WORK IN PROGRESS: PLEASE DO NOT CITE OR QUOTE

Alternative definitions of multi-way cost-effectiveness acceptability curves and frontiers

Pelham Barton

Department of Health Economics

University of Birmingham

Birmingham B15 2TT

United Kingdom

p.m.barton@bham.ac.uk

Summary

This paper is concerned with alternative definitions of the multi-way cost-effectiveness curve (CEAC) and frontier (CEAF). Examples are used which highlight two main problems of the conventional definition. First, the order of probabilities between two options in a multi-way CEAC may not be the same as the order in a bilateral CEAC considering just those two options. Second, the conventional CEAF is liable to collapse towards zero as the number of options becomes very large. An alternative definition is proposed which overcomes these problems at least in the cases illustrated. However, the alternative definition is not without problems of its own.

The conclusion of the paper is that there is no truly satisfactory definition of a multi-way CEAC, and that it is better to present uncertainty either by using measures such as value of information, which take account of the potential consequences of a non-optimal decision, or through a suitable selection of bilateral CEACs.

1. Introduction

The cost-effectiveness acceptability curve (CEAC) was introduced by van Hout and colleagues (1994) and has become widely accepted in a Bayesian context as a means of showing the uncertainty in a choice between two alternative strategies for treating a patient group. It is usually drawn to show the probability that a proposed "new" strategy is cost-effective compared to the "current" strategy, plotted against the threshold value of the incremental cost-effectiveness ratio (ICER). In the context of model-based analysis, the input parameters to the model are sampled a sufficient number of times (usually between 1000 and 10,000) and it is assumed that the proportion of model replications favouring a given option *X* is a fair estimate for the probability that option *X* will be cost-effective. The underlying assumption is that the model structure is adequate and that all relevant uncertainty is captured within the parameter distributions. This assumption will be sustained throughout the present paper.

Occasionally the bilateral CEAC is drawn showing both options. The two probabilities plotted have a sum of 1 and a mean of 0.5.

The preferred option at a given threshold ICER is not the option with the higher probability of being cost-effective, but is determined by the mean of the distribution of costs and effects (Claxton, 1999). It is well known that the option with the higher probability may not be preferred on mean cost-effectiveness: Fenwick and colleagues (2001) introduced the cost-effectiveness acceptability frontier (CEAF) as a solution to this problem.

2. Alternative interpretations of the bilateral CEAC extended to three or more options

Suppose two options *X* and *Y* for treating a given patient group are being compared. Then the following definitions (at a given threshold ICER) all give the same result:

- *(a)* the probability that *X* is cost-effective compared to *Y*;
- *(b)* the probability that *X* is the optimal option from those considered;
- *(c)* the expected number of other options compared to which *X* is preferred;

(d) the expected proportion of other options compared to which *X* is preferred.

Definition *(a)* regards the CEAC as essentially to do with a comparison of two options. When there are three or more options under consideration, it is possible to construct a range of (bilateral) CEACs. In principle, a bilateral CEAC can be drawn for any pair of options, but in practice it is up to the analyst to present an appropriate selection. It would, for example, be reasonable to present a CEAC between any pair of options for which an ICER is quoted. If an option is dominated in mean values, but has an appreciable probability of being optimal across the distribution of parameter values, then at least one bilateral CEAC including the dominated option should be given.

Definition *(b)* has been widely used as the means of extending the concept of a CEAC to three or more options. For example, Briggs and colleagues (2006, p. 134) state: "The characterizing feature of the presentation of multiple CEACs to represent multiple and mutually exclusive treatment options is that the curves sum to a probability of one vertically." This definition of a CEAC will here be referred to as Version 1.

When applied to *n* options, where $n \geq 3,$ definitions *(c)* and *(d)* differ only in that, to go from *(c)* to *(d)*, it is necessary to divide by $n-1.$ For comparability with other definitions, definition *(d)* has the advantage that the values taken are constrained between 0 and 1 regardless of the number of options. This definition will here be called the Version 2 CEAC. Unlike Version 1, it preserves the mean of the plotted values at 0.5, while the sum of the plotted values increases with *n*.

2.1 A continuum of definitions for three options

Suppose that there are three options *X*, *Y*, and *Z* for treating a patient group. At a given threshold ICER $\lambda,$ let $p_{X1}(\lambda)$ be the probability that *X* is the optimal option, let $p_{X2}(\lambda)$ be the probability that *X* is preferred to exactly one other option, and let $p_{X3}(\lambda)$ be the probability that *X* is preferred to no other option, so that $p_{X1}(\lambda)+p_{X2}(\lambda)+p_{X3}(\lambda)=1.$

(The ordering of the options is the order in terms of net monetary benefit. In the rare cases where two or more options have equal NMB for a particular parameter set at a given value of λ, the probability can be divided equally between the options.)

Then the Version 1 CEAC for *X* plots $p_{X1}(\lambda)$ as a function of λ, while the Version 2 CEAC plots $p_{X1}(\lambda)+\frac{1}{2}p_{X2}(\lambda)$ as a function of λ. It then becomes clear that it would be possible to plot $p_{X1}(\lambda)+tp_{X2}(\lambda)$ as a function of λ, for any value of *t*. Potentially sensible values are within the range $0 \le t \le 1$. It can be seen that Version 1 is at the lower extreme of this range, but Version 2 is in the middle. The CEAC defined by $t = 1$ can be called Version 3, and corresponds to a further possible definition of the bilateral CEAC, namely the probability that option *X* is not the worst of the options considered.

(Intermediate values could be defined, for example Version $1\frac{1}{2}$ would have $t = \frac{1}{4}$, and it is possible to define the bilateral CEAC in a way which generalises to give this version. However, it is not at all clear what the use for such a definition might be.)

2.2 Four or more options

Versions 1, 2, and 3 of the CEAC extend naturally to any number of options. The space of intermediate versions now becomes multi-dimensional. Details are in Appendix 1.

3. Examples

The above concepts are illustrated with a selection of examples.

3.1 Example 1

The underlying model for this example was introduced by Sadatsafavi and colleagues (2008). Small changes in the notation have been made for convenience.

In this case, there are three strategies *X*, *Y*, and *Z* for treating some patient group, where strategy *X* is the baseline option representing current practice, strategy *Y* involves the use of a new drug treatment, and strategy *Z* involves the same new treatment in combination with a second (palliative) drug, which gives small increases to both costs and outcomes in QALYs. For the purpose of this example, it is assumed that the output of the model can be approximated by Normal distributions, as shown in Table 1.

Table 1 The Sadatsafavi model

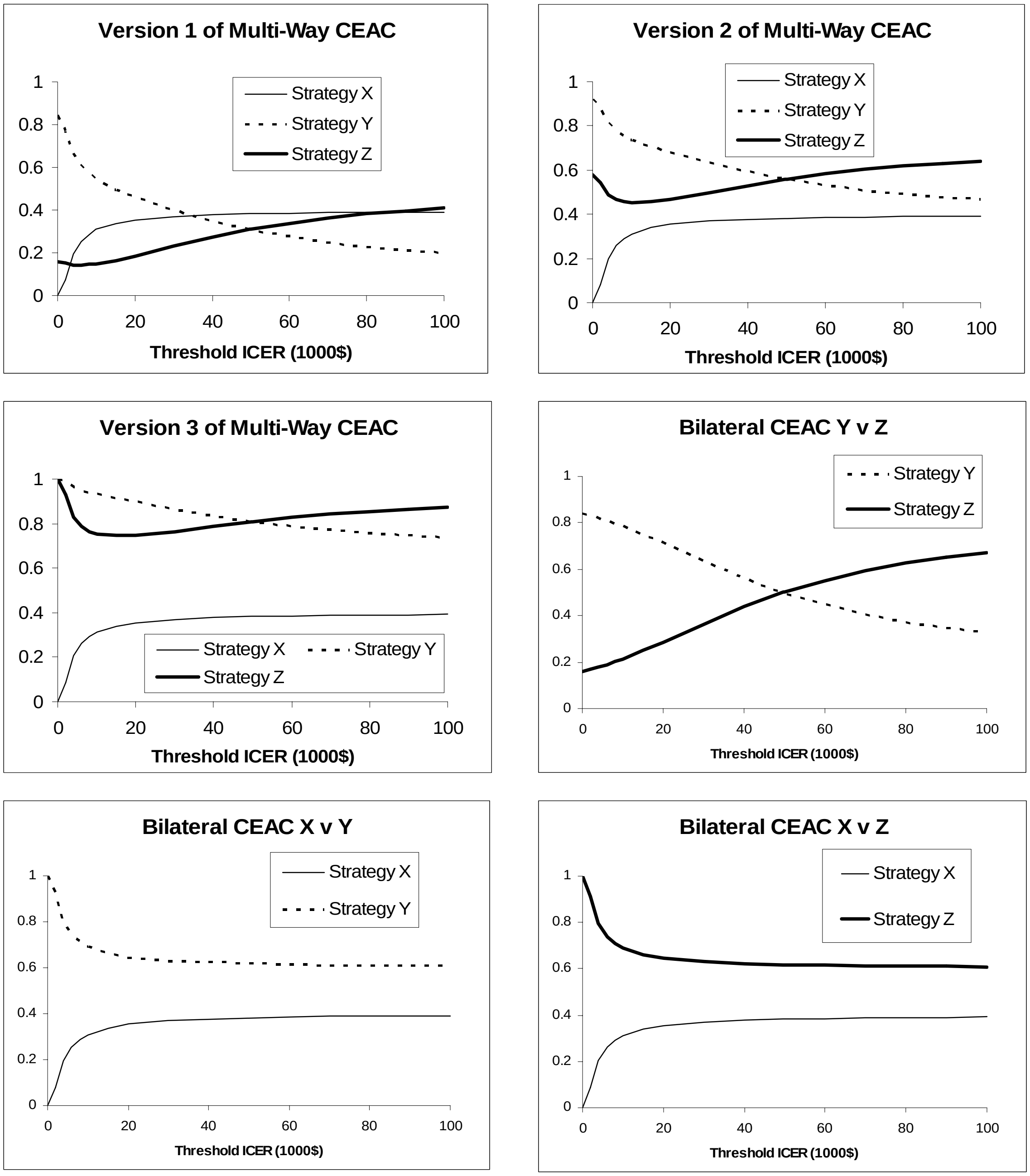
	mean	s.d.
Difference in costs (in US$) *Y* over *X*	-1000	200
Difference in QALY outcome *Y* over *X*	0.1	0.4
Difference in costs (in US$) *Z* over *Y*	50	50
Difference in QALY outcome *Z* over *Y*	0.001	0.001

Based on mean values, each of strategies *Y* and *Z* dominates strategy *X*, while the ICER between strategies *Y* and *Z* is $50,000/QALY. The various bilateral and multi-way CEACs for this model (based on 20,000 samples from the distributions in Table 1) are shown in Figure 1.

Version 1 of the Multi-Way CEAC is a reconstruction of Figure 1 in the paper by Sadatsafavi and colleagues (2008). Although strategy *X* is dominated in the mean by both the other strategies, it still has the highest CEAC for a substantial range of threshold ICERs. This is essentially because, for thresholds above $10,000/QALY, *X* is actually the preferred strategy around 40 percent of the time, while for almost all of the other 60 percent of the time, both *Y* and *Z* are preferred to *X*. The 60 percent probability is divided more or less evenly between *Y* and *Z*, so that for a substantial range of thresholds, neither *Y* nor *Z* has a 40 percent chance of being optimal.

In contrast to Version 1, Version 2 of the multi-way CEAC always shows the options in the same order as the various bilateral CEACs. The crossing point between the curves for *Y* and *Z* is in each case close to the theoretically correct value of $50,000/QALY: the small observed difference is within the range of sampling error.

Figure 1 Various CEACs for the Sadatsafavi model



Legend: In each case the vertical scale is a probability scale as defined in the appropriate part of the text. The top left-hand chart is a reconstruction of Figure 1 from Sadatsafavi *et al* (2008, p.307).

From the distribution as sampled, there is a small range of thresholds (between $49,581 and $49,785) at which the Version 1 curve for *Y* is higher than that for *Z*, although in the bilateral comparison, the curve for *Z* is higher than that for *Y*. No threshold could be found at which the Version 2 curve reversed the order of the bilateral CEAC, although there was a tiny range over which the bilateral curves coincided where the Version 2 curves did not. When a different set of pseudo-random numbers was used, a similar sized range was found at which the Version 1 curve for *Z* was higher than that for *Y*, although in the bilateral comparison, the curve for *Y* was this time higher than that for *Z*. Again it was not possible to find a reversal of the order between the bilateral curves and the Version 2 curves.

In this case, the reversal of the order between *Y* and *Z* comparing Version 1 curves to the bilateral curves may be the result of sampling error. However, it is not difficult to construct cases in terms of probability distributions where such reversal is demonstrably a genuine effect. The following example is based on one previously used by the present author (Barton, 2008).

Suppose that a model compares three options *X*, *Y*, and *Z*, where *Y* and *Z* are quite similar and *X* is somewhat different from the other two. Suppose that the order of preference in these options at some threshold ICER under different replications of the model is as follows:

0.30 *Y Z X*; 0.28 *Z Y X*; 0.23 *X Z Y*; 0.17 *X Y Z*; 0.01 *Y X Z*; 0.01 *Z X Y*.

(These probabilities have simply been constructed as final results. To adapt the Sadatsafavi model to produce results such as these would require the introduction of further correlation such that *Y* is likely to be better than *Z* when they are both better than *X*, but *Z* is likely to be better than *Y* when *X* is preferred.)

Then the Version 1 CEAC at that threshold will show probabilities as follows: *X* 0.40, *Y* 0.31, *Z* 0.29, showing (correctly) that *X* is the most likely of the three to be the preferred option. However, if we compare the options two at a time, we have the following:

X v *Y*: *X* 0.41, *Y* 0.59; *X* v *Z*: *X* 0.41, *Z* 0.59; *Y* v *Z*: *Y* 0.48, *Z* 0.52.

In all three cases, the order between options in the pairwise comparison is the reverse of the order in the three-way comparison (albeit marginally in the case of *Y* and *Z*).

When the Version 2 CEAC is calculated for this example, we have *X* 0.41, *Y* 0.535, *Z* 0.555, while for the Version 3 CEAC in this case, we have *X* 0.42, *Y* 0.76, *Z* 0.82. Each of these preserves the order of the bilateral comparisons.

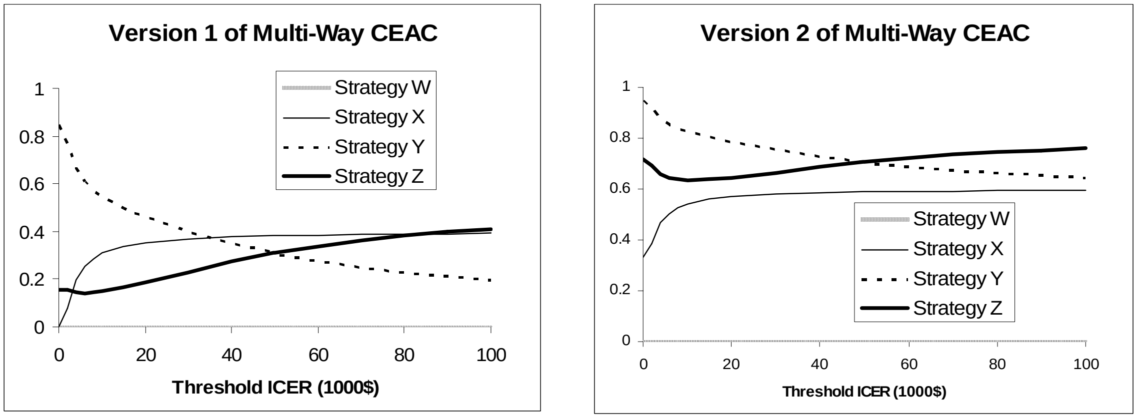
Returning to the example of Sadatsafavi and colleagues, Version 3 again preserves the order between the curves, except for very low values of the threshold ICER (including zero), where strategy *X* was the worst (in this case, most costly) strategy in all replications of the model. In this case, the Version 3 curves for *Y* and *Z* both take the value 1. (In theory, given the unlimited nature of the Normal distribution, there is a small non-zero probability that *X* is not the most costly strategy. However, it should be remembered that this is a stylised model, and the Normal distributions are taken to approximate reality. This is a case where a manageably finite number of samples is likely to approximate reality more closely than an infinite sample, or even an unmanageably large finite sample.)

3.2 Example 2

For this example, we consider the effect of adding a fourth option to the Sadatsafavi model. The new strategy *W* is designed to be clearly worse than all the other strategies. It is defined by the difference in cost (in US$) of *X* over *W* having a mean of -5000 (s.d. 200) and a QALY gain of *X* over *W* with mean 2 (s.d. 0.2). Figure 2 shows Versions 1 and 2 of the multi-way CEAC for the extended choice set. Version 3 is not shown: the curves for *X*, *Y*, and *Z* all take a constant value 1, while the curve for *W* is constant at 0.

It can be clearly seen that (apart from the additional curve for strategy *W*), the Version 1 curves are identical to those in Figure 1 for the three-way comparison, while the Version 2 curves have been compressed towards the top of the graph. While this compression preserves not only the order of the curves but also the ratio of the distances between them, the absolute values have definitely changed.

Figure 2 Multi-way CEACs for an extended version of the Sadatsafavi model.



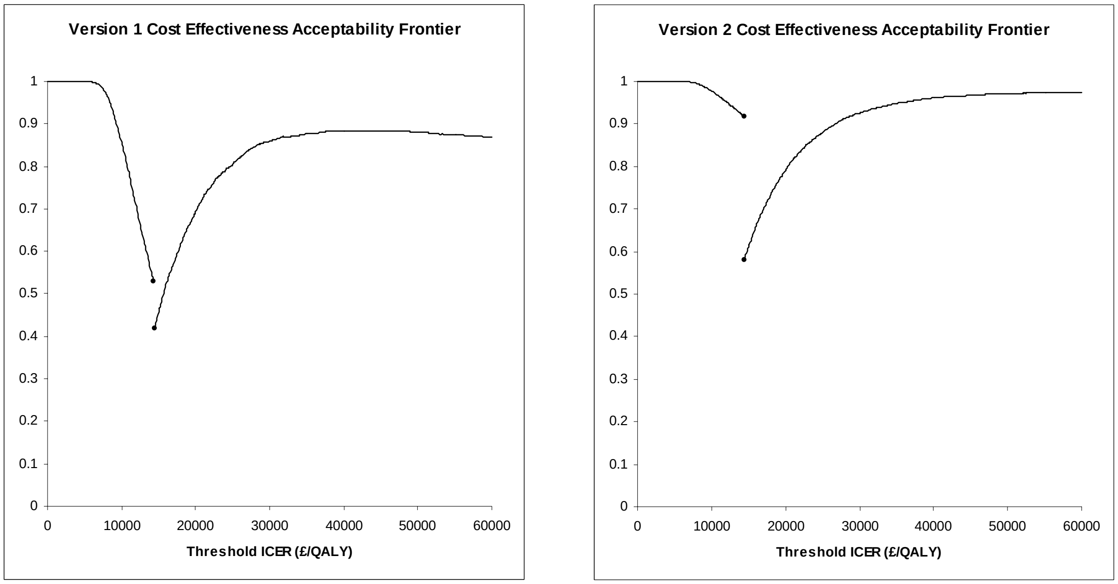
Legend: In each case the vertical scale is a probability scale as defined in the appropriate part of the text. The "curve" for Strategy *W* in each case takes a constant value of 0.

3.3 Example 3

This example considers the effect of working with a very large number of options. It is based on a model introduced by the present author in a previous HESG paper (Barton, 2008). See Appendix 2 for more details. Because of the large number of options in the model, only cost-effectiveness acceptability frontiers (CEAFs) are shown. As defined by Fenwick and colleagues (2001), the CEAF for a given decision problem is constructed by determining the optimal strategy based on the expected outcomes given current uncertainty and plotting the probability that it would remain optimal were the uncertainty to be removed. The CEAF thus typically consists of segments from two or more (Version 1) CEACs. Using Version 2 CEACs instead of Version 1 naturally gives us the Version 2 CEAF. (Version 3 and intermediate versions will not be considered for this example.)

The model used here is concerned with once-in-a-lifetime screening for a hypothetical cancer and the decision is at what age this screening should be carried out, if at all. Figure 3 shows the two versions of the CEAF for the case where the possible screening ages considered are any multiple of 16 years (from 0 to 112). The optimal policy (based on mean outcomes from the model) is no screening below £14,400/QALY and screening at age 48 from this figure up to at least £60,000/QALY, which is the highest threshold ICER plotted.

Figure 3 Two versions of the CEAF for screening ages multiples of 16 years



For very low ICERs (up to £5,200/QALY), the no screening option is preferred in all replications of the model. In that case, both versions of the CEAF take the value 1. Otherwise, the Version 2 CEAF is by definition always higher than the Version 1 CEAF.

Figure 4 Two versions of the CEAF for screening ages multiples of 8 years

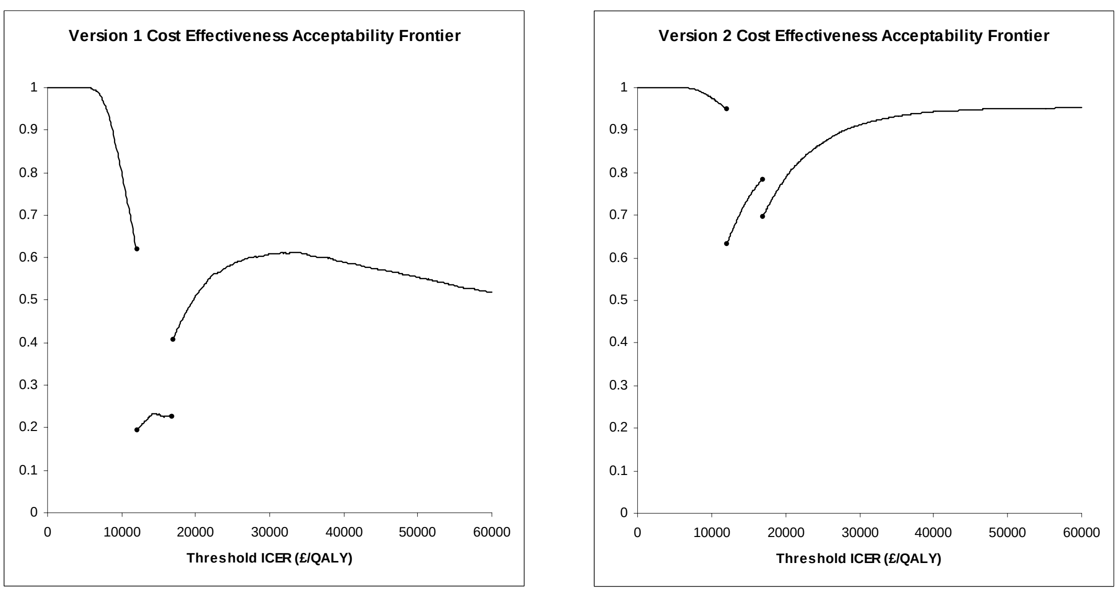


Figure 4 shows the result of including all multiples of 8 years from 0 to 120 as possible screening ages in the model. Screening at age 40 now becomes optimal for threshold ICERs from £12,100/QALY to £16,800/QALY. Unlike the Version 1 equivalent, the Version 2 CEAF can go higher with the inclusion of new options, even if the optimal strategy does not change.

Figure 5 Two versions of the CEAF for screening ages multiples of 4 years

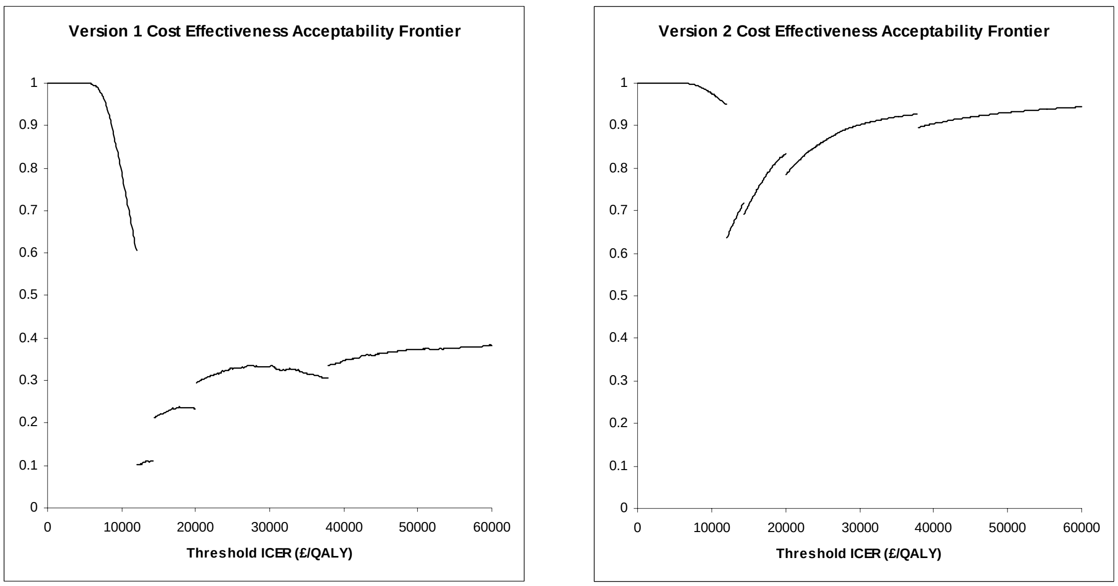
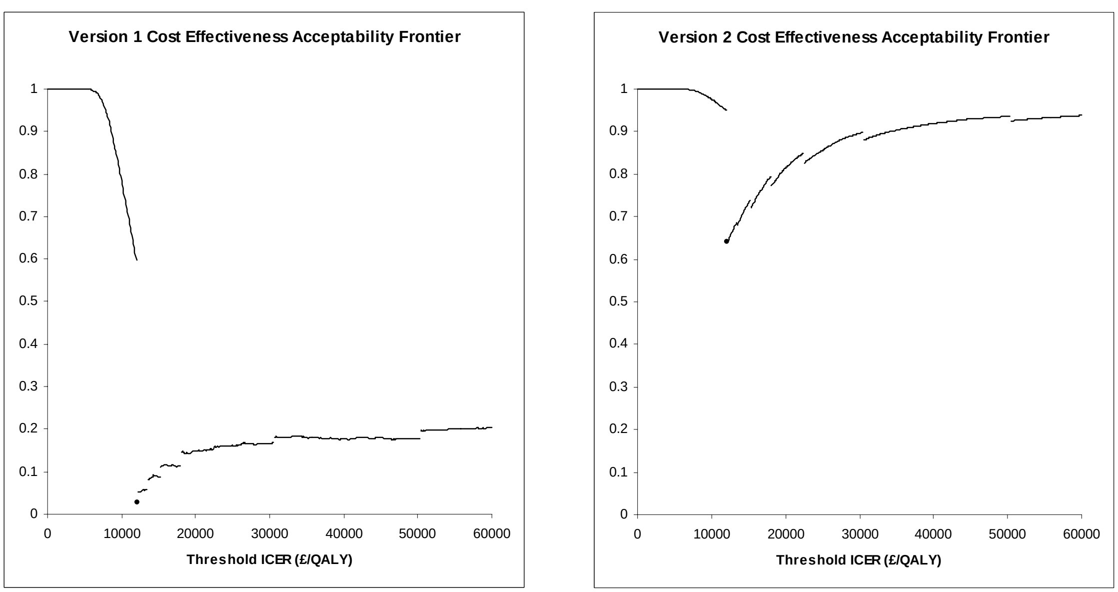


Figure 6 Two versions of the CEAF for screening ages multiples of 2 years



Figures 5 to 8 show the equivalent comparison as the number of available options is successively doubled. As previously noted (Barton, 2008) the Version 1 CEAF collapses for ICERs above £12,000/QALY, when screening at some age is preferred but the probability is divided among an increasing number of similar options. On the other hand, the Version 2 CEAF stabilises at a non-zero limit as the number of options increases.

Figure 7 Two versions of the CEAF for screening ages multiples of 1 year

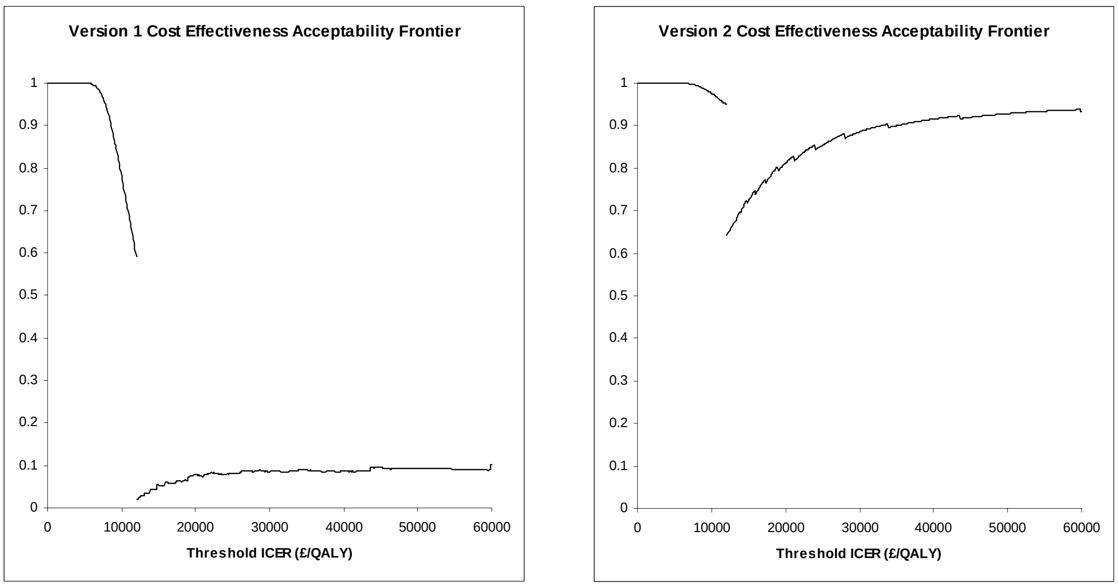
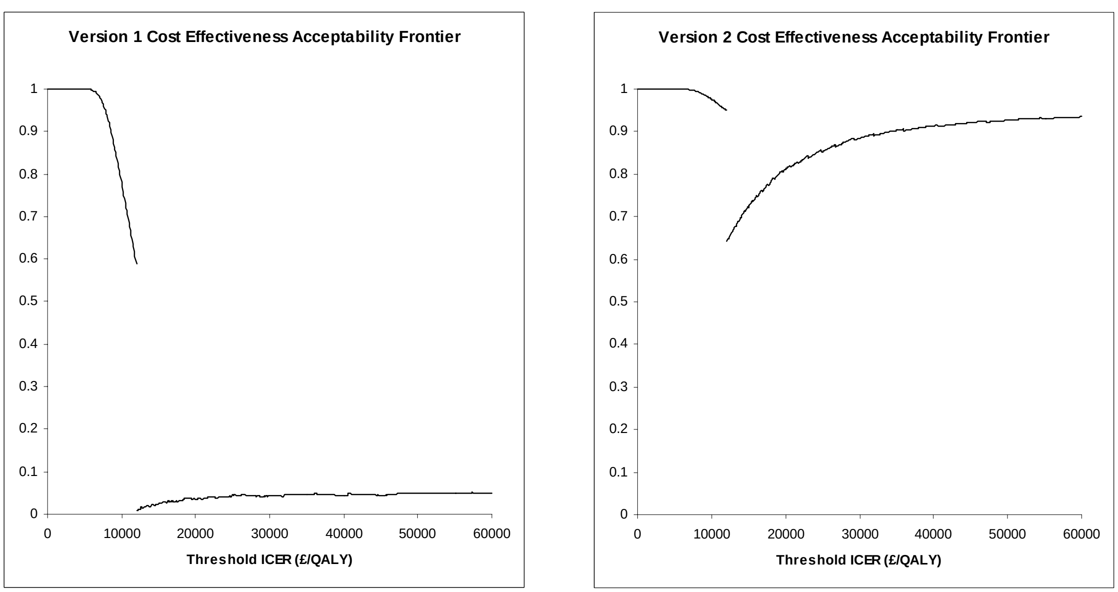
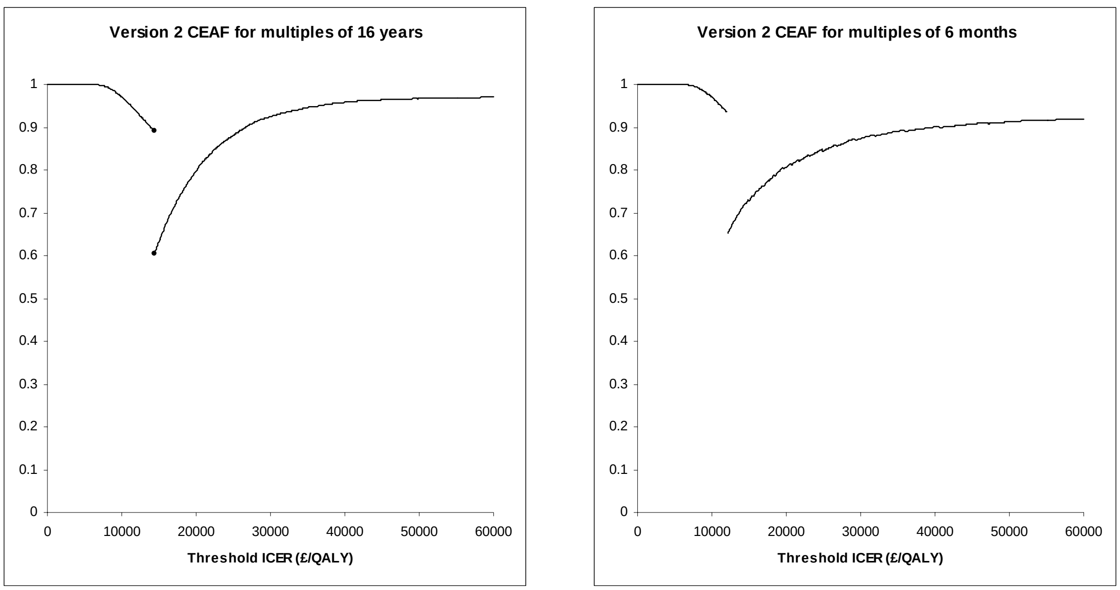


Figure 8 Two versions of the CEAF for screening ages multiples of 6 months



However, it is not clear what the limiting values in the Version 2 CEAF mean. As with the previous example, the curves are sensitive to inclusion of obviously poor options. In this case, all the curves in Figures 3 to 8 include options for screening at ages that would not be considered sensible. Figure 9 shows the effect of omitting all screening ages below the age of 32 (that is, one quarter of the positive screening options in the original model). The curves are measurably, but not hugely, lower than their equivalents in Figures 3 and 8. Omitting screening ages over 80 (as well as under 32) would be expected to multiply the magnitude of this effect by about 2.5.

Figure 9 Version 2 CEAFs omitting the possibility of screening below the age of 32



4. Discussion

A property of the Version 1 multi-way CEAC is that the order in which options appear on the curve can depend on the inclusion of other options. This is illustrated with the stylised examples in this paper, and also occurs for "real life" models of the type considered by Barton and colleagues (2008).

For all the examples shown in this paper, the reversing of order of probabilities between bilateral and Version 1 multi-way CEACs disappears when Version 2 multi-way CEACs are used instead. However, it is theoretically possible that such reversibility can occur with Version 2 curves. It remains an open question at the time of writing whether such examples can be constructed without the same reversibility occurring in the Version 1 curves.

Another point considered here is that, whereas Version 1 CEACs can collapse towards zero when the number of options becomes very large, this does not occur with Version 2 CEACs.

However, Version 2 CEACs are not without problems. Inclusion or omission of options with no chance of being preferred can change the position (though not the ordering) of Version 2 curves, while Version 1 curves are unaffected by such actions.

It was also noted that the two versions of the multi-way CEAC considered here belong to a continuum of different possible definitions. Version 1 is at one extreme of this continuum, while Version 2 is at the centroid of the continuum. The extreme furthest from Version 1 can be defined, but is not a sensible option.

4.1 A question of terminology

When considering a large number of options, the pursuit of optimality ceases to be appropriate. Rather what is sought is the distinction between good and bad options. One measure of this is that an option is good if it is better than a high proportion of other options. This notion coincides with the ordinary meaning of the word “acceptability”: it could be argued that a option that is (with certainty) better than half of the other options considered has an acceptability rating of (at least) 0.5, on a scale from 0 to 1. On this basis, the Version 2 CEAC accords more naturally with the ordinary meaning of the word “acceptability” than does the Version 1 CEAC, which might more properly be termed the “cost-effectiveness optimality curve”. It is however acknowledged that the name CEAC for Version 1 is too fully entrenched for a change to be possible now.

5. Conclusions

The alternative form of multi-way CEAC that is considered here is no more satisfactory for routine use than the conventional form. To show the uncertainty surrounding a decision with three or more options, it is necessary either to use some measure that includes the potential effect of a non-optimal decision, such as value of information, or to use a selection of bilateral CEACs. While this may involve more effort in presentation, this is preferable to presenting results in a form that is likely to mislead decision makers.

Appendix 1: The space of possible CEAC versions for four or more options

In the case of four options, the feasible functions that can be plotted for alternative forms of the CEAC are of the form $f_X(\lambda) = p_{X1}(\lambda) + t_2 p_{X2}(\lambda) + t_3 p_{X3}(\lambda)$, where $0 \le t_3 \le t_2 \le 1$. The space of all such functions can be plotted on a plane, and forms a right-angled isosceles triangle, with Versions 1 and 3, represented by the points $(0,0)$ and $(1,1)$ respectively, at the two corners furthest apart. The point midway between these points is the point $\left(\frac{1}{2}, \frac{1}{2}\right)$, which does not represent Version 2, but represents the function $f_X(\lambda) = p_{X1}(\lambda) + \frac{1}{2} p_{X2}(\lambda) + \frac{1}{2} p_{X3}(\lambda)$. This function corresponds to a notion of acceptability in which no distinction is made between second and third place in the list of four options (ordered by net benefit at a threshold ICER λ). Version 2 of the CEAC is represented by the point $\left(\frac{2}{3}, \frac{1}{3}\right)$, which is at the centroid of the triangle.

With five dimensions the feasible space is a right-angled tetrahedron with corners at the points $(0,0,0)$, $(0,0,1)$, $(0,1,1)$, and $(1,1,1)$. Again, Versions 1 and 3 are at the two corners furthest apart, while Version 2 is at the centroid of the tetrahedron. The same principle applies with six or more options, where the shape formed by the set of feasible options is now in multi-dimensional space.

Appendix 2: Specification of the cancer screening model

The model assumes that there is an asymptomatic, detectable and treatable state, but that if the cancer progresses, it becomes immediately fatal. Death from other causes is modelled by a Weibull survival curve with uncertain parameters, as are time from a modelled minimum age “at risk” to the treatable state and time from entry into the treatable state to progression. The natural history model is an individual sampling model (Barton *et al*, 2004; Brennan *et al*, 2006): for each individual the ages of death from other causes, entry into treatable state and death from cancer are sampled using the appropriate distributions.

The screening test is assumed to have sensitivity and specficity less than 100 per cent. In the case of a positive result from this test, a confirmatory test is applied which is assumed to be 100 per cent sensitive and specific. Individuals found to be in the treatable stage following the confirmatory test receive surgery, which is assemed to prevent death due to cancer, and then receive maintenance therapy for the reaminder of their life (until death from other causes). A small reduction in quality of life following surgery is also assumed.

The results in this paper are based on 5,000 replications of the model. For each replication, one set of parameters was sampled from the distributions described below. This parameter set was used to generate 100,000 individual natural histories, and the expected effect of screening was then calculated exactly for each natural history, in the interest of variance reduction. From this, the expected population cost and QALY gain from screening at any age could be estimated for that replication of the model.

For the natural history part of the model, the minimum age of onset of the treatable state and the parameters of the Weibull distributions were sampled from independent Normal distributions as shown in Table 2. Sensitivity and specificity of the screening test were set as bivariate normal on the logit scale with a negative correlation between them. Each was given a median of 0.98 and lower 95 percent limit of 0.97, with a correlation coefficient of -0.2 between the underlying normal distributions. For convenience, a similar univariate distribution was used for quality of life post-transplant, with a median of 0.9 and lower 95 percent limit 0.85. Finally, unit costs were assumed to be fixed at the values shown in Table 3. All costs and QALY gains from screening were discounted back to the time of screening using a continuous discounting function at an annual rate of 3.5 per cent.

Table 2.Parameters for time to event

Parameter		mean	s.d.
Age at death from other causes	α	8	1.5
	β	75	5
Minimum age of onset of asymptomatic cancer		30	4
Time from minimum age to actual onset of asymptomatic cancer	α	2	0.2
	β	122	20
Time from onset of asymptomatic cancer to progression of cancer	α	2	0.2
	β	22	4

Table 3 Unit costs

Item	Cost (£)
Cost of main screening test	50
Cost of confirmatory test	1000
Cost of surgical treatment	50,000
Annual cost of maintenance therapy following surgery	500

References

Barton GR, Briggs AH, Fenwick EAL. 2008. Optimal Cost-Effectiveness Decisions: The Role of the Cost-Effectiveness Acceptability Curve (CEAC), the Cost-Effectiveness Acceptability Frontier (CEAF), and the Expected Value of Perfect Information (EVPI). *Value in Health* published online 16 May 2008. doi: 10.1111/j.1524-4733.2008.00358.x

Barton P. 2008. What happens to value of information measures as the number of decision options increases? Paper presented at HESG, January 2008.

Barton P, Bryan S, Robinson S. 2004. Modelling in the economic evaluation of health care: selecting the appropriate approach. *Journal of Health Services Research and Policy* 9: 110-118.

Brennan A, Chick SE, Davies RM. 2006. A taxonomy of model structures for economic evaluation of health technologies. *Health Economics* 15: 1295-1310.

Briggs A, Claxton K, Sculpher M. 2006. *Decision modelling for health economic evaluation*. Oxford University Press.

Claxton K. 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics* 18: 341-364.

Fenwick E, Claxton K, Sculpher M. 2001. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Economics* 100: 779-787.

Sadatsafavi M, Najafzadeh M, Marra C. 2008. Acceptability curves could be misleading when correlated strategies are compared. *Medical Decision Making* 28: 306-307.

Van Hout BA, Al MJ, Gordon GS, Rutten FFH. 1994. Costs, effects and C/E-ratios alongside a clinical trial. *Health Economics* 3: 309-319.