displaced. What are more relevant are the prices of inputs which could be displaced, an important element of which is drug prices. Although branded drug prices have tended to rise, at the same time there has been generic entry on patent expiry with dramatic reductions in prices (Office of Fair Trading, 2007). As such, it is not self evident that the threshold has grown over recent years, despite real increases in the budget for health care; in any case, growth in the threshold seems much less likely in the future with the prospect of reduced budget growth, increased pressures to improve productivity and downward pressure on input prices.

Under a social decision making perspective, the allocation of budget illustrated in Figure 3 is not, and cannot, be assumed to be 'optimal'. To make claims about the optimality or otherwise of budgets poses the question: 'optimal' with respect to what? The 'what' can only be some complete (or at least separable) description of social welfare. However, the premise of a social decision making perspective is that a complete, explicit and legitimate expression of social welfare is not possible. If that premise is acceptable (it quite reasonably may not be) then all else seems to follow. All that can be said is that the budgets are allocated by a legitimate social process which is tasked with balancing ever-changing competing and contradictory claims on resources and conflicting social objectives. Therefore, the budgets allocated by this process, including allocations between sectors other than health, will not generally appear 'optimal' when compared with any specific social welfare function that might be specified. The implications within the health sector (the thresholds and the implied social rate of time preference for health) are revealed and legitimate, but can only be regarded as a partial expression of social value. In summary, it appears that both the welfarist and social decision making perspectives are internally consistent. What distinguishes them is the assumption - on which discounting and other policy questions turn - of whether a complete, explicit and legitimate expression of social welfare is possible.

References

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

Asenso-Boadi, F., Peter, T.J., Coast J. 2008. Exploring differences in empirical time preference rates for health: an application of meta-regression. *Health Economics* **17**: 235–248.

Bank of England. 2009. UK Instantaneous Implied Real Forward Curve. Hhttp://www.bankofengland.co.uk/statistics/yieldcurve/index.htm [1H October 2009]

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.

Boadway RW, Bruce N. 1984. *Welfare Economics*. Blackwell: Oxford.

Brouwer W, Culyer AJ, Job N, van Exel A, Rutten FH. 2008. Welfarism vs. extra-welfarism. *Journal of Health Economics* **27**: 325–338.

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

Claxton K, Walker S, Palmer S and Sculpher, M. Forthcoming. Appropriate perspectives for health care decisions. A report for the UK Department of Health. *Forthcoming CHE Research Report, November 2009*. Centre for Health Economics, University of York: York

Claxton K, Paulden M, Gravelle H, Bouwer W, Culyer AJ. In submission. Discounting and decision making in the economic evaluation of health care technologies. *Health Economics*, in submission.

Culyer AJ, 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 of Health and Clinical Excellence. *Journal of Health Services Research and Policy* **12**: 56-58.

Dasgupta P, Maler KG, Barrett S. 1999. Intergenerational equity, social discount rates and global warming. In P.R. Portney and J.P. Weyant, eds. *Discounting and Intergenerational Equity*. Resources for the Future: Washington DC.

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.

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

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

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.

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

Martin S, Rice N, Smith P. 2008. Does health care spending improve health outcomes? Evidence from English programme budgeting data. *Journal of Health Economics* **27**: 826-842.

McCabe C, Culyer AJ, Claxton K. 2008. The NICE cost-effectiveness threshold: what it is and what it means. Pharmacoeconomics **9**: 733-744.

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

Mookherjee D. 2006. Decentralisation, hierarchies and incentives. A mechanism design perspective. *Journal of Economic Literature* **44**: 367-390.

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

Office of Fair Trading. 2007. *The pharmaceutical price regulation scheme. An OFT market study*. Office of Fair Trading: London. Hwww.oft.gov.uk/shared_oft/reports/comp_policy/oft885.pdf [1H October 2009].

Pearce DW and Nash CA. 1981. *The Social Appraisal of Projects: A Text in Cost-Benefit Analysis*. MacMillan: London.

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**: 152-57.

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.