Whose costs and benefits? Why economic evaluations should simulate both prevalent and all future incident patient cohorts

Martin Hoyle, PhD *Research Fellow*
Rob Anderson, PhD *Senior Lecturer*

Peninsula Technology Assessment Group (PenTAG),
Peninsula Medical School,
Universities of Exeter and Plymouth, Exeter.

Correspondence and reprints to:
Dr M Hoyle
PenTAG
Institute for Health Services Research
Noy Scott House
Royal Devon & Exeter Hospital
Barrack Road
Exeter EX2 5DW
United Kingdom

Tel: +44 1392 406969
Fax: +44 1392 406401
email: martin.hoyle@pms.ac.uk

Financial support: The work reported in this paper was unfunded.

Running head: Whose costs and benefits?

NB: An earlier version of this manuscript was discussed at the UK Health Economics Study Group winter meeting in January 2008.

Word count: (excl Abstract, references, legends) = 5,749

Abstract

Background Most health technology economic evaluations simulate only the prevalent cohort, or the next incident cohort of patients. They therefore do not capture all future patient-related benefits and costs.

Objective We show how to estimate and aggregate the ICERs for both currently eligible (prevalent) and future (incident) patient cohorts, within the same model-based analysis. We show why, and in what circumstances, the prevalent and incident cohort ICERs are likely to differ.

Methods Algebraic expressions were developed to capture all components of the ICER in hypothetical cohorts of all prevalent patients and future incident patients. Numerical examples are used to illustrate the approach.

Results The ICER for the first (i.e. next) incident cohort is equivalent to the ICER for all future incident cohorts only when the discount rates for costs and benefits are the same; otherwise, when the discount rate for benefits is lower than for costs, the ICER for all future incident cohorts is lower than the ICER for the first incident cohort. Separate simulation of prevalent and incident patients treated for a hypothetical progressive chronic disease shows widely different ICERs according to which patient cohorts were included when the discount rates were equal.

Conclusions In many circumstances, both the prevalent cohort and all future incident cohorts should be modelled. The need for this approach will depend on the likely difference in the ICERs for prevalent and incident patients, the relative size of the two types of cohort, and whether costs and benefits are discounted at equal rates.

Key words: cost-effectiveness analysis, ICER, decision modelling, chronic disease, technology assessment.

Introduction

It is increasingly recognised that to inform decision-making at a regional or national level, incremental cost-effectiveness ratios (ICERs) need to be based on rigorously informed decision model-based analyses which compare the incremental costs and effects of all relevant comparators, and typically for the remainder of patients' lifetimes.[1] Also, to be consistent with the fundamental tenets of cost-benefit analysis, such models should enable the valuation of costs and benefits "in each year of the project" (p.4), or for the whole of a health technology's life.[2]

The prevalence and incidence of a disease are fundamental concepts in epidemiology. The prevalence is the number of cases in a population at a specified point in time, and the incidence is the number of new cases arising in a given period in a population.[3] We apply the equivalent concepts of the prevalent cohort and future incident cohorts to model-based cost-effectiveness analysis. We define the prevalent cohort as those patients eligible for the new technology at the time the technology is first introduced. Any given patient will be eligible from the time when the technology is first clinically appropriate (e.g. just diagnosed with multiple sclerosis and eligible for drug treatment, or when first eligible for a hip replacement) until the time when the new technology is no longer appropriate (e.g. patient dies, or the disease has reached such a severe state that the drug is no longer effective, or the patient is too old to receive a hip replacement). Next, we define the incident cohort starting *t* years in the future (i.e. *t* years after the date of a technology's introduction) as comprising those patients who first become eligible for the new technology (e.g. diagnosed) *t* years in the future.

Cost-effectiveness studies generally model either only the first incident cohort of patients or only the prevalent cohort. We argue first that model-based economic evaluations of new treatments should model the costs and benefits of all patients in the prevalent cohort and in *all* future incident cohorts over the life of the technology. We further recommend that overall cost-effectiveness should be based on all these cohorts combined, i.e. that the ICER be calculated from a weighted sum of all these costs and benefits.

The current ISPOR guidance on good practice in decision analytic modelling focuses mainly on the structure of the model, the validation of the model estimates/inputs, and the choice between alternative simulation models (e.g. Monte Carlo vs. cohort).[4] However, aside from some general encouragement to stratify models by patient sub-groups, there is no specific advice on what starting populations should go into a decision model. Nor does methods guidance from national health technology assessment agencies state what current and future populations of patients should be included in model-based analyses, e.g. UK,[5] Australia,[6] New Zealand,[7] Canada,[8] Germany.[9]

In this paper, we describe the mathematics for estimating the ICER that includes the costs and benefits for both the prevalent and all future incident cohorts. For simplicity, we consider a new technology versus a single comparator technology, but the ‘comparator technology’ could represent no treatment. Equivalent equations for more than two comparators in a net monetary benefit framework are given in the Online Appendix. The technologies can be either a drug, a medical device, or a screening program. We suggest parameters related to the structure of the patient cohorts that could be included in the probabilistic sensitivity analysis.

ICER for incident cohorts

First future incident cohort

Consider a cost-effectiveness model where future costs and benefits are modelled at discrete times (e.g. a Markov model). Suppose the incremental costs, per patient starting treatment, between the new and comparator technology (where the comparator technology could be no technology, i.e. best supportive care), in cycles 0, 1, 2,…., H are ΔK_0, ΔK_1, ΔK_2, …., ΔK_H and incremental benefits ΔB_0, ΔB_1, ΔB_2,…., ΔB_H (Table 1). The time horizon is H cycles. For clarity, given that the ΔK_j and ΔB_j are expressed per patient starting treatment, these quantities tend to zero with cycle j, as patients die. Then the ICER as currently calculated for health technology assessments for the first future incident cohort, given discount rate for costs of r^*_C and benefits r^*_B over a cycle;

$$\text{ICER (first future incident cohort)} = \frac{\sum_{j=0}^{H} {v^*_C}^j \Delta K_j}{\sum_{j=0}^{H} {v^*_B}^j \Delta B_j}$$

where we define $v^*_C = \dfrac{1}{1+r^*_C}$, $v^*_B = \dfrac{1}{1+r^*_B}$.

All future incident cohorts

Now assume, more realistically, that a new cohort of patients will become *eligible* for treatment with the new or comparator technologies at the start of each of T years in the future. The new and comparator technologies are assumed to become obsolete after T years, possibly replaced by another technology. In this paper, we present all analyses with closed-form algebra to aid understanding of the methods. However, it is of course possible to simulate each future incident cohort. In general, assume that the number of eligible patients at the start of each cohort, relative to the number of eligible patients at the start of the first year, is given by n_t, at year t, so that $n_0 = 1$. The n_t are commonly used in budget impact analyses. The n_t could increase with year t, for example to model increasing numbers of Type 2 diabetes patients in the future as obesity becomes more common. Assume further that the probability that an eligible patient is given the new technology in the t^{th} year in the future is p_t. The p_t could be described as the “rate of adoption”, “rate of uptake” or “market penetration” of the new technology, and are also commonly used in budget impact analyses. The graph of the volume of sales of a drug, i.e. the product $n_t p_t$, against year t is generally $\cap$ -shaped.[10] The annual volume of a drug sold typically increases in the first decade after drug launch, reflecting the diffusion of the new drug after launch. The annual volume of a drug sold in the second decade after launch reflects post-patent experience and declines as patients switch to newer drugs.[10;11] Then, the relative number of patients in the incident cohort starting t years in the future affected by the new technology is $n_t p_t$. By analogy with the special case of two future incident cohorts (see Online Appendix);

$$\text{ICER (all incident cohorts)} = \frac{\sum_{t=0}^{T} n_t p_t v_C^{\,t}}{\sum_{t=0}^{T} n_t p_t v_B^{\,t}} \text{ ICER (first incident cohort)} \quad \text{(Equation 1)}$$

where $v_C = \dfrac{1}{1+r_C}$ and $v_B = \dfrac{1}{1+r_B}$, and r_C and r_B are the “inter-generation” annual discount rates for costs and benefits between the current time and the time of the future incident cohorts. By contrast, r^*_C and r^*_B are the (per cycle) “intra-generation” discount rates. We further assume that undiscounted incremental costs and benefits are the same for all incident cohorts.

From Equation 1, the ICERs for all future incident cohorts combined and for the first future incident cohort are equal if the cost and benefit discount rates, r_C and r_B, are equal. Alternatively, if $r_C > r_B$, the ICER for all incident cohorts is lower (see Online Appendix). For example, in the Netherlands, where costs are discounted at 4% and benefits at 1.5% per year,[12] under certain assumptions, the ICER for all future incident cohorts combined may be about ¾ of the ICER assuming a single incident cohort (i.e. as calculated in the traditional way) (see Online Appendix).

When $n_t p_t$ is equal for all t, Equation 1 simplifies to;

$$\text{ICER (all incident cohorts)} = \left(\frac{1-v_B}{1-v_C}\right)\left(\frac{1-v_C^{\,T+1}}{1-v_B^{\,T+1}}\right) \text{ICER (first incident cohort)}$$

(Equation 2)

Now, if we assume that $n_t p_t$ follows a $\cap$ -shaped quadratic curve, as is often the case with drug sales volumes,[10] then Equation 2 is applicable again (see online Appendix). If independent estimates of $n_t p_t$ are available then they should be used in Equation 1, otherwise Equation 2 is appropriate. Equation 2 is convenient since we need only have an estimate for the single parameter T, not the $n_t p_t$ for all t.

ICER for prevalent cohort

In addition to the patients who will become eligible for the new technology in the future, there may be patients who are already eligible at the time the technology is introduced. Such prevalent patients would switch from the current to the new technology. Denoting the incremental costs and benefits of the prevalent cohort at cycle $j = 0 \ldots H$, expressed per patient at the start of the prevalent cohort, by ΔC_j, and ΔQ_j, the ICER for the prevalent cohort is;

$$\frac{\sum_{j=0}^{H} {v^*_C}^{j} \Delta C_j}{\sum_{j=0}^{H} {v^*_B}^{j} \Delta Q_j} \qquad \text{(Equation 3)}$$

ICER for incident and prevalent cohorts combined

We define N as the number of patients in the prevalent cohort that are eligible for treatment, relative to the number of patients in the first future incident cohort, and $\overline{p}$ as the probability that a patient in the eligible prevalent cohort is given the new technology, assumed constant over cycle j. Then in the general case of any number of treatments, the optimal strategy is to choose the treatment with the maximum expected net benefit[13] (see Online Appendix). Returning to the particular case of two treatments alternatives, we calculate the ICER as a “ratio of means”, in the terminology of Stinnett & Paltiel (1997).[14] In particular, the ICER equals total incremental costs divided by total incremental benefits during the whole time the technology is used:

ICER (prevalent and all future incident cohorts) =

$$\frac{\overline{p} N \sum_{j=0}^{H} {v^*_C}^{j} \Delta C_j + \left(\sum_{t=0}^{T} n_t p_t {v_C}^{t} \right) \left(\sum_{j=0}^{H} {v^*_C}^{j} \Delta K_j \right)}{\overline{p} N \sum_{j=0}^{H} {v^*_B}^{j} \Delta Q_j + \left(\sum_{t=0}^{T} n_t p_t {v_B}^{t} \right) \left(\sum_{j=0}^{H} {v^*_B}^{j} \Delta B_j \right)} \qquad \text{(Equation 4)}$$

In this equation we make the simplifying assumption that the proportion of patients in a given incident cohort that are given the new technology, p_t, does not change over cycle j. Note that if the cost and benefit discount rates are equal, then Equation 4 implies that the ICER for the prevalent and incident cohorts combined will lie between the ICER for the prevalent cohort alone and the ICER for the first future incident cohort alone.

We now introduce parameters to allow us to estimate N and $\overline{p}$. Denote the average age of patients at the start of any incident cohort as A (assumed constant over time). Suppose a patient is *eligible* for treatment with the new technology over an average period of M years, from age A to age $A+M$. To avoid confusion, note that parameter M relates to the age range of any given *patient*. It should not be confused with parameter T, which relates to the age (lifetime) of the *technology*. Costs directly associated with the technology occur during some, but not all the period of eligibility. For example, for patients in the incident cohort, the cost of a hip replacement occurs at the very start of the period of eligibility, whereas, the cost of a drug for a chronic condition might occur over the whole period of eligibility, M.

When M is small, e.g. treatments for acute infection, the costs and benefits of the incident and prevalent cohorts are similar, because the patients' initial parameters, such as the average age and average severity of condition are similar between the incident and prevalent cohorts (see below). Conversely, when M is large, for example, for long-term therapies for chronic conditions, the costs and benefits of the incident and prevalent cohorts can be substantially different for a variety of reasons. Hence the ICER for the prevalent cohort is similar to the ICER for the incident cohort for acute conditions, but can be very different for chronic conditions. On average, we would expect that patients in the prevalent cohort will be approximately half way through their treatment with the comparator technology. Correspondingly, we expect that patients at the start of an incident cohort (i.e. at the start of their treatment) to be treated for approximately twice the length of time as patients in the prevalent cohort.

If the number of patients in the prevalent cohort that are eligible for treatment, relative to the number of patients in the first future incident cohort, N, is known from the literature, then this value should be used. For example, the annual incidence of end-stage renal disease in the UK in 2003 was 5,517 patients, and the prevalence was 34,259,[15] which gives $N = 34{,}259 / 5{,}517 = 6.2$. Alternatively, we now describe how to estimate N. Denote the probability that a patient who is treated with the *comparator* technology survives from age A, at the start of an incident cohort, to age $A + t$ as $s(A, A+t)$. Such data are often available from cost-effectiveness models. Then;

$$N = n_{-1}s(A,A+1) + n_{-2}s(A,A+2) + n_{-3}s(A,A+3) + ... + n_{-M+1}s(A,A+M-1)$$

(Equation 5)

Hence when M is large, for conditions that require a long period of treatment, N is large, and when M is small, for conditions that require short-term treatment, for example acute infection, N is small.

We estimate $\bar{p}$ as the weighted average of the p_t, with the weights equal to the number of patients in the prevalent cohort t years in the future;

$$\bar{p} = \frac{\sum_{t=0}^{M-1}\sum_{i=0}^{M-1} p_t n_{-i} s(A, A+t+i)}{\sum_{t=0}^{M-1}\sum_{i=0}^{M-1} n_{-i} s(A, A+t+i)} \qquad \text{(Equation 6)}$$

where subscript -*i* refers to the incident cohort that started *i* years in the past. Now suppose the cost and benefit discount rates are equal, i.e. $v_C = v_B = v$. Then Equation 4 becomes;

ICER (prevalent and all future incident cohorts) =

$$\frac{\dfrac{\overline{p}N}{\left(\sum_{t=0}^{T} n_t p_t v^t\right)} \sum_{j=0}^{H} v^{*j} \Delta C_j + \left(\sum_{j=0}^{H} v^{*j} \Delta K_j\right)}{\dfrac{\overline{p}N}{\left(\sum_{t=0}^{T} n_t p_t v^t\right)} \sum_{j=0}^{H} v^{*j} \Delta Q_j + \left(\sum_{j=0}^{H} v^{*j} \Delta B_j\right)}$$

From which it is clear that the prevalent cohort is negligible when $\dfrac{\overline{p}N}{\left(\sum_{t=0}^{T} n_t p_t v^t\right)}$ is small. This is true when *T* is very large, or *M* is very small. We now consider three cases;

1: Parameters for both the incident and prevalent cohorts are known
2: Parameters for incident cohort only are known
3: Parameters for prevalent cohort only are known

Case 1: Parameters for incident and prevalent cohorts known

Suppose we know the model parameters for both the incident and prevalent cohorts from literature reviews of primary research. Then we can calculate ΔC_j, ΔQ_j, ΔK_j, and ΔB_j. We then calculate the ICER for the incident and prevalent cohorts combined from Equation 4, using an estimate of the p_t and hence $\overline{p}$ (as explained in the

Discussion). To calculate the ΔC_j, ΔQ_j, ΔK_j, and ΔB_j directly, we would need data from two types of clinical trial. One trial (or trial subgroup) with patients from an incident cohort, i.e. newly diagnosed, and another trial with patients from the prevalent cohort. This would be especially useful if patients respond differently to a new technology according to previous treatments received, for example, for corticosteroids for asthma.[16]

If the prevalent cohort is large relative to the incident cohort, the range of values of input parameters, such as patient age and disease severity, for patients in the prevalent cohort may be wide. In this case, it may be preferable to allow for such heterogeneity of input parameters in the cost-effectiveness model which is used to generate the ΔC_j and ΔQ_j for the prevalent cohort. For example, the model could be run for each of a range of patient ages, and the ΔC_j and ΔQ_j estimated as a weighted average of the incremental costs and benefits for each model run, with weightings proportional to the probability density function of each age (e.g. as in Dewilde & Anderson 2004).[17]

Case 2: Parameters for incident cohort only known

Suppose we know the parameter values for the incident cohort only, e.g. if the clinical trial(s) were based on incident cohorts of patients only. We now outline a method for estimating the incremental costs and benefits for the prevalent cohort, ΔC_j, ΔQ_j. As above, we then calculate the ICER for the incident and prevalent cohorts combined from Equation 4. In the Online Appendix, we describe an alternative method for estimating ΔC_j and ΔQ_j, where we estimate the parameter values that specify the characteristics of patients at the start of the prevalent cohort. Although this second method is simpler to implement than the first method, it is slightly less accurate because we assume no variability in the input parameters of the prevalent cohort.

Returning to the first method, suppose the costs in the incident cohort, expressed per patient at the start of the incident cohort, are K_j and K'_j at cycle j = 0…H for the new and comparator technologies respectively (Fig. 1). As above, we assume that these costs are the same across all incident cohorts. We cannot simply assume that the future costs with the new technology for the incident cohort that started in year t (i.e. in the past, so that t is negative), $K_{t,j}$, j cycles since the start of the incident cohort, are given by K_j, because this would assume (incorrectly) that patients had been treated with the new technology in the past. Instead, in the Online Appendix, we show how to estimate the $K_{t,j}$ by an algorithm, which can be coded as a macro. The prevalent cohort costs and benefits for the new technology at cycle j = 0…H are calculated as $C_j = \sum_{t=-(M-1)}^{-1} K_{t,j-t} n_t \Big/ N$ and $Q_j = \sum_{t=-(M-1)}^{-1} Q_{t,j-t} n_t \Big/ N$ and for the comparator technology as $C'_j = \sum_{t=-(M-1)}^{-1} K'_{j-t} n_t \Big/ N$ and $Q'_j = \sum_{t=-(M-1)}^{-1} Q'_{j-t} n_t \Big/ N$ (Fig. 1).

Case 3: Parameters for prevalent cohort only known

In the Online Appendix, we describe a method to estimate the incremental costs and benefits for the incident cohort, ΔK_j, ΔB_j, given that we know the parameter values, e.g. average age, for the prevalent cohort only. As above, once we estimate ΔK_j, ΔB_j, we calculate the ICER for the incident and prevalent cohorts combined from Equation 4.

Example of application

Here, we apply the methods described above to an example cost-effectiveness model of a new maintenance drug versus an existing comparator drug to treat a chronic progressive condition. Details of the model structure and results are given in the Online Appendix, however we provide a brief description here. We assume that the new drug will be used in the health system for the next $T = 30$ years, and that the probability that a patient eligible for treatment takes the new drug at time t, p_t, follows a $\cap$-shaped quadratic curve. The relative number of patients in the incident cohort, n_t, is assumed equal over time t. The new drug reduces the rate of disease progression. Non-drug costs increase and utilities decrease with increasing disease severity. The average age at diagnosis, i.e. at the start of an incident cohort, $A = 30$ years, and we assume a certain distribution across disease severity states for patients in the incident cohort. Patients were modelled from age 30 to death or age 100. This gives $M = 70$ years over which patients are eligible to be treated with the new drug.

We estimate that the prevalent cohort is $N = 47$ times the size of a single incident cohort (Equation 5), and the average age of patients in the prevalent cohort is approx. 56 years, compared to $A = 30$ years in the incident cohort. As expected, patients are at a more advanced stage of illness in the prevalent cohort compared to the incident cohort. The ICER for the first incident cohort was calculated as £25,000 per quality-adjusted life year (QALY). Given that the cost and benefit discount rates were assumed equal, the ICER for all future incident cohorts combined was also £25,000 / QALY. The total discounted costs and benefits for the prevalent cohort were calculated using the algorithm described in the Online Appendix (Fig. 2). The ICER for the prevalent cohort alone was substantially higher, at £94,000 / QALY, and for both the prevalent and all incident cohorts combined, £57,000 / QALY.

Discussion

In this paper we have argued that the cost-effectiveness of a treatment should be assessed in relation to all patients whose costs and benefits will be affected; both those currently eligible and those who will become eligible for the new treatment in the future. On average, patients in the prevalent cohort will be older and will typically be at a more advanced stage of disease than patients in the incident cohort. Furthermore, the more life-years over which the technology is applicable for patients (e.g. maintenance therapies for chronic conditions), the greater these differences. In summary, the suggestions in this paper are particularly important to implement in cost-effectiveness analysis in any of the following circumstances:

- for long-term therapies for chronic conditions (particularly for chronic progressive conditions), e.g. Alzheimer's disease, multiple sclerosis, cystic fibrosis, diabetes, eczema, rheumatoid arthritis.
- when the discount rates for costs and benefits differ.

In these cases, the ICER as calculated in this paper for all affected patients may differ substantially from the ICER as traditionally calculated (for the next incident cohort). In particular, we have described a simplified but realistic example cost-effectiveness analysis of a chronic progressive condition, assuming equal cost and benefit discount rates. In this example, the ICER as calculated by our method is 2.3 times the ICER as traditionally calculated by assuming just a single incident cohort, and 0.6 times the other traditional method of assuming a single prevalent cohort.

We have shown that when the discount rates for costs and benefits differ, it is particularly important to estimate the costs and effectiveness of all future incident cohorts. While many health economists,[18] and most country's official guidance for the cost-benefit analysis of health technologies,[12] recommend equal discount rates for costs and benefits the matter is by no means settled. Some suggest r_C should be greater than r_B.[18-20] In particular, Brouwer et al (2005)[19] recommend r_C = 3.5% and r_B = 1.5%, and Gravelle & Smith (2001)[18] suggest that r_C should be 2-5% greater than r_B. There remain some countries where different discount rates are recommended for health care economic evaluations (e.g. Netherlands: r_C = 4%, r_B = 1.5%; and Belgium: r_C = 3%, r_B = 1.5%; source, ISPOR website[12]).

An obvious question is: when the prevalent cohort is not negligible, when is the ICER for the prevalent cohort greater than the ICER for the first future incident cohort, and vice versa? We suggest an answer to this question for three types of conditions-with-treatments. First, we have shown that for the example cost-effectiveness model of a continuous treatment for a *progressive* chronic disease, the prevalent cohort ICER is substantially greater than the incident cohort ICER, because at each cycle, the ratio of incremental costs to incremental benefits is greater for the prevalent cohort (Fig. 2 online Appendix). This may be a typical result for a progressive chronic condition, supported by economic evaluations in cardiology.[21] Nevertheless, this question warrants further analysis, particularly since a contrary result has been found in a cost-effectiveness study of a cholesterol-lowering statin.[22] In this study, the incremental cost per life year gained was lower for older patients than for younger patients. The difference in the ICERs was due to higher incremental costs in the younger age group, but similar incremental life years gained. Whilst these two patient groups did not correspond to incident and prevalent cohorts, this result does suggest that the prevalent cohort ICER may, in some cases, be lower than the incident ICER, given that patients in the prevalent cohort are, on average, older than those in the incident cohort.

Second, we consider a continuous treatment for a non-progressive chronic condition, such as asthma. Suppose there are two health states A and B, and patients are in the worse state A (e.g. poorly controlled asthma) under the comparator drug and the better state B under the new drug. Suppose further that life expectancy is independent of the drug and that costs are a function just of the drug (higher for the new drug) and whether the patient is in state A (higher) or state B (lower). Further, suppose that patient utility is a function of just the state, and is higher in state B than in state A. In this case, the ratios of incremental costs and benefits $\frac{\Delta K_j}{\Delta B_j}$ and $\frac{\Delta C_j}{\Delta Q_j}$ are constant over cycle *j* and are the same for the incident and prevalent cohorts. Hence the prevalent cohort ICER equals the incident cohort ICER.

Third, consider the scenario where the majority of costs are incurred up front for chronic conditions. This is particularly appropriate for medical devices, such as cardiac pacemakers for heart conditions and cochlear implants for deafness. Again, suppose there are two health states A and B, and suppose that patients are in the worse state A under the comparator technology and in the better state B under the new technology. Again, suppose that life expectancy is independent of the technology. Suppose the cost of the technology, e.g. cost of cochlear implant itself plus cost of implantation surgery, is incurred in the first cycle, and is greater for the new than the old technology. Health state costs can be higher or lower in state A than in state B. Patient utility is again solely a determined by health state. In this case, for the incident and prevalent cohorts, the ratios of incremental costs and benefits $\frac{\Delta K_j}{\Delta B_j}$ and $\frac{\Delta C_j}{\Delta Q_j}$ are high in the first cycle, and far smaller in all future cycles. The ratios for the two cohorts are equal by cycle. However, given that patients are older in the prevalent than in the incident cohort, and will therefore use the technology for fewer years, in the prevalent cohort, there will be fewer cycles with low incremental cost/benefit ratios. Hence, the prevalent cohort ICER will be greater than the incident cohort ICER.

Another question is whether the ICER calculated according to our approach will be greater or smaller than the ICER as traditionally calculated. In general, it is not possible to say: some technologies will appear more cost-effective, and others less cost-effective. Consider first the case when the prevalent cohort is negligible compared to the incident cohort, for example with treatments for acute conditions. Then, if the cost and benefit discount rates are equal, the ICER will not change. Alternatively, if the discount rate for costs is greater than the rate for benefits, the ICER will be less than traditionally calculated. Now, assume that the prevalent cohort is not negligible. In previous model-based cost-effectiveness analyses, either;

1: all patient-related parameters (e.g. average age, average disability level) refer to the prevalent cohort, or

2: all patient-related parameters refer to the incident cohort, or

3: some parameters refer to the prevalent cohort and the rest to the incident cohort.

Again, assuming equal discount rates, in the expected scenario that the prevalent cohort ICER is greater than the incident cohort ICER, the combined ICER as calculated here would be lower than the ICER calculated in case 1, greater than in case 2, and uncertain in case 3. Conversely, in the less likely event that the prevalent cohort ICER is lower than the incident cohort ICER, then these conclusions are reversed. However, in a review of model-based cost-effectiveness analyses, we found very few studies that explicitly state whether model parameters were derived from incident or prevalent cohorts. Therefore, our analysis suggests that the ICER as calculated in previous cost-effectiveness analyses may be substantially different from the ICER as calculated according to the methods of this paper. As a side issue, note that we have assumed that the costs and benefits in all future incident cohorts are equal. This assumption would be violated if, for example, one component of the costs is predicted to increase in the future at a different rate to the other components of the costs. Then we must adjust Equations 1, 2 and 4 appropriately.

One disadvantage of our suggested methods is that they require estimation of additional model parameters. The following algorithm may allow the analyst to decide when it is necessary to implement our suggested methods. First, if the cost and benefit discount rates differ, our suggested method should be followed. Specifically, we must estimate the *relative* sizes of the affected patient populations (n_tp_t) for each year in the future up to year $t = T$ (Equation 1). If such data is not available, we suggest above that n_tp_t can be assumed a quadratic function of year t. We then require only an estimate of the lifetime of the new technology, T (Equation 2). Variability in n_tp_t and/or T should be incorporated in the probabilistic sensitivity analysis. The values of n_tp_t, T, and the variability in these quantities could be estimated by analysing trends in the volumes of sales of similar technologies in the past.

Next, what if the cost and benefit discount rates are equal? When the size of the prevalent cohort is negligible compared to the size of the incident cohort, then the ICER for the prevalent cohort and all future incident cohorts combined can be approximated by the ICER for the first future incident cohort alone. However, when the prevalent cohort is not small, the analyst should first compare the ICERs for the prevalent and incident cohorts. Given that the ICER for both types of cohort combined lies between the ICER for the prevalent cohort and the ICER for the first future incident cohort when the cost and benefit discount rates are equal (see analysis), if the two ICERs are similar, then the ICER for the prevalent cohort and all future incident cohorts combined can be approximated by the ICER for either the prevalent cohort or the ICER for the first future incident cohort. If the ICERs for the prevalent cohort and first future incident cohort are not similar, then our method for calculating a combined ICER should be used.

The proposed method requires estimates of n_t and p_t separately for each year in the future up to year $t = T$ in order to estimate $\bar{p}$ (Equation 6). However, without relevant data, it is reasonable to assume that the n_t are equal for all t. The p_t are then estimated as described in the estimation of $n_t p_t$ above. Next, we must estimate the size of the prevalent cohort relative to the size of the first future incident cohort, N, and patient-related parameters, such as the average age and average disability status for both the incident and prevalent cohort. Uncertainty in N should also be reflected in the probabilistic sensitivity analysis. Given that the ICER can be greatly altered by use of our proposed methods, the extra effort in estimating these parameters and in adjusting the cost-effectiveness analysis is justified. Nonetheless, we are mindful of the extra analytical effort and data requirements that are implied by our methods. We have therefore also provided some practical tools for estimating the costs and benefits for incident or prevalent patient cohorts when full data on the other type of cohort is unavailable. Ideally, however, cost-effectiveness analyses in these situations should be grounded in rigorous empirical studies which yield separate effectiveness estimates and other data from both incident, newly eligible, patients and those prevalent patients who are switching to the new treatment.

Given that the clinical and cost-effectiveness of a health technology can differ by patient subgroup, national guidance recommends assessing cost-effectiveness separately by patient subgroup (England,[5] Australia,[6] New Zealand,[7] Canada,[8] Germany[9]). The characteristics of patients in the subgroup should be identified on the basis of an *a priori* expectation of differential clinical or cost effectiveness due to known, biologically plausible mechanisms, social characteristics or other clearly justified factors.[5] Disease severity is an example of such an *a priori* factor. For example, consider a chronic progressive disease, with cost-effectiveness assessed for one mild disease subgroup and a severe disease subgroup. As already explained, patients are on average more severely ill in the prevalent cohort than in the incidence cohort. Therefore we might expect the proportion of patients in the severe disease subgroup that are in the prevalent cohort to be higher than the proportion of patients in the mild disease subgroup that are in the prevalent cohort. In the extreme case, the severe subgroup might represent only patients in the prevalent cohort, and the mild subgroup only patients in the incident cohort. In this case, the ICER for the severe subgroup would equal the prevalent cohort ICER (Equation 3), and the ICER for the mild subgroup would equal the ICER for all future incident cohorts combined (Equation 1). In this special case, the technology might be deemed cost-effective for patients in the incident cohorts, but cost-ineffective for patients in the prevalent cohort, or visa versa. Of course, cost-effectiveness is often assessed without splitting patients into subgroups according to disease severity. For example, in the NICE appraisal of natalizumab for multiple sclerosis, patients in all Expanded Disability Status Scale levels from 0 (mild) to 10 (death) were combined to calculate a single estimate of cost-effectiveness (NICE 2007).[23] In this case, the ICER should be estimated as in Equation 4.

We have already outlined two possible areas for future research: the general conditions under which the prevalent cohort ICER is greater than the incident cohort ICER, and vice versa; and the estimation of the sizes of future incident cohorts, and the product life-time of a given technology, and their variability by analysis of trends in the volumes of sales of similar technologies in the past. Now we suggest the following additional areas of research. First, we have shown that cost-effectiveness is influenced by our methods when applied to an example simplified model. Our methods could be applied to other existing cost-effectiveness models to explore their influence on cost-effectiveness. Second, cost-effectiveness for our example model was rather dependent on the specific method used to estimate the costs and benefits for the prevalent cohort. It would be interesting to investigate this for real cost-effectiveness models. Third, we have suggested how clinical effectiveness in our model may be parameterized from trial data. We encourage investigation of the availability of such clinical data for ‘real world’ models. Fourth, in the previous paragraph, we describe how the proportion of patients in a patient subgroup that are in the prevalent cohort may depend on the subgroup. We recommend investigating the extent to which patient subgroups differ in this respect in real decision problems. Finally, we have assumed that undiscounted incremental costs and benefits are the same for all incident cohorts. Whilst we suggest that this is a reasonable assumption without evidence to the contrary, we encourage investigation into how factors such as the future prices of the health technology[24], future changes in the median age at diagnosis, future changes in life expectancy and relative treatment effectiveness may influence this assumption.

At present, most economic evaluations of health technologies simulate only the first incident cohort. In this paper, we have argued that model-based economic evaluations should simulate the costs and benefits for *all* people who will be affected by a given health policy decision. In particular, we have (a) demonstrated how to calculate the incremental cost-effectiveness of new health technologies when including the costs and benefits associated with either the current prevalent cohort or the future incident cohorts of patients, or both types of cohort together, and (b), using these equations, we have described the circumstances under which the ‘combined cohorts ICER’ is likely to differ from the ICER for the next incident cohort of patients.

An Excel spreadsheet implementing the example cost-effectiveness model is available from the authors on request.

Acknowledgements

We are grateful for useful comments on our manuscript made at the UK Health Economics Study Group winter meeting in January 2008. We also thank Ken Stein and three reviewers of this manuscript for their helpful suggestions.

Reference List

(1) Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models - A suggested framework and example of application. Pharmacoeconomics 2000; 17(5):461-477.

(2) Layard R, Glaister S. Introduction. In: Layard R, Glaister S, editors. Cost-benefit analysis. 2nd edition ed. Cambridge, UK.: Cambridge University Press; 1994. 1-56.

(3) Beaglehole R, Bonita R, Kjellström T. Basic epidemiology. 1993. Geneva, World Health Organisation.

(4) Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C et al. Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ISPOR task force on good research practices-modeling studies. Value in Health 2003; 6(1):9-17.

(5) National Institute for Health and Clinical Excellence. Guide to the Methods of Technology Appraisal. 2008. London, NICE.

(6) PBAC. Guidelines for preparing submissions to the PBAC. Part II: Guidelines for preparing the main body of a major submission. Section D: Economic

evaluation for the main indication. 2008. Australia, Australian Government. Department of Health and Ageing. URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/pbacguidelines-index~pbacguidelines-part2~pbacguidelines-part2_d

(7) PHARMAC. Prescription for Pharmacoeconomic Analysis (PFPA). Methods for cost-utility analysis. 2007. New Zealand, Pharmaceutical Management Agency. URL: http://www.pharmac.govt.nz/EconomicAnalysis/pharmacoeconomics

(8) CADTH. Guidelines for the economic evaluation of health technologies. 2006. Ottawa, Canada., Canadian Agency for Drugs and Technologies in Health. URL: http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf

(9) IQWIG. Methods for Assessment of the Relation of Benefits to Costs in the German Statutory Health Care System. 2008. Germany, Institute for Quality and Efficiency in Health Care. URL: http://www.iqwig.de/download/08-01-24_Methods_of_the_Relation_of_Benefits_to_Costs_Version_1_0.pdf

(10) Danzon P, Kim J. The life cycle of pharmaceuticals: a cross-national perspective. 2002. Whitehall, UK., Office of Health Economics. URL: http://hc.wharton.upenn.edu/danzon/html/books_and_monographs.htm

(11) Salomon JA, Weinstein MC, Goldie SJ. Taking account of future technology in cost effectiveness analysis. British Medical Journal 2004; 329(7468):733-736.

(12) ISPOR. Country-specific guidelines. 2008. International Society for Pharmacoeconomics and Outcomes Research. URL: www.ispor.org/PEguidelines/COMP1.asp

(13) Briggs A, Sculpher M, Claxton K. Decision modelling for health economic evaluation. Oxford, UK.: Oxford University Press; 2006.

(14) Stinnett AA, Paltiel AD. Estimating CE ratios under second-order uncertainty: The mean ratio versus the ratio of means. Medical Decision Making 1997; 17(4):483-489.

(15) Ansell D, Feest T. UK Renal Registry Report 2004. 2005. Bristol, UK, UK Renal Registry.

(16) Shepherd J, Rogers G, Anderson R, Main C, Thompson-Coon J, Hartwell D et al. Systematic review and economic analysis of the comparative effectiveness of different inhaled corticosteroids and their usage with long-acting beta(2) agonists for the treatment of chronic asthma in adults and children aged 12 years and over. Health Technology Assessment 2008; 12(19):XI-+.

(17) Dewilde S, Anderson R. The cost-effectiveness of screening programs using single and multiple birth cohort simulations: A comparison using a model of cervical cancer. Medical Decision Making 2004; 24(5):486-492.

(18) Gravelle H, Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. Health Economics 2001; 10(7):587-599.

(19) Brouwer WBF, Niessen LW, Postma MJ, Rutten FFH. Need for differential discounting of costs and health effects in cost effectiveness analyses. British Medical Journal 2005; 331(7514):446-448.

(20) Gravelle H, Brouwer W, Niessen L, Postma M, Rutten F. Discounting in economic evaluations: Stepping forward towards optimal decision rules. Health Economics 2007; 16(3):307-317.

(21) van Hout BA, Goes ES, Grijseels EWM, van Ufford MAQ. Economic evaluation in the field of cardiology: Theory and practice. Progress in Cardiovascular Diseases 1999; 42(2):167-173.

(22) Tonkin AM, Eckermann S, White H, Friedlander D, Glasziou P, Magnus P et al. Cost-effectiveness of cholesterol-lowering therapy with pravastatin in patients with previous acute coronary syndromes aged 65 to 74 years compared with younger patients: Results from the LIPID study. American Heart Journal 2006; 151(6):1305-1312.

(23) NICE. Multiple sclerosis - natalizumab: Appraisal consultation document. 2007. National Institute for Health and Clinical Excellence, London, UK. URL: http://guidance.nice.org.uk/page.aspx?o=418937

(24) Hoyle M. Future drug prices and cost-effectiveness analyses. Pharmacoeconomics 2008; 26(7):589-602.

Table 1. Key parameters.

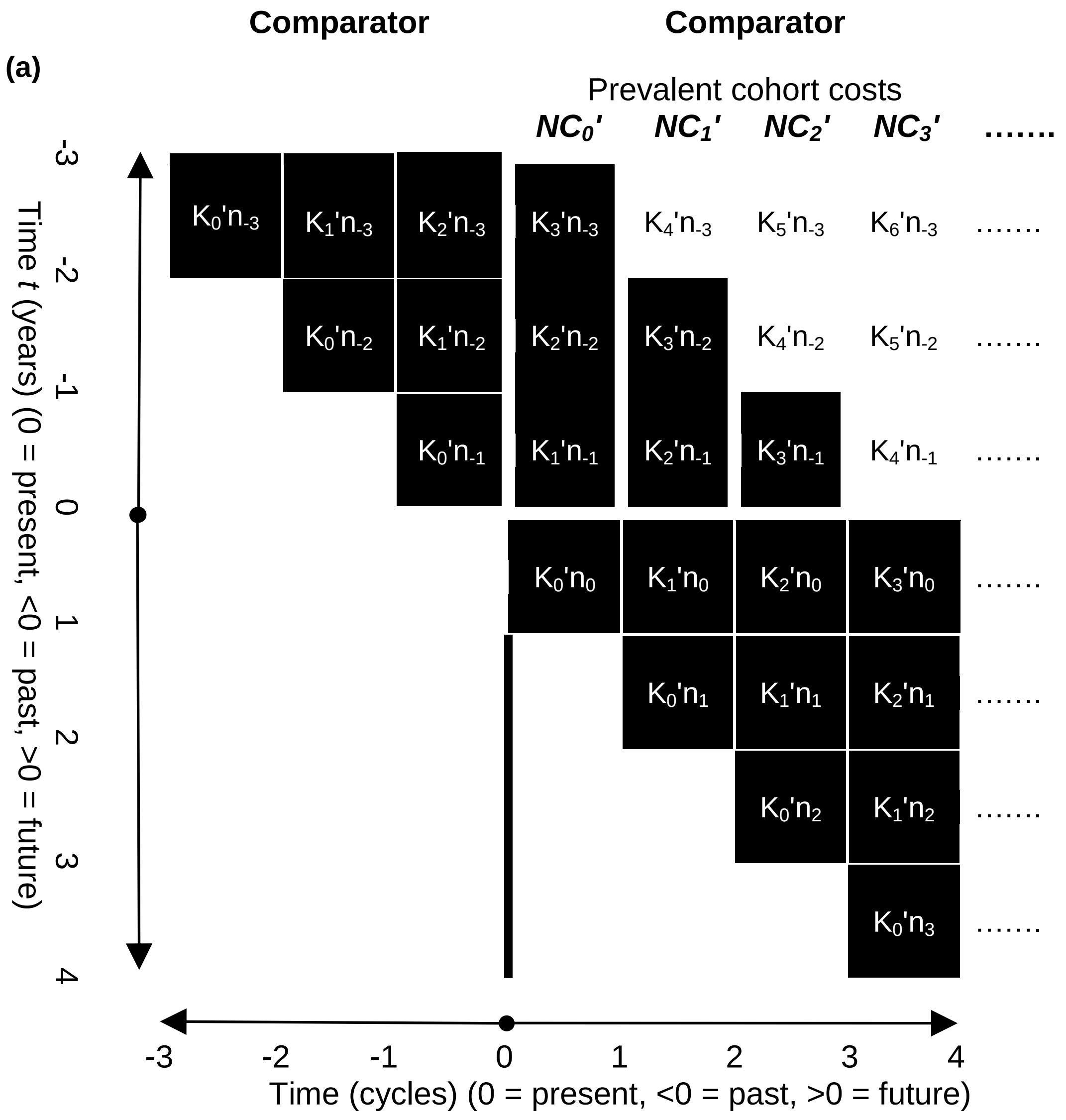
Parameter	**Definition**
ΔK_j, ΔB_j	incremental incident cohort cost and benefit between the new and comparator technology at cycle $j = 0, 1, 2, \ldots, H$, expressed per patient at the start of the incident cohort
ΔC_j, ΔQ_j	incremental prevalent cohort cost and benefit at cycle $j = 0, 1, 2, \ldots, H$, expressed per patient at the start of the prevalent cohort
K_j, K'_j	incident cohort cost per patient for the new and comparator technology at cycle $j = 0, 1, 2, \ldots, H$
C_j , Q_j	prevalent cohort cost and benefit per patient for the new technology at cycle $j = 0, 1, 2, \ldots, H$
C'_j , Q'_j	prevalent cohort cost and benefit per patient for the comparator technology at cycle $j = 0, 1, 2, \ldots, H$
$K_{t,j}$, $Q_{t,j}$	future costs and benefits per patient with the new technology for the incident cohort that started in year t (in the past, so that t is negative), j cycles since the start of the incident cohort
H	time horizon of each incident cohort in cycles
r_C , r_B	“inter-generation” annual cost and benefit discount rates
v_C , v_B	$= \frac{1}{1+r_C}$, $= \frac{1}{1+r_B}$
r^*_C , r^*_B	“intra-generation” cost and benefit discount rates over a cycle

v^*_C , v^*_B	$= \dfrac{1}{1+r^*_C}$, $= \dfrac{1}{1+r^*_B}$
T	expected lifetime of new technology in years
n_t	number of patients eligible for the new technology at the start of the incident cohort starting in year $t = -H, \ldots , -2, -1, 0, 1, 2, \ldots, T$, relative to the number of eligible patients at the start of the first year
p_t	probability an eligible patient is given the new technology $t = 0 \ldots T$ years in the future
$\overline{p}$	probability that a patient in the eligible prevalent cohort is given the new technology
N	number of patients in the prevalent cohort that are eligible for treatment, relative to the number of patients in the first future incident cohort
A	average age of patients at the start of the incident cohort
$s(A, A+t)$	probability a patient who is treated with the comparator technology survives from age A to $A + t$
M	number of years over which patients are eligible to be treated with the new technology

Figure 1. Prevalent and incident cohort costs for (a) the comparator and (b) the new technology. Incident cohorts are shown as separate rows. For simplicity, one cycle equals one year in this example. Here, the technology is applicable on average to a given patient for M = 4 years (4 black cells in each row), and the prevalent cohort comprises M - 1 = 3 incident cohorts. The future prevalent cohort comparator and new technology total costs at cycle j, NC_j' and NC_j equal the sum of the costs in the respective highlighted boxes. In (b), all costs before the assessment time (time zero) refer to the comparator technology, because the new technology was not used then. Costs directly associated with the technology occur in some, but not all the black cells. For simplicity, we display costs only four years into the future, whereas the expected technology lifetime, T, will probably be much longer.

Figure 2. Undiscounted costs (£) over time in the example cost-effectiveness model. (a) displays the per patient comparator drug costs showing separately all incident cohorts that started in the past. The costs in the future, i.e. to the right of the vertical line, comprise the costs of the prevalent cohort. For clarity, a single example incident cohort is displayed in bold. Costs initially rise as disease becomes more severe, thus incurring higher health state-related costs. Costs eventually fall to zero as patients die. (b) displays the same data for times in the past, but costs for the new drug in the future, i.e. for the new drug costs in the prevalent cohort. (c) displays comparator drug costs. In (c), the downward sloping line represents total costs in the prevalent cohort (summing over costs in all incident cohorts that started in the past), and the upward sloping line represents total costs in all future incident cohorts. To demonstrate scale, the incident cohorts that make up these quantities, some of which are shown in (a), are just visible at the bottom of the graph. We assume that there are the same number of patients in all incident cohorts.

Figure 1.


Comparator
Comparator
(a)
Prevalent cohort costs
NC_0' NC_1' NC_2' NC_3' …….
$K_0'n_{-3}$ $K_1'n_{-3}$ $K_2'n_{-3}$ $K_3'n_{-3}$ $K_4'n_{-3}$ $K_5'n_{-3}$ $K_6'n_{-3}$ …….
$K_0'n_{-2}$ $K_1'n_{-2}$ $K_2'n_{-2}$ $K_3'n_{-2}$ $K_4'n_{-2}$ $K_5'n_{-2}$ …….
$K_0'n_{-1}$ $K_1'n_{-1}$ $K_2'n_{-1}$ $K_3'n_{-1}$ $K_4'n_{-1}$ …….
$K_0'n_0$ $K_1'n_0$ $K_2'n_0$ $K_3'n_0$ …….
$K_0'n_1$ $K_1'n_1$ $K_2'n_1$ …….
$K_0'n_2$ $K_1'n_2$ …….
$K_0'n_3$ …….
-3 -2 -1 0 1 2 3 4
Time t (years) (0 = present, <0 = past, >0 = future)
-3 -2 -1 0 1 2 3 4
Time (cycles) (0 = present, <0 = past, >0 = future)

(b)

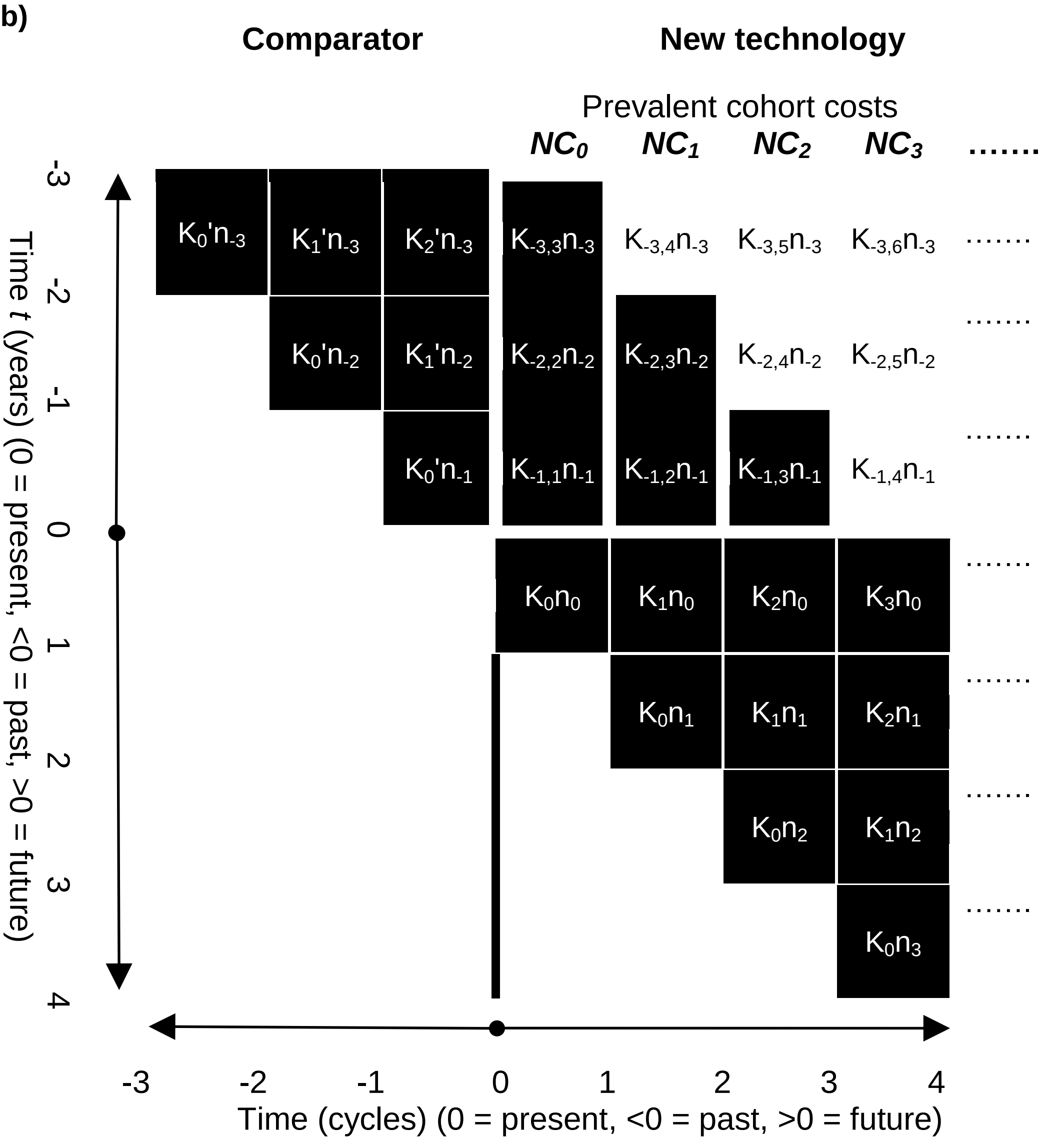
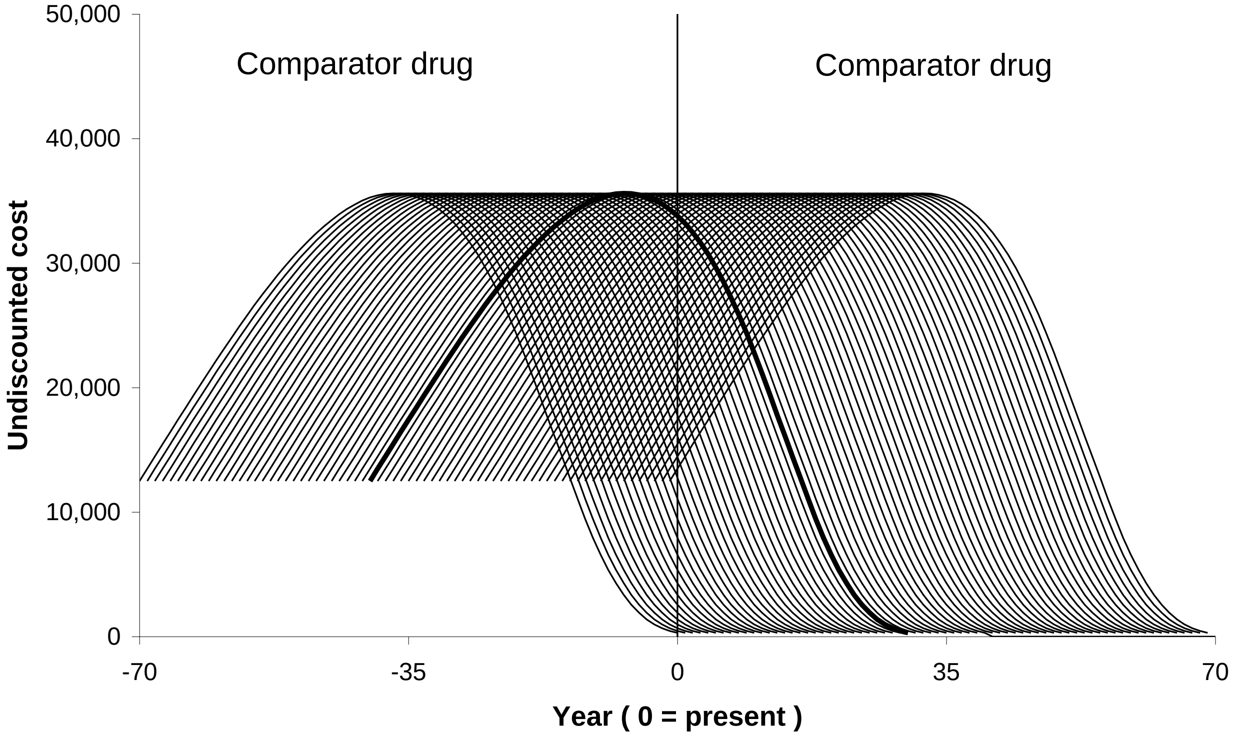
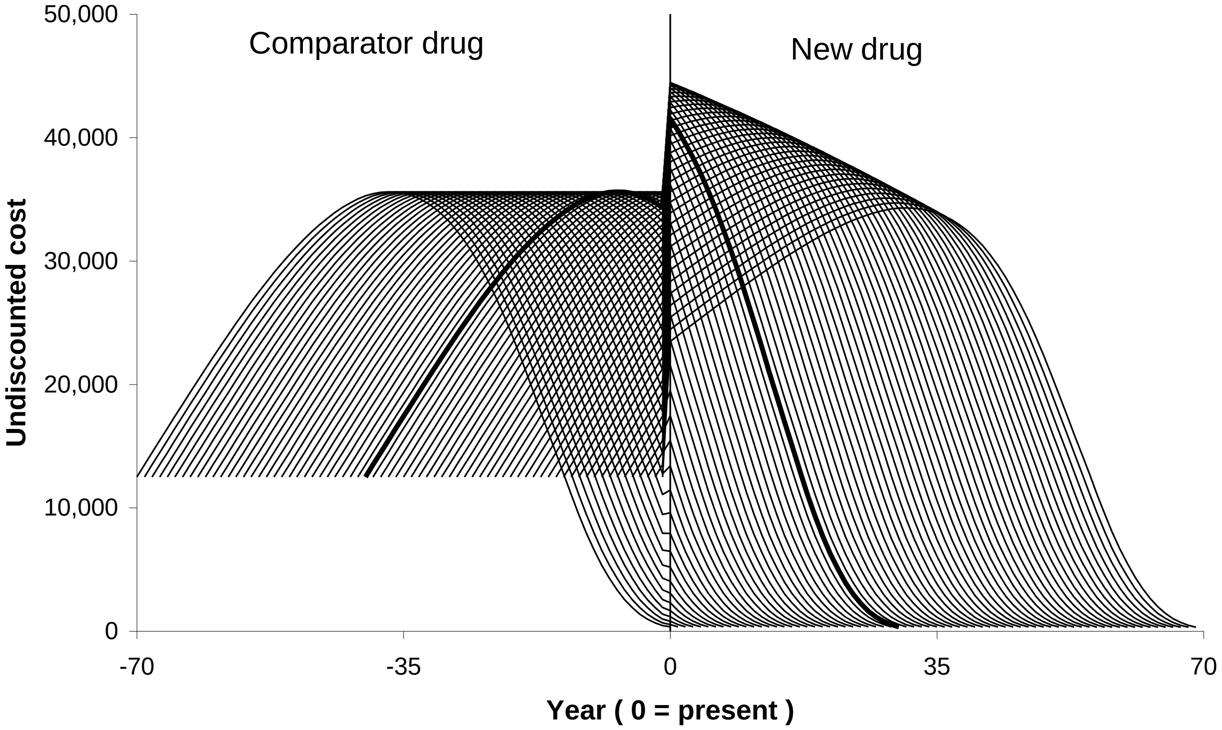

Comparator
New technology
Prevalent cohort costs
NC_0
NC_1
NC_2
NC_3
…….
$K_0'n_{-3}$
$K_1'n_{-3}$
$K_2'n_{-3}$
$K_0'n_{-2}$
$K_1'n_{-2}$
$K_0'n_{-1}$
$K_{-3,3}n_{-3}$
$K_{-3,4}n_{-3}$
$K_{-3,5}n_{-3}$
$K_{-3,6}n_{-3}$
$K_{-2,2}n_{-2}$
$K_{-2,3}n_{-2}$
$K_{-2,4}n_{-2}$
$K_{-2,5}n_{-2}$
$K_{-1,1}n_{-1}$
$K_{-1,2}n_{-1}$
$K_{-1,3}n_{-1}$
$K_{-1,4}n_{-1}$
K_0n_0
K_1n_0
K_2n_0
K_3n_0
K_0n_1
K_1n_1
K_2n_1
K_0n_2
K_1n_2
K_0n_3
Time t (years) (0 = present, <0 = past, >0 = future)
-3
-2
-1
0
1
2
3
4
Time (cycles) (0 = present, <0 = past, >0 = future)

Figure 2.

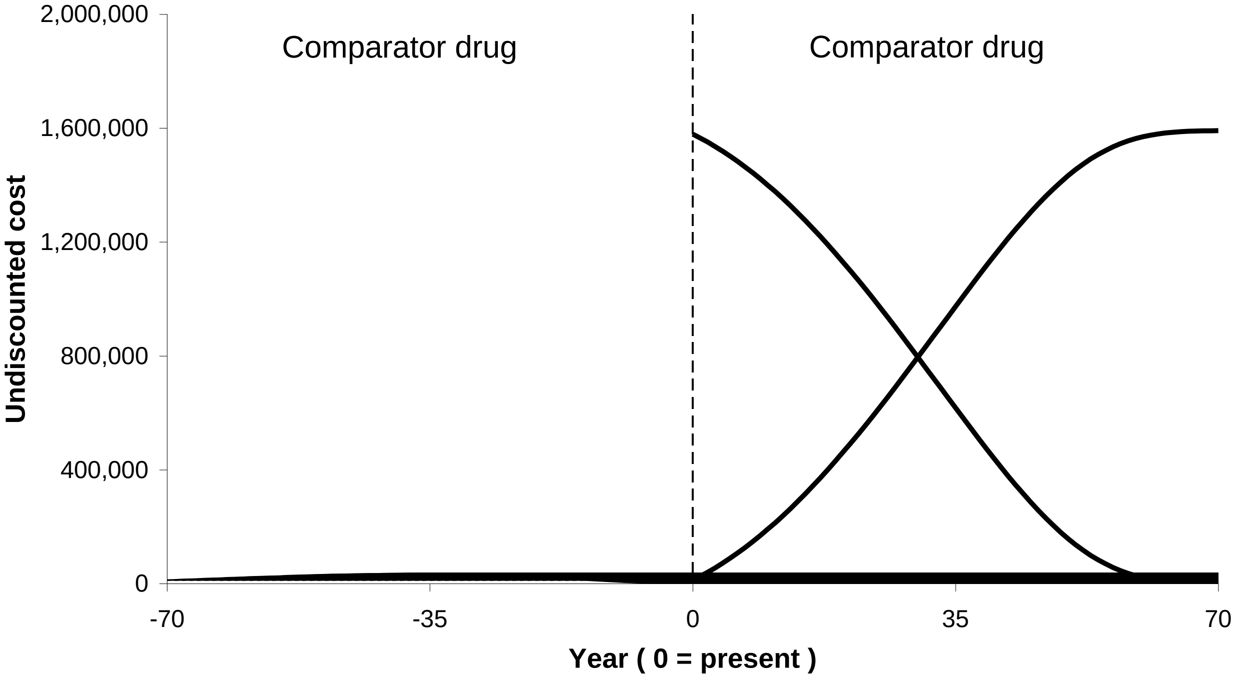
(a)


Comparator drug
Comparator drug
50,000
40,000
30,000
20,000
10,000
0
Undiscounted cost
-70
-35
0
35
70
Year (0 = present)

(b)


Undiscounted cost
50,000
40,000
30,000
20,000
10,000
0
Comparator drug
New drug
-70
-35
0
35
70
Year (0 = present)

(c)


Undiscounted cost
2,000,000
1,600,000
1,200,000
800,000
400,000
0
Comparator drug
Comparator drug
-70
-35
0
35
70
Year (0 = present)