Time for a change?

Single cohort (individual) vs multiple cohort (population) CEA

Please note that this is a work in progress and any constructive comments are welcomed.

Background

CEA (cost effectiveness analysis) is usually performed per individual/single cohort, ie treatment begins at same time for all individuals, which we might call a *single cohort model*. This method has been criticised on the grounds that it assumes some value for *lambda*, i.e. the marginal willingness to pay for a QALY and does not value the *opportunity cost* (Gafni and Birch 2006). In order to estimate the opportunity cost the total number of patients treated is needed (Sendi and Maiwenn 2003). However, we will further argue in this paper that the total size of the cohort is insufficient. In fact there are grounds for explicitly considering that not all individuals are treated at the same time i.e. time 'zero', but that there is *incidence* ie new individuals arriving each period e.g. each year, which we might call a *multiple cohort model*. The crucial condition for this to be true is that there is value to decision making of showing the cost and effectiveness consequences at a *finite* and not only *infinite* or *patient lifetime* time horizon.

Therefore aim of this paper is to show firstly that there might be value, from a decision making perspective, in using a finite time horizon or, more precisely, *single* finite time horizon. Then, it will be shown, using hypothetical scenarios that, given a finite time horizon, a multiple cohort model is preferable in terms of estimating total (over all individuals funded from the budget) cost and effectiveness as well as the ICER.

Finite versus infinite time horizon

Before proceeding to the exposition of multiple vs single cohort models let us first examine some of the key assumptions often made in CEA currently.

1) Future cost and benefit

Why do we count the costs and benefits that are incurred not only immediately (on adoption), but *over time*? The answer appears to be obvious, but most fundamentally because those costs and benefits *accrue to* (affect) the decision maker (and NOT the patient unless the patient is the decision maker). But a question remains regarding:

2) Time horizon

Why might we use infinity? One answer is that 'any other duration' is arbitrary (and therefore subject to potential 'selection bias' and, of course, any 'cut-off' implies the loss of counting of some future consequences. One other solution that is in keeping with recommendations for economic evaluation alongside trials is to use a variety of time horizons (Ramsey, Willke et al. 2005). However, the 'modelling equivalent of these ISPOR guidelines suggests that a lifetime (of the patient population) horizon is probably required if there is difference in survival. Actually, for a given single patient cohort, lifetime could be considered as equivalent to infinity in order to be certain that everyone in the cohort has died. Of course, practically, since no one has lived beyond a particular age this is unnecessary. Also, usually most technologies are considered has having a finite lifetime.

However, in estimating the *expected value of perfect information* (EVPI), although a finite technology lifetime is assumed and multiple cohorts occur over that time, each cohort is *followed up* i.e. all consequences of all cohorts are counted equally (Ades, Lu et al. 2004). 'Follow up' means that even when a technology (or perhaps, more precisely, the comparison of technologies) is assumed to have a finite life the costs and effectiveness of even the cohort that enters in the final year of the technology's life are counted beyond that year. Thus in effect there is no single time horizon: as the analysis counts costs and effects for each cohort for the same duration as all of the other cohorts, which could be lifetime. This contrasts the notion of drawing the line at some single finite time, which could be described as implying 'no follow-up' (beyond the technology lifetime).

In summary, the issue of time horizon in terms of finite versus infinite might be better expressed as one of:

A) 'no follow up' (where there is a single cut-off point such that consequences for all cohorts are counted up to the same time horizon, such as technology lifetime) versus
B) 'follow up' (where each cohort 'runs' to a different cut-off, i.e. the consequences of each cohort are counted regardless of e.g. technology lifetime)

We would argue that follow up does not *guarantee* accuracy. Generally, we would argue that if the valuation of consequences (and their uncertainty) is *subjective* then each decision maker who might use such a valuation might like to make their own *judgement*. Therefore, one cannot, *a priori* without making an assumption, state that there is no value in showing the consequences for single finite time horizons. In particular, we would question the need to always and only present the results with follow-up of cohorts beyond a finite time horizon such that all consequences of all cohorts are counted. For example, any investment in a new health care technology might be shown to be cost saving (as well as health improving) 'in the long run'. The question is can we assume that knowing the (single) point in time that it will take to break even is irrelevant to decision making. Indeed the argument has been made elsewhere that the decision making perspective should dictate the extent of costing (Brouwer, van Excel et al. 2006), which might be extended to include the time horizon.

Therefore, the more particular and relevant situation that this paper addresses is what might happen at a finite (or finite points) in the future that means that the costs or effects of all cohorts be counted at this point (or those points). Although this paper cannot provide a list of all such circumstances, it does seem reasonable to make some suggestions. For example:

1) Consistency with technology lifetime. Since follow up implies that cohorts continue on the same 'pathway' including any implied treatment, would it not be reasonable to consider a scenario where any savings or benefit after the end of the technology were not counted?
2) Price decrease due to 'coming off patent'. This is quite common and to some extent predictable. Each cohort would incur the price change at different points in their lifetime.
3) Effect of demand for equipment on cost where demand is uncertain. This example will be explored in the context of an HTA on surgical treatments for Benign Prostatic Enlargement (BPE) below.

We conclude this introduction by stating that we cannot rule out there being value to a decision maker of knowing what will happen at some single finite time horizon. Below we will show that, given such a finite time horizon and except under a restrictive condition, a *multiple cohort* as opposed to the traditional *single cohort model* in economic evaluation produces are more accurate estimate of cost (and by logical extension, effects and the ICER). This will be shown graphically and without formal proof at this stage. The next section provides the exposition of the models. A further hypothetical example will then be presented to show a change at a finite point in the future that differentially affects cohorts. One final example shows how a multiple cohort model was necessary for estimating capital cost in treating BPE when such a cost is a function of uncertain demand.

Exposition

Single cohort

Consider a technology for N individuals in the cohort such that costs are incurred each year, k. Formally:

$$SCy = N.C_k.(1/1+r)^{k+1}$$

Where r is the discount rate.

Therefore, the incremental costs of technology a minus b are:

$$ISCy = N.(Cak-Cbk).(1/1+r)^{k+1}$$

And total incremental costs over time horizon Y=y-1 years per individual are the sum over all years from k=0 to k=y:

$$TISCy = N.\Sigma(C_{ak}-C_{bk}).(1/1+r)^{k+1}$$

Multiple cohorts

Consider that in year 0 there are N new patients to be treated, each therefore at year 0. In year 1 these N patients then are in their next year, i.e. year 1 for them. However, additionally, more *new* patients now begin treatment, ie at year 0 for them. The following year, the first cohort are in year 2, the second in year 1 and a new cohort is now at year 0. In order to simplify the exposition, although generalisation to more complex (realistic) cases can be made, we will assume that the number of new patients per year, N, is constant over time. Generalisation to any time increment, eg months or days are also unnecessary for exposition here.

Therefore the cost in year zero of the evaluation (j=0) is simply the year zero cost of the first cohort (k=0) at time 0, ie:

$MC0 = N.C_{0k}$

The cost in year 1 of the evaluation (j=1) is the sum of the cost of year 1 of cohort 1 and year 0 of cohort 2:

$MC1 = N.(C_1+C_0).(1/1+r)$

Generally, the cost in year j, where k is between 0 and j, is given by:

$MCy = N.(\Sigma_k C_{jk}).(1/1+r)^{j+1}$

Therefore, incremental cost of a-b at year y is given by:

$IMCy = N.(\Sigma_k (C_{ak}-C_{bk})).(1/1+r)^{j+1}$

Therefore total incremental cost over y years, where j is between 0 and y and k is between 0 and j, is given by:

$TIMCy = N.\Sigma_j((\Sigma_k(C_{ak}-C_{bk})).(1/1+r)^{j+1})$ (2)

Comparing with equation (1):

$TISCy = N.\Sigma_k((C_{ak}-C_{bk}).(1/1+r)^{k+1})$ (1)

Graphically, if we have incidence of N per year then the pattern the incremental costs (or effects) will be as in table below:

WITHOUT FOLLOW-UP

		Cohort year						
Year from technology introduction		0	1	2	3	4	5	6
	0	N	0	0	0	0	0	0
	1	N	N	0	0	0	0	0
	2	N	N	N	0	0	0	0

	3	N	N	N	N	0	0	0
	4	N	N	N	N	N	0	0

The label 'no follow-up' is indicated in that after the 5th year no further cohort years are counted. 'Cohort year' should be contrasted with 'Year from technology introduction'. The former counts time from when an individual begins one of the compared technologies whereas the latter counts time from when any new technology is introduced (more precisely the start of the technology comparison).

Looking at the table, moving down, time from technology introduction increases and, as it does, each year more cohorts begin the technology i.e. enter cohort year 0. In year 0 there are none in any cohort year beyond the first because no-one has entered their second year yet. One year post technology introduction, those who began 'treatment' in year 0 now enter year 1 and show up as N in the box indexed by cohort year 1 and year from introduction 1. If one wants to follow a single cohort then the table needs to be read diagonally right and down.

Therefore, generally, for a time horizon of Y years, in year y there will be y lots of N who have been through cohort year 0, and thus incurred the year 0 costs. Similarly there will be y-1 lots of N who have been through cohort year 1 and thus incurred the year 1 costs, etc.

It can be seen that the total number of cohorts (discounted or not) is a multiple of N that depends only on time horizon Y. It can also be shown (proof not provided) that if incremental cost does not vary by cohort year i.e. $(c_{ak}-c_{bk})' = (c_{ak}-c_{bk})''$ for all cohort years k then total incremental multiple cohort cost is a multiple of total incremental single cohort cost that only depends on Y. Therefore, under this restriction, single cohort analysis would be equivalent to multiple cohort analysis.

However, if incremental cost does vary by cohort year then single will not equal multiple (proof not given), assuming no follow-up. Intuitively and as can be seen from the table this is because for any finite period with no follow-up there are always more individuals in the earlier treatment stages, precisely one more cohort in stage 0 than in stage 1 and one more in stage 1 than stage 2 and so on. Of course, as time horizon lengthens the total number of cohorts at any stage becomes larger *relative* to this difference and therefore the single cohort total approaches the multiple cohort total multiplied by the number of cohorts. In fact, there will be equality if each cohort is 'followed up' for their lifetime, as shown below:

WITH FOLLOW-UP

		Cohort year						
Year (from technology introduction)		0	1	2	3	4	5	6
	0	N	0	0	0	0	0	0
	1	N	N	0	0	0	0	0
	2	N	N	N	0	0	0	0
	3	N	N	N	N	0	0	0
	4	N	N	N	N	N	0	0
	5	0	N	N	N	N	N	0

	6	0	0	N	N	N	N	N
	7	0	0	0	N	N	N	N
	8	0	0	0	0	N	N	N
	9	0	0	0	0	0	N	N
	10	0	0	0	0	0	0	N
	11	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0

In this example, the technology also has a lifetime of 5 years i.e. up to and including year 4. However, instead of only including consequences of cohorts up to this time, cohorts are followed up such that each and every consequence of each and every cohort that received the technology is counted. The table would need to be extended to the right and down to include every cohort year until death.

Finally, it can be shown that the difference between multiple and single cohort models is generalisable for any technology lifetime including infinite and also with discounting. The next section gives some simple numerical examples with a comparison of technologies.

Hypothetical example to show difference in recommendation between multiple and single cohort models

In this example, a single finite time horizon of eleven years is assumed. For simplicity the discount rate is set to zero. Consider three technologies, where, again for simplicity, 'three' is more effective than 'one', which is more effective than 'two' and with the following costs per individual over 11 years:

	single cohort (cost per individual)			**incremental analysis**	
technologies	**one**	**two**	**three**	**one-two**	**three-one**
0	2000	100	200	1900	-1800
1	10	100	200	-90	190
2	10	100	200	-90	190
3	10	100	200	-90	190
4	10	100	200	-90	190
5	10	100	200	-90	190
6	10	100	200	-90	190
7	10	100	200	-90	190
8	10	100	200	-90	190
9	10	100	200	-90	190
10	10	100	200	-90	190
total over 11 years	**2100**	**1100**	**2200**	**1000**	**100**

Incremental costs over 11 years show that, given the effectiveness ordering, the recommendation would be to adopt any one of the technologies, depending on further information (traditionally some measure of willingness to pay per unit of

effectiveness, or, more correctly, the budget and the incremental costs and effectiveness of other independent technologies).

However, the results for multiple cohorts, each one, for simplicity having N=1 are below:

	population (multiple cohorts-one new individual per year)				
technologies	**one**	**two**	**three**	**three-two**	**three-one**
0	2000	100	200	100	-1800
1	2010	200	400	200	-1610
2	2020	300	600	300	-1420
3	2030	400	800	400	-1230
4	2040	500	1000	500	-1040
5	2050	600	1200	600	-850
6	2060	700	1400	700	-660
7	2070	800	1600	800	-470
8	2080	900	1800	900	-280
9	2090	1000	2000	1000	-90
10	2100	1100	2200	1100	100
total over 11 years	**21550**	**6600**	**13200**	**6600**	**-8350**

Clearly, the recommendation would be different. In particular, 'one' can be ruled out since it is both less effective and the more costly than 'three'. Whether 'one' or 'two' should be adopted still depends on the opportunity cost. Of course, multiple cohort cost is a multiple of the single cohort cost. Notice that this multiple is the same and varies only with time, but only for technologies 'two' and 'three'. As stated above, this is because cost per cohort year for technologies 'two' and 'three' does not vary over time.

Of course, we have assumed a single finite time horizon and no discounting. If we apply discounting, but leave the time horizon unchanged then there will also be a change in rank. It is only if we permit each cohort to continue for the same period that this effect disappears and then the total cost (or effectiveness) ratio of multiple to single becomes the same for all technologies.

Example to estimate demand related capital cost-the case of BPE

A CEA was conducted as part of an HTA to compare sequences of surgical treatments for benign prostatic enlargement (BPE). A time horizon of 10 years was chosen since this was the period over which current technologies were believed to remain relevant. One of the issues in this analysis was the there were sequences of treatments, the length of the sequence depending on rate of initial failure and later relapse Variable costs could have been handled using a single cohort model. However, one sequence included treatment with a new piece of laser equipment for a procedure called Holmium laser enucleation of prostate (HoLEP), but only on failure or relapse. Therefore the demand for HoLEP (and therefore need to purchase) was endogenous to the model and in particular a function of total numbers treated. Therefore, a multiple

cohort model enabled calculation of the number of new patients needing HoLEP each year and equipment only purchased according to the estimated capacity.

Discussion

The traditional method of performing cost effectiveness analyses is to estimate each consequence, cost or effectiveness, per individual, whether using cohort or individual based (Monte Carlo) simulation. Population based analyses are generally only used in limited circumstances, one being to account for individual interactions such as due to queuing or disease transmission or for estimating the total budgetary impact. Of course, the latter case has been argued to be the correct method in order to estimate opportunity cost, as opposed to calculating an incremental cost effectiveness ratio. Even in EVPI calculations, when the technology is considered as having a finite time horizon and multiple cohorts considered, the net present value is still estimated using follow up for all cohorts each cohort having a cut off at the same point in its lifetime, but different points from technology introduction.

However, this paper has questioned this approach if there is any value in knowing the results of analyses over single finite time horizons thus counting the cost or effect for all cohorts at different points in each of their lifetimes. Examples were suggested as to when this would occur. For example, in sequences of treatments for BPE where total cost (and therefore average individual cost) is a function of uncertain demand. What has been demonstrated is that under this condition (except under the restriction that neither cost nor effectiveness varies over an individual's lifetime) that the multiple cohort model is more accurate. In the BPE example, at each single point cost of equipment accumulates as members of each cohort fail early treatment or relapse, thus also requiring a multiple cohort analysis. Whether any recommendation according to the single cohort model would be incorrect is an empirical matter and depends on the opportunity cost.

The implications of this these conclusions could be far reaching in that they add a question to raditional approaches to CEA. Perhaps it is 'time for a change'.

As stated at the beginning, this is a work in progress and therefore there follow some questions for discussion.

Questions

1) Is this method of framing the problem correct?
2) What do we do about finite technology lifetimes?
3) Should we always follow up cohorts to end of life (infinity)?
4) Is the length of the time horizon a matter for the decision maker/decision making context?
5) Should we try different discount rates, in particular increasing over time?
6) Is there another method of estimating demand related cost?

7) Is there another method of estimating the effect of price changes?
8) What do we call changes that occur at some finite time in the future-system changes?
9) Should we ignore such changes or is there another way of dealing with them?

References

Ades, A. E., G. Lu, et al. (2004). "Expected value of sample information calculations in medical decision modeling." Medical Decision Making **24**: 207-227.

Brouwer, W. B. F., N. J. A. van Excel, et al. (2006). "A dollar is a dollar is a dollar-or is it?" Value in Health **9**(5): 341-347.

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

Ramsey, S., R. Willke, et al. (2005). "Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA task force report." Value in Health **8**(5): 521-533.

Sendi, P. and J. A. Maiwenn (2003). "Revisiting the decision rule of cost effectiveness analysis under certainty and uncertainty." Social Science and Medicine **57**: 969-974.