Uncertainty and the Decision-Maker: Assessing and Managing the Risk of Undesirable Outcomes

Gafni A[1,2], Walter SD[1], Birch S[1,2,3]

1. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.
2. Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, Ontario, Canada
3. School of Community Based Medicine, University of Manchester, UK.

Corresponding author:

Amiram Gafni, Ph.D., Professor

Centre for Health Economics and Policy Analysis

McMaster University

Communications Research Laboratory (2nd Floor – 208)

1280 Main Street West

Hamilton, Ontario, Canada L8S 4K1

Phone: 905 – 525 9140 ext: 22681

E-mail: gafni@mcmaster.ca

Abstract

We present an approach to rank order new programs in ways that accommodate uncertainty of different outcomes occurring, based on the size and nature ('bad' or 'good') of those outcomes. This represents an improvement on the way uncertainty has been accommodated in existing approaches (e.g., threshold approach to CEA). We illustrate the approach using the decision-making plane (DMP), which explicitly incorporates opportunity costs and relaxes the assumptions of perfect divisibility and constant returns to scale of the cost-effectiveness plane. The nature of the bad (or good) outcome is determined by the quadrant that it falls into (i.e., a 'quadrant effect') and it magnitude by its location within the quadrant (i.e., 'within quadrant effect'). By explicitly defining the loss function the process of accepting (or rejecting) a new program becomes transparent. We illustrate the approach using two different loss functions. We show that by recognizing that not all bad (or good) outcomes are equal and the choice of the loss function can result in different ranking of resource allocation options Further implications of the proposed approach are discussed.

1. Introduction

In today's economic climate developing methods to determine whether to adopt a new intervention, which is typically more effective and more expensive than existing interventions, is a priority for decision-makers. It stems from the realization that health care resources are limited and are not sufficient to support everything that might be done (i.e., the concept of scarcity). Hence, healthcare systems are forced into making choices between different uses of these resources. The discipline of economics, and in particular in its application to healthcare (i.e., health economics) has been advocated as a mode of thinking about resource allocation in health. Economic evaluation of healthcare programs involves 'ensuring that the value of what is gained from an activity outweighs the value of what has to be sacrificed' (Williams, 1983). Hence it relates directly to the concept of opportunity costs.

Cost- Effectiveness Analysis (CEA) has been presented as a method to help decision makers allocate available resources in a way that maximizes the health benefits produced (e.g., Weinstein and Stason, 1977; Gold et al, 1996) and is the most widely used analytic framework for comparing healthcare interventions from an economics perspective. The analytic tool of CEA is the incremental cost effectiveness ratio (ICER), which is compared with a threshold ICER (λ). A new program, which is more effective and more costly (the typical case), but with an ICER less than λ is deemed to be cost-effective and hence recommended for adoption because the average price per unit of health gain produced by the program, as indicated by the ICER, is 'acceptable'. In the past two decades it has been recognized that both costs and effects are stochastic, and hence whether the ICER is actually less than λ cannot be determined with certainty. Net health benefit (NHB) and cost-effectiveness acceptability curves (CEAC) have been presented as methods to determine the likelihood that the ICER is less than λ (The 'acceptable' average price per unit of health gain) (Jakubczyk and Kainski, 2010). These methods however do not define what is an acceptable level of risk, which is characterized by the probability that ICER is less than λ. This is left to the interpretation of the decision maker. It is also assumed that all cases where the ICER is lower than λ are seen as of "equal value" by the decision maker. In other words, the only "threat" that the uncertainty of the outcome presents is whether the ICER is greater or lower than λ.

It has been shown that only under strong assumptions of perfect divisibility and constant returns to scale of all programs does the ICER approach identify interventions associated with an efficient use of resources (Weinstein and Zeckhauser, 1973; Birch and Gafni, 1992, 1993). Under these assumptions, λ represents the shadow price of the budget constraint or the opportunity cost of resources in the margin. As a result, λ is a function of, *inter alia*, the size of the budget, and is equal to the ICER of the last program adopted, where programs are adopted in line with their ICER values, with lower ICER programs having higher priority (Gafni and Birch, 2006). Because the information on the ICERs of all programs is incomplete, however, λ cannot be determined from the information available to decision makers (Gafni and Birch, 2006). This has not prevented researchers from claiming the 'cost-effectiveness' of a new program based on an arbitrary determination of the value of λ. But this implies the availability of an indeterminate stream of additional resources at a constant marginal opportunity cost to fund all new programs with ICER<λ (Birch and Gafni, 1992, 1993; Sendi and Briggs, 2001, Sendi et al, 2002). This is inconsistent with the concept of opportunity cost on which economic evaluation is based (Williams, 1983).

A proposed alternative approach, which accommodates the concept of opportunity cost, has been presented in the literature (Birch and Gafni, 1992; Gafni and Birch, 1993) and extended to account for uncertainty (Sendi, Gafni and Birch,2002). Sendi et al (2002) used the 'decision making (DM) plane' which replaces the CE plane of the ICER approach and incorporates information on opportunity costs. This approach relaxes the assumptions of perfect divisibility and constant returns to scale. Neither does it assume that there is an indeterminate stream of additional resources, available at a constant marginal opportunity cost, to fund all new programs with ICER<λ. This approach explicitly identifies the source of the additional resource requirements of the new programme and bases the recommendation regarding the adoption of the programme of a direct comparison of benefits produced with benefits forgone.

Like any situation that involves uncertainty outcomes cannot be guaranteed. There is a non-zero probability that 'bad' outcomes can occur (e.g., that the adoption of the program will result in a net loss in health benefits). Hence we must ask whether the risk of 'bad outcomes' is acceptable to decision makers. Sendi et al (2002) suggested that "one possibility would be to limit this risk to an arbitrary level, say 0.05, similar to the arbitrary decision of accepting a 5% Type I error in hypothesis testing". This suggestion assumes that (i) this level of risk is acceptable to the decision maker and (ii) that all 'bad outcomes' are equal. These are assumptions similar to those used in uncertainty analysis (e.g., CEAC). While these approaches can handle different levels of risks (i.e., a change in the first assumption) they cannot accommodate the case where the decision maker does not view all 'bad outcomes' as equal.

In this paper we develop an alternative approach to rank order different ways of funding a new program, taking into account the risk that 'bad outcomes' can occur and the nature and magnitude of these 'bad outcomes'. The approach involves (a) identifying a loss function that does not require that all 'bad outcomes' (or 'good outcomes') are assumed to be equal and (b) defining how the expected loss will be taken into account. In this way the process of accepting (or rejecting) a new program becomes transparent. In the next section we present the decision rule and the 'DM plane' for analyzing the decision problem and introduce the concept of a loss function to determine the acceptability of potential changes in resource allocations. We illustrate the use of the approach using two possible loss functions in a numerical example. The implications of the proposed approach are discussed in the final section.

2. The Decision Rule and the Incorporation of a Loss Function

The decision rule used here is the modified approach developed by Birch and Gafni (1992). This modified approach requires that the additional health improvements of the proposed new program are compared with the health improvements produced by an *identifiable* existing program (or combination of programs) that must be forgone to generate sufficient funds to support the new program. Only if the additional health improvements of the proposed program exceed the forgone health improvements of the existing program(s) does the new program represent an improvement in efficiency. Note that this approach does not require any assumptions of perfect divisibility and constant returns.

Using the notation from Sendi et al (2002) and limiting the presentation to the case of identifying one program, B, instead of a set of programs, to be cancelled to free up the additional resources required by the new program, A, we need to find a program B such that:

$$\Delta C(A) \leq \Delta C(B)$$

and

$$\Delta E(B) < \Delta E(A).$$

where $\Delta C(A) = C_A - C_a$, the incremental cost of A, where 'A' is the new program and 'a' denotes the treatment that was provided before A (i.e., the new program) was introduced. $\Delta C(B) = C_B - C_b$ is the incremental savings (or resources released) by cancelling B, where B is the currently funded program and 'b' denotes the treatment that would be provided to those patients currently receiving B should B be cancelled. Similarly, $\Delta E(A) = E_A - E_a$ is the incremental effectiveness gained from A as compared to the current program a, $\Delta E(B) = E_B - E_b$ is the incremental effect forgone by cancelling B, leaving these patients to be treated by program b.

The problem is illustrated using the 'decision making (DM) plane' (Sendi et al, 2002), a graphical framework (Figure 1) that replaces the CE plane of the ICER approach and incorporates information on opportunity costs. The vertical axis represents the net impact on total costs (i.e., $\Delta C(A) - \Delta C(B)$), while the horizontal axis represents the net impact on total health gains (i.e., $\Delta E(A) - \Delta E(B)$). Thus the DM plane, in our example, includes information on four (or more) programs and hence describes the impact on the healthcare system by taking into account the opportunity cost of the more costly new program specifically by identifying the changes required in the system to support it. The four quadrants represent four types of potential outcomes. The southeast quadrant (I) is consistent with a more efficient allocation of resources (i.e., cancelling B to support A results in a net gain in total health effects and net reduction in total costs). The southwest quadrant (II) reflects situations where both total health effects and total costs are reduced. In the northwest quadrant (III) there is an increase in total costs and a reduction in total health effects. Finally, in the northeast quadrant (IV) there is an increase in both total health effects and total costs. Any evaluation of a program yielding results that fall completely into the southeast quadrant (including the horizontal and vertical axes with the exception of the origin) satisfies the efficiency improvement algorithm described earlier.

While the goal is to be in the southeast quadrant (I), where there is an unambiguous improvement in efficiency, this is not guaranteed because of uncertainty in costs and outcomes of the programs involved. Any other potential outcome (i.e., ending in quadrants II, III or IV) represents having also a 'bad outcome component'. Yet, not all outcomes are likely to be seen as "equally bad" (or "good"). Their relative undesirability will depend on (i) in which quadrant they are located (i.e., the ' quadrant effect') and (ii) their locations within the quadrant (i.e., 'within quadrant effect'). For example, adopting a new program that falls in quadrant III, representing an increase in total costs and a reduction in net health gains (i.e., two 'bad outcome components), is likely to be seen as worse than those falling in quadrants II or IV (where there is only one 'bad outcome component'; there is a reduction in net health effects (quadrant II) or an increase in total costs (quadrant IV). These represent a lose- lose situation versus a lose – gain (or gain – lose) situations in the other two quadrants. Also the specific location of a potential realization within the quadrant will determine how 'bad' this outcome would be viewed. For example, being close to the origin of quadrant III represents a small increase in total costs and small reduction (or loss) in net health effects. Being far from the origin of quadrant III represents a bigger increase in total costs and larger reduction in net health gains. This would have different implications for the decision-maker. It is also reasonable to assume that both the likelihood of a 'bad outcome' occurring and its magnitude will affect the decision whether to adopt a new program or not. Thus in addition to assessing 'how bad' a potential outcome may be, it is also important to know how likely it is to happen.

An approach to incorporate both the likelihood of a bad outcome occurring and its magnitude into the decision making process is the use of a loss function. A *loss function* maps an event onto a real number representing the "loss" associated with it. The "loss" due to bad outcomes can be quantified using, for example, the decision maker's utility function, providing transparency in the way the loss from potential bad outcomes is valued and hence supporting decision-maker accountability (Williams and Cookson, 2000).

In the next section we illustrate the use of a loss function and two approaches to take into account the potential losses (and gains). The first one builds on Sendi et al (2002) requiring that for every potential program for cancellation (e.g., program B), using the decision rule above, at least X% (e.g., 80%) of the potential results will fall in quadrant I (quadrant of unambiguous improvement in efficiency). Amongst all programs that meet this constraint, the one with the smallest expected loss will be chosen. Note that this approach takes into account only bad outcomes. The second approach expands this approach to incorporate both bad and good results given that not all good results produce the same value. For example, bigger gains in total health are better ceteris paribus. Hence in this approach we also attribute values to the potential gains (i.e., using a gain function) and calculate the net gain (i.e., the total expected gain minus the total expected loss). The decision maker might still impose the constraint described above (as in the first approach). In this case, among all programs that meet this requirement, the one with the largest expected net gain will be chosen.

3. Numerical illustration

The first step is to determine the loss and gain functions. In this numerical example additive loss and gain functions f and g (respectively) were assumed, defined in each of the quadrants of the decision making plane (Figure 1) as follows:

Quadrant	Loss	Gain
I	$f_I = 0$	$g_I = \|y\|^{\alpha 2} + \|x\|^{\alpha 1}$
II	$f_{II} = \|x\|^{\alpha 1}$	$g_{II} = \|y\|^{\alpha 2}$
III	$f_{III} = \|y\|^{\alpha 2} + \|x\|^{\alpha 1}$	$g_{III} = 0$
IV	$f_{IV} = \|y\|^{\alpha 2}$	$g_{IV} = \|x\|^{\alpha 1}$

where: x = ΔE(A) – ΔE(B); y= ΔC(A) – ΔC(B); α_1 and α_2 are exponents. In general any value of the exponents can be adopted. Values > 1 imply progressively increasing rate of loss (and gain) as one moves away from a zero value for the difference in incremental costs or incremental effectiveness. In our numerical illustration we used α_1=α_2=2. The values of x and y were standardized by their respective standard deviations. This enhances comparability, and avoids, for instance, the difficulty of cost differences (measured in dollars) dominating effect differences (measured in years) simply because of the units involved.

The expected loss was calculated for a variety of scenarios. We assumed that the estimates of mean incremental cost (y) and mean incremental effectiveness (x) have a bivariate normal distribution, with means μ_y and μ_x, standard deviations σ_y and σ_x, respectively, and correlation ρ. The bivariate normal distribution is the obvious choice to make because, by virtue of the Central Limit Theorem, the estimates of means for x and y will have approximately normal distributions in sufficiently large samples, even if the distributions of the original x and y values are not normal.

The overall expected loss (L) and net gain (G) were calculated by integrating the loss function or the difference in gain and loss functions over the bivariate normal distribution *h*. For example the net gain (G) was calculated as follows:

$$G = \int_{-\infty}^{0}\int_{0}^{\infty}(g_I - f_I)\cdot h(x,y)dxdy + \int_{-\infty}^{0}\int_{-\infty}^{0}(g_{II} - f_{II})\cdot h(x,y)dxdy$$
$$+ \int_{0}^{\infty}\int_{-\infty}^{0}(g_{III} - f_{III})\cdot h(x,y)dxdy + \int_{0}^{\infty}\int_{0}^{\infty}(g_{IV} - f_{IV})\cdot h(x,y)dxdy$$

where

$$h(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - \frac{2(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]/[2(1-\rho^2)]\right]$$

is the bivariate normal probability distribution function for x and *y.* The expected gain G was calculated for various scenarios by altering the values of $\mu_{x,}, \mu_y, \sigma_x, \sigma_y$, and *ρ.* We only considered scenarios where approximately (80%) of the distribution *h* was in quadrant I. For each scenario we evaluated the loss and gain functions over the different quadrants of the DM plane.

Table 1 provides examples of the total expected losses and net gains for different programs for potential cancellation (i.e., B programs as defined above) in order to free up enough resources to implement a new program (i.e., A program as defined above). We provide 13 cases (or scenarios). The idea was to see how, compared to a base case, relative changes in $\mu_{x,}, \mu_y, \sigma_x, \sigma_y,$ and ρ will affect the expected loss and net gain. Case 1 is the base case. In cases 2 to7 we increased the value of μ_x and reduce the value of μ_y compared to the base case. In cases 8 to 10 we did the opposite (i.e., reduced the value of μ_x and increased the value of μ_y). In case 11 we changed the values of the standard deviation compared to the base case. For each case we provide the values of $\mu_{x,}, \mu_y, \sigma_x, \sigma_y,$ and ρ. Given the values of these parameters, we can calculate the probability of being in each quadrant of the decision making plane as well as the expected total loss and the expected net gain.

We note that having chosen additive loss and gain functions, the gain/loss is functionally independent of the correlation of the costs and effects (i.e., ρ) and this result does not rely on the normality assumption that we have adopted; furthermore this result would remain true for any choice of the joint distribution of *x and y* in the decision-making plane, and for any choice of the loss and gain functions. In other words, when the gains and losses are separable into additive components that are each functions of the costs and effects, the overall gain and loss are independent of the correlation between the costs and effects, regardless of the form of their joint distribution, and regardless of which gain and loss functions are adopted. We kept cases 12 and 13 in Table 1 to illustrate the case of total loss and net gain being independent of the correlation in the specific instance of a bivariate normal distribution of costs and effects, and the absolute quadratic gain and loss functions, as was used in our other numerical examples..

The results in Table 1 illustrate that changes in the values of $\mu_{x,}, \mu_y, \sigma_x, \sigma_y,$ affect the values of the total loss and net gain. While we have chosen to present cases where the probability of being in quadrant I (i.e., the desired quadrant) is approximately 80%, the probability of being in the other quadrants changes in each case. This has an obvious impact on the total loss and net gain values. We also tried to see if any pattern emerge, and thus in most of our examples (i.e., cases) we changed only μ_x and μ_y and kept the other parameters constant. Comparing cases 1 and 2-10 shows that changes in μ_x and μ_y resulted in changes in the total loss and net gain in opposite directions (i.e., when comparing cases 2-7 and 8-10). It also seems from the examples in Table 1 that when the net loss is lower the net gain is higher and thus choosing one or the other approach might not matter. However this is not true. For example, comparing cases 9 and 10 we see that the expected loss is greater in case 10 but here the expected net gain is also bigger. Hence, the two criteria for decision making (i.e., expected loss and expected net gain) do not always result in the same ranking. In what circumstances these two criteria will result in the same ranking is an interesting question, but one which is beyond the scope of this paper. Finally, it is interesting to note that when comparing cases 1, 12 and 13, while the total loss and the net gain are not affected by ρ (as explained earlier) the probability of being in quadrant I (an unambiguous improvement) varies (0.807; 0827; 0.798 respectively). If that probability is an important consideration to the decision-maker, then the value of ρ will be important in determining the acceptability of each case.

4. Discussion

In this paper we have highlighted the problem of a decision maker who faces a non-zero probability that 'bad' outcomes can occur, but where not all 'bad' outcomes are regarded as the same. Hence alternatives for funding new programs need to be ranked to take into account the different probabilities of 'bad outcomes' and the nature and magnitude of these outcomes. In this paper we suggest an approach that allows the decision maker to explicitly and transparently take into account differences in bad outcomes in the context of decisions about resource allocations. This follows the decision rule, suggested by Birch and Gafni (1992, 1993), which is consistent with the concept of opportunity costs without the need for unrealistic assumptions. To the best of our knowledge this has not been considered previously in the literature.

There can be various reasons why different bad outcomes are valued differently. One reason is equity considerations. For example, bad outcomes can occur to different individuals or population groups. However, as explained elsewhere (Gafni and Birch, 2006), when performing an economic evaluation of health care interventions, the preferred equity position must be incorporated into the objective function (e.g., to 'favour' outcomes accruing to one particular individual or social groups) or incorporated as an additional constraint (e.g., to ensure equal availability of services, irrespective of outcome) in order to ensure that the equity goals are being pursued efficiently. Another reason can be simply a 'quantity effect'. For example, Mehrez and Gafni (1990) discuss the case where a plane crash causing 200 deaths has a bigger effect in the media than 200 people dying in different car accidents, 'one at a time' over a week. Quantity effects are captured in our approach in the "within quadrant effect" described earlier. However, we are not dealing with the question why different bad (or good) outcomes are valued differently; our method simply provides a mechanism for incorporating such differences into the decision making process about resource allocation.

Our numerical examples illustrate the importance of taking into account not all bad (and good) outcomes are equal. In all the cases presented in Table 1 the probability of being in quadrant I (i.e., the desired quadrant) is approximately 80%. Hence if all the bad (good) outcomes were valued equally a decision maker who would choose the alternative that maximizes the probability of being in quadrant I will be indifferent between these options. Yet, when all bad and good outcomes are not valued the same, there is a clear ranking which depends on the loss/gain functions used.

To determine whether a decision maker values all bad and good outcomes the same is a simple task. One can simply describe several potential outcomes in the decision making plane, which differ with respect to their cost and effect consequences, and ask the decision maker which one is preferred. Determining the loss/gain function of an individual decision maker or of society is a more complex task. While the complexity associated with allocation of healthcare resources is acknowledged, this is not a reason to avoid being transparent about how this complexity is accommodated in specific ways. Only through such transparency can decision-makers be held accountable for the use of healthcare resources.

Figure 1: The Decision Making Plane

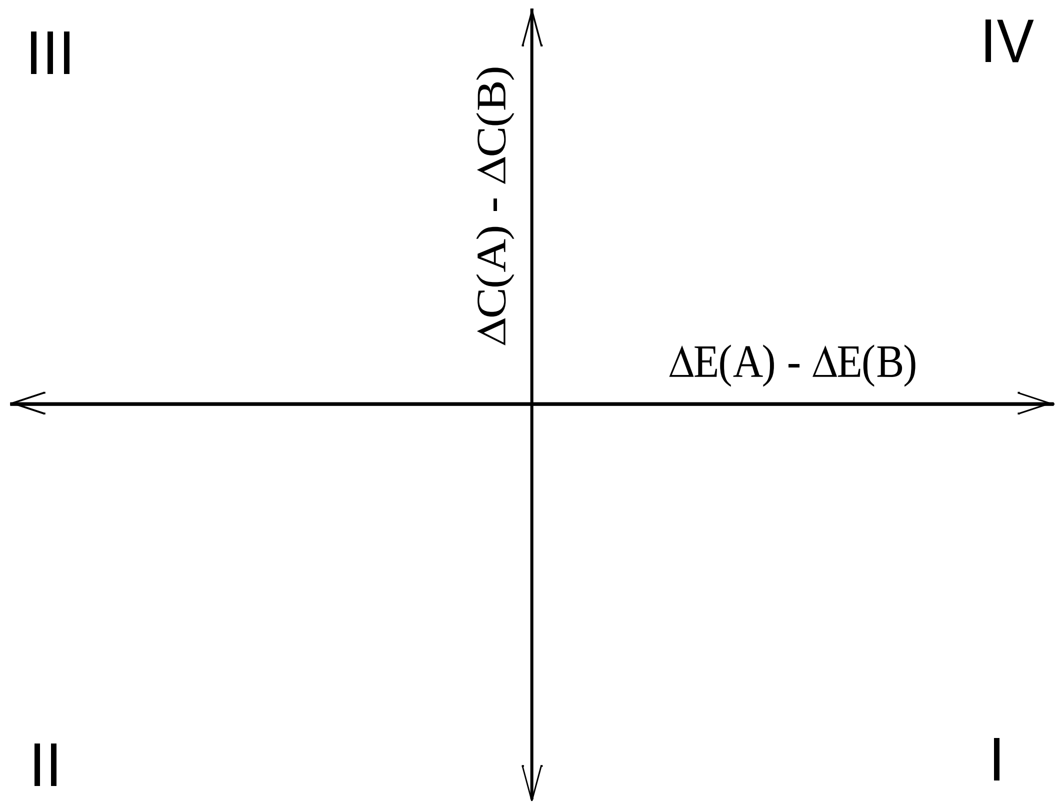

Table 1: Examples of the total expected losses and net gains associated with different programs that can be cancelled to free up resources to implement a new program

Case #	μ_x	μ_y	σ_x	σ_y	ρ	$P(Q_1)$	$P(Q_4)$	$P(Q_3)$	$P(Q_2)$	*Total Loss*	**Total Net Gain**
1*	2.2	-1.5	2	1	0	0.807	0.058	0.009	0.127	0.264	11.563
2	2.5	-1.25	2	1	0	0.800	0.094	0.011	0.094	0.212	12.388
3	2.75	-1.15	2	1	0	0.801	0.074	0.011	0.114	0.179	13.527
4	3.0	-1.1	2	1	0	0.807	0.127	0.009	0.058	0.152	14.907
5	3.25	-1.01	2	1	0	0.800	0.044	0.008	0.148	0.140	16.303
6	3.5	-0.97	2	1	0	0.800	0.033	0.007	0.159	0.128	17.937
7	4.0	-0.92	2	1	0	0.803	0.019	0.004	0.175	0.113	21.621
8	2.0	-1.7	2	1	0	0.804	0.152	0.007	0.037	0.315	11.260
9	1.75	-2.2	2	1	0	0.798	0.188	0.003	0.011	0.398	12.106
10	1.7	-2.7	2	1	0	0.800	0.197	0.001	0.003	0.417	14.345
11	2.2	-1.5	1	-1.7	0	0.800	0.186	0.003	0.011	0.284	10.411
12	2.2	-1.5	2	1	-0.5	0.827	0.037	0.029	0.106	0.264	11.563
13	2.2	-1.5	2	1	0.5	0.798	0.066	0.001	0.135	0.264	11.563

Where: x = ΔE(A) – ΔE(B); y= ΔC(A) – ΔC(B); 1* = base case

References

Birch S, Gafni A. 1992. Cost effectiveness/utility analysis: do current decision rules lead us to where we want to be? *Journal of Health Economics* 11:279-296.

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

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

Gold MR, Siegel JE, Russel LB, Weinstein MC. 1966. *Cost-effectiveness in Health and Medicine.* Oxford University Press: New York.

Jakubczyk M, Kaminski B. 2010. Cost-effectiveness acceptability curves – caveats quantified. *Health Economics* 19:955-963.

Mehrez A, Gafni A. 1990. Resource allocation, equity and public risk: dying one at a time versus all together. *Socio-Economic Planning Sciences* 24:285-294.

Sendi P, Briggs A: 2001. Affordability and cost-effectiveness: decision-making on the cost-effectiveness plane. *Health Economics* 10:675-680.

Sendi P, Gafni A, Birch S. 2002. Opportunity costs and uncertainty in the economic evaluation of health care interventions. *Health Economics* 11:23-31.

Weinstein MC, Zeckhauser R. 1973. Critical ratios and efficient allocation. *Journal of Public Economics* 2:147-157.

Weinstein MC, Stason WB. 1977. Foundation of cost effectiveness analysis for health and medical practices. *New England Journal of Medicine* 296:716-721.

Williams A. 1983. The economic role of 'health indicators'. In *Measuring the Social Benefits of Medicine,* Teeling SE (ed), Office of Health Economics: London, 63-67.