HEALTH ECONOMICS
ECONOMIC EVALUATION
Health Econ. **14**: 655–667 (2005)
Published online 27 January 2005 in Wiley InterScience (www.interscience.wiley.com). **DOI:**10.1002/hec.973

Optimal allocation of resources over health care programmes: dealing with decreasing marginal utility and uncertainty

Maiwenn J. Al[a,*], Talitha L. Feenstra[b] and Ben A. van Hout[c]
[a] *Institute for Medical Technology Assessment, Erasmus Medical Center, Rotterdam, The Netherlands*
[b] *Department for Health Services Research CZO, National Institute of Public Health and the Environment, The Netherlands*
[c] *Julius Center for Patient Oriented Research, Utrecht Medical Center, The Netherlands*

Summary

This paper addresses the problem of how to value health care programmes with different ratios of costs to effects, specifically when taking into account that these costs and effects are uncertain. First, the traditional framework of maximising health effects with a given health care budget is extended to a flexible budget using a value function over money and health effects. Second, uncertainty surrounding costs and effects is included in the model using expected utility. Other approaches to uncertainty that do not specify a utility function are discussed and it is argued that these also include implicit notions about risk attitude. Copyright © 2005 John Wiley & Sons, Ltd.

Keywords economic evaluation; budget-restrictions; uncertainty; allocation; decision rules

Introduction

Cost-effectiveness analyses often conclude by stating that the obtained cost-effectiveness ratio is well within the acceptable limit and it is thus advisable to implement the therapy under investigation on a wide scale. An example can be found in the Dutch guidelines on cholesterol-lowering therapy. Supported by a cost-effectiveness analysis, the guidelines advise to treat all patients with a 22% or higher 10-year risk for a cardiovascular event. This type of advice is problematic. Although it has been estimated that for such a strategy the costs per life-year gained are less than EUR 20 000, no indication is given where the budget for treatment may be found. Given the number of untreated individuals in the Netherlands, the additional costs are estimated at EUR 100 million, a considerable amount. Thus, the question arises whether budget should be allocated to the new therapy. A related problem concerns the uncertainties surrounding this cost-effectiveness ratio. A point estimate of EUR 20 000 may be deemed acceptable in a league table, but this may not hold true for the uncertainty margins. Fear of being confronted with a higher than expected cost-effectiveness ratio when implementing this strategy may induce the decision maker to value the cost-effectiveness lower than the certain outcome. More generally, the question arises how the budget should be allocated when programmes differ not only in terms of their cost-effectiveness but also in terms of the uncertainty surrounding both costs and effects. These questions are addressed here using a formal mathematical framework.

This framework collapses the intricate process of priority setting in health care into a simple optimization problem, with a single societal

*Correspondence to: IMTA, Erasmus MC, PO Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: Al@bmg.eur.nl

Received 22 April 2003
Accepted 24 August 2004

decision maker setting priorities over a set of health care programmes with given costs and health-effects. Of course, in reality single societal decision makers do not exist and neither do all programmes have clear-cut measurable health effects and costs, nor is the aggregation of health-effects over individuals or programmes a simple matter of adding up. However, in this paper, we want to focus on the underpinnings of cost-effectiveness so the metaphor of a single decision maker turns out to be a very helpful tool. It has been used by a number of authors to analyse the underlying rules of cost-effectiveness [1,2]. We extend this line of work by addressing the possibilities of new programmes entering the set of options and of uncertainty on programme costs and effects. In the discussion section, the questions whose preferences are and should be reflected in the optimization problems and how these might be found are shortly discussed.

Next, the mathematical rules of cost-effectiveness as introduced by Weinstein and Zeckhauser involving maximisation of health effects for a given total budget are summarized [1,3]. As we will see, these rules (that do not account for uncertainty) imply that the required budget for a new cost-effective therapy has to be taken from the existing less cost-effective treatments in the health care sector. A pragmatic alternative to the standard model is the use of a fixed trade-off between costs and health gains. However, as argued by Gafni and Birch, this approach would lead to an ever-increasing health care budget [2].

Of course, in practice both the total budget and the fixed critical ratio will be chosen with at least some reflection about their consequences. They will have to be adjusted regularly to reflect changes in the set of programmes available, population needs and budgetary restrictions. If not, over time the critical ratio resulting from the fixed budget approach and the fixed critical ratio will diverge. The adjustment to current circumstances involves the implicit weighting of the different interests at stake.

Following this, a rule similar to the one derived by Weinstein and Zeckhauser is explored (still assuming certainty) which can be derived to support a flexible trade-off between budgets. Thus the implicit weighting involved in the adjustment of budgets or thresholds is made explicit. The model clarifies the trade-off between more health care programmes and increases in the health care budget at the cost of money available for other purposes.

Then, an approach including uncertainty is presented that combines the results discussed previously with the theory of expected utility. This approach requires the explicit formulation of a utility function, expressing risk attitude regarding money and health effects. Based on the ideas from the previous section we discuss how to handle uncertainty without explicitly formulating such a utility function. Implied notions about risk attitude are clarified. Finally, we discuss our results and compare them with other recent work on uncertainty and budget allocation [4–6].

The classical approach

The classical approach to support thinking in terms of cost-effectiveness ratios was introduced in 1973 by Weinstein and Zeckhauser, who addressed the problem of maximising health gains within a given budget mathematically [1]. They assumed (as we do) independence of programmes, constant returns to scale, perfect divisibility and a finite, fixed size of each programme. Throughout this paper, we will assume that the baseline health care has already been funded and that the programmes under review represent policy changes to this baseline. Furthermore, we assume that in the optimum (for each approach) some available programmes are not (completely) implemented. Conditions can be adjusted to take mutually exclusive or indivisible programmes into account [7–9]. Finally, the problems presented here, like most applications of cost benefit analysis, unavoidably follow a partial equilibrium approach [10]. That is, the costs and effects of each programme are assumed to be independent of the allocation decisions that are considered.

Denote, for n health care programmes, the total costs of each programme i by c_i and its health effects by e_i. Let d_i, $0 \leq d_i \leq 1$, represent the fraction of the ith programme undertaken and B the fixed total budget. The budget allocation problem can be formalised as

$$\begin{array}{ll} \max_{d_i} & \sum_{i=1}^{n} d_i e_i \\ \text{subject to} & \sum_{i=1}^{n} d_i c_i \leq B \end{array}$$

The optimal solution is characterised by

$$\frac{c_i}{e_i} < \alpha \rightarrow d_i^* = 1$$
$$\frac{c_i}{e_i} > \alpha \rightarrow d_i^* = 0$$
$$\frac{c_i}{e_i} = \alpha \rightarrow d_i^* = \pi \quad (1)$$

Here α and π have to be chosen such that the budget constraint is met with equality. The parameter α can be interpreted as the maximum cost-effectiveness accepted within the given budget. In this situation, the decision rule is to undertake programmes with a lower ratio of costs to effects than α, while programmes with a higher ratio are not undertaken. Note that if the budget increases, all else remaining equal, the critical ratio α increases, which means that more programmes become acceptable. For very high critical ratios, the budget constraint is not binding and all available programmes can be financed.

Suppose a new programme is introduced with a ratio of costs to effects $c_{\mathrm{new}}/e_{\mathrm{new}}$ smaller than the current critical ratio α. A new optimal allocation of the budget must be found, whereby some of the programmes currently implemented are now (partially) abandoned, because the budget is needed to finance the new programme. The new critical ratio will be smaller than or equal to α. Thus, for a fixed budget, the greater the number of existing efficient programmes, the lower the critical ratio of costs to effects, i.e. the higher the requirements for new programmes to enter.

In practice, the dependency between the critical ratio and the available budget is often neglected, and the critical ratio is assumed to be fixed at some value γ [11–13]. According to this approach the allocation of budget over programmes is given by the following conditions:

$$\frac{c_i}{e_i} < \gamma \rightarrow d_i^* = 1$$
$$\frac{c_i}{e_i} > \gamma \rightarrow d_i^* = 0$$
$$\frac{c_i}{e_i} = \gamma \rightarrow 0 \leq d_i^* \leq 1 \quad (2)$$

In that case, the introduction of a new intervention has no effect on the critical ratio of costs to effects. If $c_{\mathrm{new}}/e_{\mathrm{new}} < \gamma$, a new allocation with a larger amount of money spent on health care and more health effects is obtained. No current programmes are reduced. Note that this would in theory lead to unlimited growth of the health care budget if new, more efficient programmes continue to become available (as also noted by Gafni and Birch [2]).

Examples

As a theoretical example, consider the four programmes in Table 1, and assume that a budget of EUR 750 000 is available.

Assume that initially programmes 1, 2 and 3 are available. The top half of Table 2 shows the optimal budget allocation for the fixed budget approach and the fixed ratio approach. When the fixed ratio approach is used with a critical ratio of 250, the decision maker is indifferent to all fractions of programme 2. As soon as a marginally larger critical ratio is chosen all programmes will be implemented, while a marginally smaller critical ratio will lead to the implementation of programmes 1 and 3 only. Clearly, total costs and total effects will vary according to the fraction chosen for programme 2.

Now suppose that programme 4 is introduced. The new optima are presented in the bottom half of Table 2. For a fixed budget, total health effects increase by shifting budget from programmes 2 and 3 to the new programme 4, and a new, lower, critical ratio results. For a fixed critical ratio of 250, programme 4 is fully implemented and, again, the decision maker is indifferent to all fractions of programme 2. The total costs are now higher than the original budget. This is, however, combined with higher health effects.

As a practical example, consider the introduction of cholesterol-lowering therapy on a large scale. Suppose that the health care budget in the Netherlands was sufficient to fund all programmes up to a limit of EUR 20 000/life-year gained. For a fixed budget, programmes have to be cancelled to free the budget needed for cholesterol-lowering therapy until everything fits the total budget. Moreover, a new critical ratio results, for example EUR 19 000/life-year gained. This example raises a

Table 1. Costs and health effects for four fictitious health care programmes

Programme	Total costs	Total health effects	Ratio costs to effects
1	400 000	1900	211
2	450 000	1800	250
3	300 000	1300	231
4	225 000	1000	225

Table 2. Budget allocation over three or four fictitious health care programmes, for both the fixed budget approach and the fixed ratio approach

Approach	Optimum	Total costs	Total health effects	Critical ratio
Fixed budget	$d_1^* = 1,\ d_2^* = 1/9,\ d_3^* = 1$	750 000	3400	250
Fixed ratio	$d_1^* = 1,\ d_2^* = 0\ \text{to}\ 1, d_3^* = 1$	700 000–1 150 000	3200–5000	250
Fixed budget	$d_1^* = 1,\ d_2^* = 0,\ d_3^* = 5/12,\ d_4^* = 1$	750 000	3442	231
Fixed ratio	$d_1^* = 1,\ d_2^* = 0\ \text{to}\ 1,\ d_3^* = 1,\ d_4^* = 1$	725 000–1 425 000	4200–6000	250

number of questions: for instance, which health care programmes are on the border, or what is the critical ratio (as derived from the total health care budget), or even what is the total budget? The last question is already difficult to answer, let alone the second and the first. This may be one of the reasons for defining an acceptable critical ratio without consideration of the interaction between this ratio and the total health care budget. When such a fixed ratio is used, the budget would have to be increased.

In summary, under the fixed budget approach costs, of course, equal the budget, while the critical ratio varies with the introduction of new health programmes. On the other hand, under the fixed ratio approach, costs may well overrun the budget.

A flexible trade-off

The standard approach starts with a given health care budget. The determination of this health care budget may be seen as the result of negotiations in which, implicitly, a comparison is made between the effects of spending money for health care versus alternative purposes, e.g. education.

One formulation of the health care budget allocation problem that allows for more flexibility than a fixed budget is to use a value function defined both over health and money as the objective of a maximisation problem. Value is derived from budget available for purposes competing with health care spendings and from health. This value function represents the decision maker's preference structure for these attributes [14]. It is assumed that the value function is characterised by decreasing marginal value from both money and from health effects.

Once a value function V is defined on budget and health effects, the goal is to find fractions d_i that maximise value, given n health care programmes with costs c_i and effects e_i. This can be written as

$$\max_{d_i}\ V(M-c,e)$$

$$\text{subject to}\ \ 0 \le d_i \le 1.$$

Here $c=\sum_i d_i c_i,\ \ e=\sum_i d_i e_i$, and M is the total budget of the decision maker. This budget M should be interpreted as the maximum additional amount of money the decision maker would be willing to spend on health. In reality, the optimum will never include spending the full budget M, since by assumption M is larger than the maximum amount that could be spent, if all programmes were fully implemented. The amount of money $M-c$ is available for purposes other than the financing of the interventions considered and is hence valued positively by the decision maker.

The optimal solution is characterised by

$$\begin{aligned}
&\frac{c_i}{e_i}<\frac{V_2(M-c,e)}{V_1(M-c,e)} \rightarrow d_i^* = 1\\
&\frac{c_i}{e_i}>\frac{V_2(M-c,e)}{V_1(M-c,e)} \rightarrow d_i^* = 0\\
&\frac{c_i}{e_i}=\frac{V_2(M-c,e)}{V_1(M-c,e)} \rightarrow 0 \le d_i^* \le 1
\end{aligned} \tag{3}$$

Here, $V_1=\partial V/\partial(M-c)$ and $V_2=\partial V/\partial e$, the marginal derivative of the value function to money and health effects, respectively. Thus, V_2/V_1 is the marginal rate of substitution of money for health. Note that decreasing marginal value implies that $V_{11}<0$ and $V_{22}<0$, with $V_{11}=\partial^2 V/\partial(M-c)^2$ and $V_{22}=\partial^2 V/\partial e^2$.

The conditions in Equation (3) are similar to the optimality conditions in consumer theory where the relative price, here the ratio of costs to effects c_i/e_i, equals the marginal rate of substitution [15]. If programme i's ratio of costs to effects is higher than the decision maker's marginal rate of substitution, it is better to spend no money on programme i since, for the decision maker, spending an additional euro on programme i gains less in terms of value for health effects obtained than the value of the additional euro and vice versa.

There is also a similarity to the conditions presented in the previous section (see Equations (1) and (2)). The marginal rate of substitution (V_2/V_1) takes the place of the critical ratio α (γ). That is, the critical ratio is defined by the value function, which is characterised by decreasing marginal values for money and health. Except for corner solutions, in the optimum, $V_2/V_1 = c_{\text{last}}/e_{\text{last}}$ or, in words, the marginal rate of substitution is equal to the largest (i.e. least favourable) ratio of costs to effects in the optimal set of programmes. From this follows a specific budget for health care.

It is interesting to note that the approach with a fixed critical ratio can be represented by a linear value function, that is, with a fixed trade-off between costs and effects. This is written as $V = (M - c) + \gamma * e$, with $V_1 = 1$ and $V_2 = \gamma$, that is, the marginal value of money and health effects is constant. Thus, the fixed critical ratio approach should only be used when such value function adequately describes the decision maker's preferences concerning money and health effects.

Consider again the introduction of a new intervention. If V_2^*/V_1^* is the marginal rate of substitution in the current optimum, and programme 'last' is the last programme to be (partially) implemented, $V_2^*/V_1^* = c_{\text{last}}/e_{\text{last}}$. Suppose that a new programme is introduced, with a ratio of costs to effects of $c_{\text{new}}/e_{\text{new}}$. In order to be at least partially implemented, this ratio must satisfy the condition $c_{\text{new}}/e_{\text{new}} \leq c_{\text{last}}/e_{\text{last}}$. Whether or not the new programme is fully implemented will depend on the size of the programme. Independent of the size of the new programme, $V_2^{\text{new}}/V_1^{\text{new}} \leq V_2^*/V_1^*$, i.e. the critical ratio never increases as a new programme becomes available. Since the budget is flexible, adjustment to the new situation takes place not only through changes in the d_i, but also through a change in the health care budget. The therapies on the border, which would have been cancelled in the standard approach, are now evaluated against additional money for other purposes.

Examples

For the theoretical example, consider again the allocation of budget over programmes 1, 2 and 3 mentioned in the previous section (see Table 1). Assume a decision maker has a non-linear value function that can be written as $V(M - c, e) = (M - c)^{0.75}e^{0.6}$. Furthermore, assume that the total budget M is 2 000 000. The optimum solution is $d_1^* = 1$, $d_2^* = 0.3$, $d_3^* = 1$, and the critical ratio is 250. In this optimum, the total costs are 835 000, with total health effects of 3740.

Compared to the example given in the previous section the critical ratio in this example is the same as for the fixed budget approach, but the fraction of programme 2 implemented and the budget are somewhat larger. The preferences stated in the value function above indicate a willingness to increase the budget in exchange for more health effects.

As a practical example, consider again a situation with a critical ratio of approximately EUR 20 000. For a value function with decreasing marginal values, the introduction of cholesterol-lowering therapy might push some therapies that are already implemented out of the health care system. However, within this framework, these programmes might also compete with other non-health care expenditures. The changes in the optimal fractions are smaller than they would have been in case of a fixed budget. Moreover, while in the fixed budget approach the critical ratio might decrease to EUR 19 000 it might now only decrease to EUR 19 500, while the budget increases with e.g. EUR 25 million.

Decisions under uncertainty using the flexible trade-off

Until now we have assumed that costs and effects for all programmes are known with certainty, without having any example for which this is true. In general, the estimates of costs and effects are surrounded by uncertainty, and the magnitude of the uncertainty may influence the valuation of a health care programme. To include such risk attitude into decision making, we will combine the results from the previous sections with the well-known concept of expected utility. The expected utility approach implies the definition of a utility function over the set of uncertain outcomes ($M-c$,e) that satisfies the expected utility property, i.e. the utility of a stochastic event is represented by the expected value of the utilities of all possible realisations of the event [14,15].

Again, $c = \sum_i d_i c_i, e = \sum_i d_i e_i$, and M is the total budget of the decision maker. Let $x_i = (c_i, e_i)$ be stochastic and assume that x_i are independent for all i.

The general goal is now to find d_i so as to maximise expected utility

$$\mathrm{Max}_{d_i}\ E[U(M - c, e)] \text{ subject to } 0 \leq d_i \leq 1$$

The first order conditions for an interior optimum can now be derived

$$E\left[\frac{\partial U}{\partial d_i}\right] = 0 \quad \text{for } 0 < d_i < 1$$

Using similar expressions for corner solutions, and defining $U_1 = \partial U / \partial (M - c)$ and $U_2 = \partial U / \partial e$ as the marginal utility of money (i.e. the budget available for purposes competing with health care spendings) and the marginal utility of health effects, respectively, we find the following characterisation of the optimum

$$\begin{aligned} &E[e_i U_2 - c_i U_1] > 0 \rightarrow d_i^* = 1 \\ &E[e_i U_2 - c_i U_1] < 0 \rightarrow d_i^* = 0 \\ &E[e_i U_2 - c_i U_1] = 0 \rightarrow 0 < d_i^* < 1 \end{aligned} \tag{4}$$

These conditions are similar to those under certainty (Equation (3)), but use expected values. When assuming a linear utility function $U = (M - c) + a^* e$ with $U_1 = 1$ and $U_2 = a$ they can be rewritten as

$$\begin{aligned} &\frac{E(c_i)}{E(e_i)} < a \rightarrow d_i^* = 1 \\ &\frac{E(c_i)}{E(e_i)} > a \rightarrow d_i^* = 0 \\ &\frac{E(c_i)}{E(e_i)} = a \rightarrow 0 < d_i^* < 1 \end{aligned} \tag{5}$$

As expected, this means that for a linear utility function one can simply compare the expected values of costs and effects to a critical ratio. However, such a function would assume risk neutrality as well as constant marginal utility of money and health effects.

When using utility functions that assume decreasing marginal utility or risk aversion, in general, $E[e_i U_2 - c_i U_1]$ cannot be simplified and the decision of whether or not a new programme will be added to an existing set of programmes cannot be made by a simple comparison of $E(c_{\mathrm{new}})/E(e_{\mathrm{new}})$ with a critical ratio. This means that no simple decision rule can be derived from the characterisation of the optimum given in Equation (4). Instead, numerical optimisation methods can be used to find the optimum. For certain combinations of utility function and probability density function it is possible to simplify Equation (4) further; an example is presented in Appendix A. In the simplified conditions, the variance and covariance appear.

An important question is whether it is reasonable to assume that decision makers are risk averse and whether they will be risk averse for health costs, health effects, or both. Claxton claims that opportunity costs of making wrong decisions based on the mean are symmetrical and thus, decisions should be based only on the net mean benefit [16]. Conversely, Arrow and Lind have shown that for public investment decisions, it is rational to be risk neutral towards costs, but when substantial benefits accrue to individuals directly, these individuals' risk attitude should be taken into account [17]. The availability of health programmes may affect a large part of an individual's utility. Thus, it may be argued that it is rational for decision makers who use the societal perspective to be risk averse regarding health effects. For other decision makers, e.g. HMO managers or insurers, it may be argued that both costs and health effects accrue privately, with different individuals receiving benefits and paying costs. Then, the appropriate procedure would be to include risk attitude regarding both costs and health effects.

Similar arguments have been put forward by Ben-Zion and Gafni and Zivin *et al.* [18–20]. For instance, Ben-Zion and Gafni point to the difficulties of risk sharing for health interventions [18]. Assuming some insurance arrangement, it is unclear if the money collected from 'successful' patients will always be sufficient to compensate the less 'successful' patients. However, they go on to propose solutions based on the idea that preferences of individuals over health and money and uncertainty in these should be reflected, while in our paper, societal preferences and risk attitudes are considered.

Example

Assume that a decision maker is constantly risk averse and has utility function

$$U(M - c, e) = k_1(1 - \exp^{-\delta(M-c)}) + k_2(1 - \exp^{-\gamma e})$$

with $k_1 = 0.49 * 1.0186; \quad \delta = 0.000002$
$k_2 = 0.51 * 1.125; \quad \gamma = 0.00022$

where δ and γ are parameters reflecting risk aversion if positive and risk proneness if negative, and k_1 and k_2 are scaling factors. For clarity, the assumed values of k_1 and k_2 are written as a product of two factors, a weighing factor, giving relative weight of costs and effects in the utility function and a scaling factor, chosen such that on the domain considered, the utility is between zero and one. The total budget M is 2 000 000. It is furthermore assumed that total costs and effects of the programmes follow normal distributions and (to simplify calculations) we will assume that they are uncorrelated. Table 3 presents the parameters of the distributions.

To see how the optimum may change for changes in the degree of uncertainty, we varied the standard deviation of the costs of programme 1 (Table 4).

Clearly, as the uncertainty about the total costs of programme 1 increases, the decision maker is less willing to spend money, which results in lower expected total costs and lower expected health effects. Also, for a sufficiently high standard deviation, programme 1 is implemented only partially, even though its ratio of costs to effects is the lowest among the three programmes. It is interesting to see that, for a standard deviation of 200 000, both programmes 1 and 2 are implemented partially. Under certainty, this will never be optimal for programmes with different ratios of costs to effects.

To conclude, the optimal budget allocation not only depends on the ratio of costs to effects and the size of the programmes, but also on the (co)variance of costs and effects. As shown in the example above, under uncertainty and risk aversion, it may be optimal that a programme with a more favourable ratio than all other programmes under consideration is implemented only partially. Under certainty, that would never be the case. This may occur if the variance of costs and effects in this programme is large relative to that of the other programmes. Being risk averse implies a willingness to give up some effects or budget in return for more certainty. Since marginal utility is not independent of total costs or effects, the first order conditions cannot be simplified in terms of ratios of expected costs and effects. Hence, the ratio of expected costs to effects alone does not inform whether a new programme should be (partially) implemented. A new optimum must be derived, taking into account expected costs and effects of this programme and all other programmes, as well as the (co)variances of costs and effects.

Table 3. Mean costs and health effects (+variances) for three fictitious health care programmes

Programme	μ_c	μ_e	σ_c	σ_e	μ_c/μ_e
1	400 000	1900	120 000	125	211
2	450 000	1800	40 000	90	250
3	300 000	1300	70 000	110	231

Decisions under uncertainty using the classical approach

The expected utility approach is general and includes most other approaches as special cases. However, this general approach requires the specification of a utility function by the decision maker who may be unwilling or unable to make such a specification. Therefore, some alternative approaches are discussed below, which are the most straightforward extensions towards uncertainty of the optimisation problems presented in earlier sections.

Table 4. Changes in optimum budget allocation with changes in uncertainty for programme 1

σ_c programme 1	Optimum	Total expected costs	Total expected health effects
120 000	$d_1^* = 1,\ d_2^* = 0.9,\ d_3^* = 1$	1 105 000	4820
150 000	$d_1^* = 1,\ d_2^* = 0.89,\ d_3^* = 1$	1 100 500	4802
200 000	$d_1^* = 0.97,\ d_2^* = 0.89,\ d_3^* = 1$	1 088 500	4745
250 000	$d_1^* = 0.81,\ d_2^* = 1,\ d_3^* = 1$	1 074 000	4639
350 000	$d_1^* = 0.68,\ d_2^* = 1,\ d_3^* = 1$	1 022 000	4392

The first approach is similar to the fixed budget approach, as formulated in the earlier. Instead of maximising deterministic health effects for a fixed budget, for instance expected health effects can be maximised under the constraint that the probability that the total costs will exceed the budget is smaller than a given small percentage, say, 5%. This may be written as

$$\text{maximise} \sum_{i=1}^{n} d_i E(e_i)$$

$$\text{subject to } P\left(\sum_{i=1}^{n} d_i c_i \leq B\right) \geq 0.95$$

An implicit assumption on preferences regarding additional costs and effects under uncertainty is made here. This assumption implies a large risk aversion for costs and risk neutrality for health effects. To see this, first note that only the expected health effects are used, without any reference to the uncertainty around the estimate. Second, assuming that total costs are normally distributed (this assumption is based on the Central Limit Theorem), total costs are lower than expected costs with a 50% probability. Under risk neutrality the probability that total costs remain within the budget constraint equals 50%. The more risk averse regarding costs, the higher the required probability that total costs remain within the budget constraint, with an – unattainable – extreme of 100%. Within this range, 95% represents a high risk aversion regarding costs.

However, as we have discussed in the previous section, it will rarely be rational for a decision maker to be risk neutral towards health effects and risk averse towards costs, it is more likely that the reverse is true. For such a decision maker, who is risk neutral regarding costs and risk averse regarding health effects, the following approach may be more relevant. The decision maker now minimises expected costs under the constraint that the probability that total effects exceed some aspiration level L is at least, say, 95%. This can be written as

$$\text{minimise} \sum_{i=1}^{n} d_i E(c_i)$$

$$\text{subject to } P\left(\sum_{i=1}^{n} d_i e_i \geq L\right) \geq 0.95$$

If (e_i, c_i) have a bivariate normal distribution with mean costs μ_{c_i}, mean effects μ_{e_i}, variance in costs $\sigma_{c_i}^2$, and variance in effects $\sigma_{e_i}^2$, then this problem is equivalent to

$$\text{minimise} \sum_{i=1}^{n} d_i \mu_{c_i} \text{ subject to } \mu^e - z_{0.95}\sigma^e \geq L$$

where $z_{0.95}$ denotes the 95th percentile of the standard normal distribution,

$$\sigma^e = \sqrt{\sum\nolimits_{i=1}^{n} d_i^2 \sigma_{e_i}^2} \quad \text{and} \quad \mu^e = \sum_{i=1}^{n} d_i \mu_{e_i}$$

The first order conditions for optimality are given by

$$\frac{\mu_{e_i}}{\mu_{c_i}} + \frac{2z_{0.95}\sigma_{e_i}}{\mu_{c_i}} \geq \frac{1}{\lambda} \rightarrow d_i = 1$$

$$\frac{\mu_{e_i}}{\mu_{c_i}} < \frac{1}{\lambda} \rightarrow d_i = 0$$

$$\text{else} \rightarrow 0 < d_i < 1$$

Both approaches can also be written as the maximisation of a specific expected utility function (see Appendix B). Thus, they are a special case of the expected utility formalisation. The critical ratio $1/\lambda$ is found as the shadow price belonging to the constraint that $\mu^e - z_{0.95}\sigma^e \geq L$. If the effects are more uncertain, this constraint becomes stricter. The result is a critical ratio that, for the same aspiration level, is higher for larger uncertainty. Note that in this case, a higher critical ratio implies that fewer programmes are implemented because the first order conditions are now stated in terms of ratios of expected effects to costs.

Example

Using the information from Table 3, and assuming an aspiration level for health effects of 3000, this approach leads to the following optimum: $d_1^* = 1.0$, $d_2^* = 0.70$, $d_3^* = 0.056$. In this optimum, the expected total costs are 732 836 and the expected total effects are 3233. If instead we disregard uncertainty about total health effects and simply minimise expected costs, given that the expected effect are above the aspiration level, the optimum is $d_1^* = 1.0$, $d_2^* = 0.0$, $d_3^* = 0.846$.

If we compare these two results, we see that the standard deviations of the effects of programmes 2 and 3 are given more weight than the ratio of costs to effects, indicating a strong risk aversion.

This becomes even clearer if the aspiration level is increased to 4000. Using the risk averse approach, the optimum is now $d_1^* = 0.937$, $d_2^* = 0.965$, $d_3^* = 0.577$, with expected total health effects of 4267. Under risk neutrality, the optimum is $d_1^* = 1.0$, $d_2^* = 0.444$, $d_3^* = 1.0$. Apparently, at this aspiration level, even programme 1, with the lowest ratio of costs to effects, is partially abandoned in favour of programme 2, with the highest ratio of costs to effects, in order to decrease the uncertainty about the total health effects. It is also interesting to see that now all programmes are partially implemented. As mentioned earlier, under certainty, this will never be optimal for programmes with different ratios of costs to effects.

A third approach is implied by the calculation of confidence intervals and acceptability curves for cost-effectiveness ratios [21,22]. When these methods are used for the ratio of total costs to total effects, they can be seen as a straightforward extension towards uncertainty of the fixed trade-off formulation presented earlier. The idea is that the decision maker wants to be reasonably sure that the ratio of costs to effects of a health programme is below the critical ratio. There is a clear link with the situation under certainty, where we know for certain whether or not a ratio is smaller than the critical ratio. Similar to the fixed trade-off approach this is a pragmatic approach, which was developed without reference to an underlying optimisation problem. Analogous to the former approach, here too an implicit assumption is made about preferences concerning additional costs and effects under uncertainty. Note that, in this situation, the decision maker requires a certain level of certainty for each individual programme (instead of for the portfolio of programmes together) and that not only the variance of costs but also the variance in health effects and the covariance of costs and effects are now taken into consideration.

Here again, it is clear that on a scale from 50% certainty (reflecting risk neutrality) to 100% certainty (reflecting absolute risk aversion), requiring 95% certainty that the ratio of costs to effects is below the critical ratio reflects a very strong risk aversion.

Example

Based on Table 3, using Fieller's theorem, the following 90% two-sided confidence intervals are found for the ratios of costs to effects: programme 1 [106, 320]; programme 2 [210, 294]; programme 3 [140, 331] [21,23]. Assuming the decision maker wants 95% certainty that the ratio of costs to effects is below a certain limit, the order in which to implement programmes would be: 2, 1, 3. If, again, a threshold of 250 is used, no programme would be implemented. This reflects the very strong risk averse attitude regarding both costs and effects implied by requiring 95% certainty of sufficiently low ratios.

Discussion

This paper addresses the question how a decision can be made between different health programmes with different ratios of costs to effects, specifically when taking into account the uncertainty surrounding the estimates of costs and effects. It is suggested that attitudes regarding risk should be considered as well as the idea of decreasing marginal value of money and health effects. In order to formalise this problem, we first considered the reasoning behind the use of cost-effectiveness ratios as introduced by Weinstein and Zeckhauser [1]. They derived that the optimal budget allocation is found by comparing the ratio of costs to effects with a critical ratio, and if the current ratio is smaller, the programme should be implemented. The value of the critical ratio is determined by the available budget, and the efficiency of available health care programmes. We contrasted this with the fixed trade-off approach, in which again a critical ratio is defined, but now that ratio is fixed, and not dependent on the budget. Of course, it seems likely that in practice the fixed critical ratio is chosen with at least some reflection about the available budget. However, over time the critical ratio resulting from the fixed budget approach and the fixed critical ratio will diverge. We have shown before that the fixed critical ratio can also be explained as indicating the preferences of the decision maker for a fixed trade-off between costs and effects.

Because it seems unlikely that a decision maker will have a true fixed budget or a true fixed ratio, we extended the standard approach of maximising health gains for a given budget to maximising the value derived from health gains and the budget available for purposes outside the health sector competing with health care spendings. When a value function is used, addition of a new health programme presents a choice between abandoning existing programmes and increasing the health care budget according to the decision maker's preferences. Moreover, it was shown that accepting a health programme without removing other health programmes may only be justified when assuming a very specific value function: one with constant marginal value for either money or health effects. Thus, the fixed trade-off approach is a special case of the flexible trade-off approach. In the latter, it is still possible that the budget will keep increasing as more efficient programmes become available. However, due to the assumed decreasing marginal utility, these increases in the budget will be limited, and the definition of what is considered an efficient programme will become stricter.

Using a value function including both health effects and money to decide between health programmes proves particularly fruitful when the ordering of programmes for which the outcomes are surrounded with uncertainty is addressed. We have shown that, if uncertainty is taken into account, the optimum budget allocation may differ from the allocation that would have been derived for the situation with no uncertainty surrounding costs and effects. However, for that purpose we assumed that a utility function for both money and health effects was defined. In practice, it may not be easy for a decision maker to specify preferences for uncertain outcomes so explicitly. Thus, we also addressed some other approaches to handling uncertainty.

The intuitive generalisation of the fixed budget approach is to maximise the expected health effects while staying within budget with some specified probability. However, the implications of such a goal may not adequately reflect the decision maker's true preferences, because the goal embodies the assumption that the decision maker is risk neutral regarding health and risk averse regarding costs. This will in general not be a rational risk attitude, and it has been suggested that the reverse will often be rational: risk aversion towards health effects and risk neutral towards costs. Thus, we have described a goal consistent with this, i.e. to minimise expected costs while achieving an aspiration level for health effects with some specified probability. It should be remarked that both approaches ignore interdependencies between a health program's costs and effects. Also, both approaches can be rewritten as a special case of the expected utility approach.

The generalisation of the fixed trade-off formalisation derives confidence intervals around the ratios of costs to effects and implements programmes only if the upper limit of the interval is below a critical ratio. Although these intervals are very informative, using them to decide on the implementation of health programmes may imply a rather strong risk aversion regarding both costs and effects. Therefore, one may conclude that the alternative approaches have their own implicit utility functions. These functions suggest that all decision makers have the same type of preference concerning uncertain situations, which is unlikely to be the case. This stresses the need for further research into the risk attitude of decision makers.

Other approaches to the problem of budget allocation under uncertainty have been published. For example, O'Brien and Sculpher propose formulating the problem of budget allocation under uncertainty as a portfolio selection problem [6]. In portfolio analysis the ultimate goal is to maximise expected return on investment and minimise uncertainty. These two goals will often be conflicting; hence a trade-off needs to be made between expected return on investment and variance of return on investment. That is, their approach also requires a utility function describing preferences and risk attitude regarding two attributes. It may be problematic to use the variance of the incremental cost-effectiveness ratio, because it is theoretically undefined.

Meltzer uses expected utility to model the choice as to how much of a specific treatment should be financed [5]. His model differs from ours in the assumption that all money not used for the specific treatment is used for non-medical consumption, and there is a fixed budget constraint. A more important difference from our analysis is that Meltzer goes on to assume perfect insurance at the population level. That is, perfect markets in claims contingent on states of the world exist and individuals can use insurance to adjust their consumption and income in different states of the world to their risk attitude. Therefore, medical events (states of the world) do not affect the marginal utility of income under perfect insurance.

Under the perfect insurance condition, Meltzer finds optimality conditions in terms of the ratio of expected costs to expected effects. The conditions we find for a risk neutral decision maker are similar to Meltzer's conditions. This is logical, since the assumption of perfect insurance and the assumption of risk neutrality both ignore the (co)variance in health effects and costs.

In his conclusion, Meltzer notes that: 'the issue of how risk should be assessed in policy decisions deserves further consideration because other assumptions about preferences concerning risk or about insurance would lead to different conclusions about many methodological issues in cost-effectiveness analysis.' Our model is rather general and allows for various assumptions about risk preferences. However, in contrast to Meltzer, our model describes the preferences of a single decision maker, which means that there can be no question of a market in contingent claims. An idea for future research would be to try to extend our model and start from the consumer perspective.

Finally, Hutubessy *et al.* propose stochastic league tables [4]. For a given budget, a stochastic league table presents the probability that a certain intervention i is in the optimum (in terms of our model, has $d_i^* = 1$). The stochastic league table approach does not show how the decision maker's risk attitude affects priority setting. Rather, it provides decision makers with information so that they can choose between uncertain cost-effectiveness outcomes and other values, for instance equity considerations. Similarly, the stochastic league table approach does not show how the decision maker's preferences about money and health affect priority setting. It is not discussed whether new programmes will be financed from a larger budget or a reduction in existing programmes, or both. Instead a range of league tables is presented for different budgets. Finally, it is important to realise that the stochastic league table approach does not take into account the joint distribution of costs and effects.

It is clear that to achieve optimal budget allocation much information is required: information on costs, effects and size for each health care programme (including information about uncertainty) and information on goals, preferences and risk attitude of decision makers. Information on expected costs and effects per patient is typically collected in cost-effectiveness studies. When those studies are properly done, they will also include information on the degree of uncertainty, either through statistical analysis or through a probabilistic sensitivity analysis. Information on the size of the programme may be collected through epidemiological studies. Since the estimated size will also be surrounded by uncertainty, the uncertainty around total costs and total effects is compounded of uncertainty about the average costs and effects per patient and the total number of patients eligible for the health care programme.

Two approaches can be followed to obtain information on goals, preferences and risk attitude of decision makers [10]. Firstly, this information may be inferred from choices made between alternative courses of public action in the past. The second method is direct, by asking the decision makers what their preferences are. Ideally this is done in an iterative process, in which the decision maker states preferences, upon which the implications of these preferences are worked out. This may lead to a adjustment of the preferences. Throughout the paper, the value functions and utility functions used were assumed to represent a hypothetical health care decision maker's preferences over health and money. Still the question arises whose preferences are to be reflected in the optimization problem and whether the objective functions used (the value function or utility function) should be interpreted as social welfare functions and what requirements on their characteristics might be implied. These types of questions address the welfare theoretic underpinnings of cost effectiveness analysis. An overview of the questions involved and the routes for solutions can be found in the first two chapters in Drummond and McGuire [24,25].

Although it may not be easy for decision makers to formulate goals, preferences and risk attitude, the fact that this information is required should be seen as an advantage of the described approaches. Currently, decision makers do weigh information on total costs, total effects and the amount of uncertainty when making decisions, and the approaches discussed here compel decision makers to make this explicit. It is not intended that the model should prescribe what the decision makers should decide. Rather, the model has two applications. First, it compels decision makers to be explicit about their preferences and informs them about the consequences of certain choices. Second, decisions can be analysed retrospectively, confronting the decision makers with the risk attitude reflected in their choices.

Other attributes, such as equity and ethics, may of course also play a role in the decision-making process. An advantage of the expected utility approach is that the function may easily be extended to include such attributes. As such, this approach presents a broad general framework for budget allocation that better describes reality than the currently assumed model of maximising health effects for a fixed budget.

Acknowledgements

Part of the work of Maiwenn Al was funded through the Pfizer Health Economics Fellowship Programme. Additionally, the work of Maiwenn Al and Talitha Feenstra was funded through MW-NWO grant 945-10-029. We thank Han Bleichrodt, Mark Koopmanschap, Xander Koolman and two anonymous reviewers for their constructive comments.

Appendix A

For certain combinations of utility function and probability density function it is possible to simplify Equation (5) further. As an example, suppose the decision maker is constantly risk averse regarding both money and health effects, and assume mutual utility independence. This is reflected by the following utility function

$$\begin{aligned} U(M-c,e) = {} & k_1(1-\exp^{-\delta(M-c)}) + k_2(1-\exp^{-\gamma e}) \\ & + k_3(1-\exp^{-\gamma e})(1-\exp^{-\delta(M-c)}) \end{aligned}$$

where δ and γ are parameters reflecting risk aversion if positive and risk proneness if negative, and k_1, k_2 and k_3 are scaling factors.

Furthermore, assume that for each intervention costs and effects follow a bivariate normal distribution. The expected utility $E[U(M-c,e)]$ can be found using the fact that if $x \sim \mathrm{N}(\mu, \sigma^2)$ then $\exp^x$ follows a lognormal distribution and $E(\exp^x) = \exp(\mu + \frac{1}{2}\sigma^2)$. Maximising this expected utility we find the following characterisation of the optimum

$$\begin{aligned} & R_1^i + R_2^i \geq 0 \to d_i^* = 1 \\ & R_1^i \leq 0 \to d_i^* = 0 \\ & \text{else} \to d_i^* = -\frac{R_1^i}{R_2^i} \end{aligned}$$

with

$$\begin{aligned} R_1^i = {} & -(k_1+k_3)\exp^{S_1}\delta\mu_{c_i} + (k_2+k_3)\exp^{S_2}\gamma\mu_{e_i} \\ & + k_3\exp^{S_3}(\delta\mu_{c_i} - \gamma\mu_{e_i}) \end{aligned}$$

$$\begin{aligned} R_2^i = {} & -(k_1+k_3)\exp^{S_1}\delta^2\sigma_{c_i}^2 + (k_2+k_3)\exp^{S_2}\gamma^2\sigma_{e_i}^2 \\ & + k_3\exp^{S_3}(\delta^2\sigma_{c_i}^2 + \gamma^2\sigma_{e_i}^2 - 2\gamma\delta\sigma_{ce_i}) \\ S_1 = {} & -\delta\left(M - \sum d_i\mu_{c_i}\right) + 1/2\delta^2\sum d_i^2\sigma_{c_i}^2 \\ S_2 = {} & -\gamma\sum d_i\mu_{e_i} + 1/2\gamma^2\sum d_i^2\sigma_{e_i}^2 \\ S_3 = {} & S_1 + S_2 - \delta\gamma\sum d_i^2\sigma_{ce_i} \end{aligned}$$

Note that compared to the situation under certainty, the variances and covariances of costs and effects of all programmes (the σ's) now appear in these conditions. Further note that if the variances converge towards zero, the conditions become similar to those outlined.

Appendix B

Maximising $\sum_{i=1}^n d_i E(e_i)$ subject to $P\left(\sum_{i=1}^n d_i c_i \leq B\right) \geq 1-\alpha$ is equivalent to expected utility maximisation using the utility function $U(M-c,e) = \sum_{i=1}^n d_i e_i + G(M-c)$ with

$$G(M-c) = \begin{cases} 0 & \text{if } P\left[\sum_{i=1}^n d_i c_i - B > 0\right] - \alpha \leq 0 \\ -\infty & \text{if } P\left[\sum_{i=1}^n d_i c_i - B > 0\right] - \alpha > 0 \end{cases}$$

and likewise minimising $\sum_{i=1}^n d_i E(c_i)$ subject to $P\left(\sum_{i=1}^n d_i e_i \geq L\right) \geq 1-\alpha$ is equivalent to expected utility maximisation using the utility function $U(M-c,e) = M - \sum_{i=1}^n d_i c_i + G(e)$ with

$$G(e) = \begin{cases} 0 & \text{if } P\left[\sum_{i=1}^n d_i e_i - L < 0\right] - \alpha \leq 0 \\ -\infty & \text{if } P\left[\sum_{i=1}^n d_i e_i - L < 0\right] - \alpha > 0 \end{cases}$$

References

1. Weinstein M, Zeckhauser R. Critical ratios and efficient allocation. *J Public Econ* 1973; **2**: 147–157.
2. Gafni A, Birch S. Guidelines for the adoption of new technologies: a prescription for uncontrolled growth in expenditures and how to avoid the problem. *Can Med Assoc J* 1993; **148**: 913–917.
3. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programmes. *Health Serv Res* 1972; **7**: 118–133.
4. Hutubessy RW, Evans DB, Barendregt JJ *et al.* Stochastic league tables: communicating cost-effectiveness results to decision-makers. *Health Econ* 2001; **10**: 473–477.

5. Meltzer D. Addressing uncertainty in medical cost-effectiveness analysis. Implications of expected utility maximazation for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. *J Health Econ* 2001; **20**: 109–129.
6. O'Brien BJ, Sculpher MJ. Building uncertainty into cost-effectiveness rankings: portfolio risk-return tradeoffs and implications for decision rules. *Med Care* 2000; **38**: 460–468.
7. Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. *J Health Econ* 1996; **15**: 641–653.
8. Laska EM, Meisner M, Siegel C. Statistical inference for cost-effectiveness ratios. *Health Econ* 1997; **6**: 229–242.
9. Birch S, Donaldson C. Applications of cost-benefit analysis to health care: departures from welfare economic theory. *J Health Econ* 1987; **6**: 211–225.
10. Sugden R, Williams A. *The Principles of Practical Cost-Benefit Analysis*. Oxford University Press: Oxford, 1978.
11. Centraal begeleidingsorgaan voor de intercollegiale toetsing. *Behandeling en preventie van coronaire hartziekte door verlaging van de plasmacholesterol-concentratie*. Centraal begeleidingsorgaan voor de intercollegiale toetsing: Utrecht, 1998.
12. Niessen LW, Dippel DWJ, Limburg M. Berekening van kosten na beroerte en kosteneffectiviteit van 'stroke units' en van secundaire preventie, zoals aanbevolen in de herziene CBO-richtlijn 'Beroerte' [Calculation of costs and cost effectiveness of stroke units and of secundary prevention in patients after a stroke, as recommended in the revised CBO guideline 'Stroke']. *Nederlands Tijdschrift voor Geneeskunde* 2000; **144**: 1959–1964.
13. Health WCoMA. *Macroeconomics and health: investing in health for economic development*. Report of the Commision on Macroeconomics. World Health Organization: Geneva, 2001.
14. Keeney RL, Raiffa H. *Decisions with Multiple Objectives*. Cambridge University Press: Cambridge, 1993.
15. Varian HR. *Microeconomic Analysis* (2nd edn). W.W. Norton & Company: New York, 1984.
16. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999; **18**: 341–364.
17. Arrow K, Lind A. Uncertainty and the evaluation of public investment decisions. *Am Econ Rev* 1970; **60**: 364–378.
18. Ben-Zion U, Gafni A. Evaluation of public investment in health care: is the risk relevant? *J Health Econ* 1983; **2**: 161–165.
19. Graff Zivin J. Cost-effectiveness analysis with risk aversion. *Health Econ* 2001; **10**: 499–508.
20. Graff Zivin J, Bridges J. Addressing risk preferences in cost-effectiveness analyses. *Appl Health Econ Health Policy* 2002; **1**(3): 135–139.
21. Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios:an example from a randomized trial. *Stat Med* 1996; **15**: 1447–1458.
22. van Hout BA, Al MJ, Gordon GS *et al*. Costs, effects and c/e-ratios alongside a clinical trial. *Health Econ* 1994; **3**: 309–319.
23. Fieller EC. Some problems in interval estimation. *J Roy Stat Soc Ser B* 1954; **16**: 175–185.
24. McGuire A. Theoretical concepts in the economic evaluation of health care. In *Economic Evaluation in Health Care*, Drummond M, McGuire A (eds). Oxford University Press: Oxford, 2001.
25. Tsuchiya A, Williams A. Welfare economics and economic evaluation. In *Economic Evaluation in Health Care; Merging Theory with Practice*, Drummond M, McGuire A (eds). Oxford University Press: Oxford, 2001.

