Mathematical programming for the optimal allocation of health care resources

Health Economics Study Group Winter 2005, Oxford UK

David Epstein[1], Zaid Chalabi[2], Mark Sculpher[1], Karl Claxton[1]

1- Centre for Health Economics, University of York

2- London School of Hygiene and Tropical Medicine

Version 3: 17 November 2004

Introduction

The establishment of the National Institute for Clinical Excellence offers the opportunity for a new approach to priority setting in England and Wales. NICE was established to provide clear national guidance 'as part of the process of improving the quality of healthcare across England and Wales'; given the limited resources available, this implies a requirement only to recommend healthcare which is cost-effective. However, NICE is ambiguous about its role in rationing healthcare. NICE undertakes technology appraisals, in which the committee makes recommendations about a set of healthcare technologies, which take into account the cost of additional quality adjusted life years (QALYs) by calculating the incremental cost effectiveness ratio (ICER). The threshold of willingness to pay for additional health benefits is not precisely defined but usually in the range £20000 to £40000[1]. The use of the ICER approach has a serious drawback. The method does not explicitly consider whether the recommendation is affordable, given its overall cost to the health service, and does not explicitly identify the treatments that will have to be displaced or alternatives forgone (the opportunity cost of the recommendation).

There have been recent attempts to formulate the allocation of resources across health care programmes as optimisation problems using Mathematical Programming (MP) techniques[2]. Only a few have attempted to apply these to real-world policy choices[3-6]. This work has been restricted to choosing between treatments in a given disease area.

The aims of this study are

- To formulate a mathematical programming framework to allocate resources within and between healthcare programmes
- To provide computational methods to solve the resulting deterministic MP problems
- To apply the mathematical framework to a stylised but relevant policy problem using real data
- To show that equity concerns can be represented as constraints
- To show the opportunity cost of equity concerns varies between patient groups

The approach taken in this study recognises certain important features of the policy problem. Resources are allocated in a publicly funded system by planning mechanisms. Ideally all treatments for all population groups in all healthcare programmes would be compared against one another to maximise overall health benefits, subject to budget, equity and other constraints. This is unlikely to be feasible. A more limited but relevant policy problem is suggested by the recognition that funding should be made available to NHS bodies in order to implement the recommendations made by NICE. Maynard recommends that NICE receives an annual, top sliced budget and is required to fund all advice within that expenditure envelope [7 8]. The consequence of such a policy would be that it would no longer be efficient for NICE to make recommendations for each appraisal in isolation from the others. In order to ensure that its recommendations were affordable within the overall financial envelope, all options would have to be compared against one another in a mathematical programme. A decision to implement a particular healthcare treatment will make claims on future budgets. This arises both from the commitment to treat the prevalent population over future years, and because the decision covers future incident populations, who may also be treated over several years. This study aims to handle this complexity by allowing a separate budget constraint for each future year.

Methods

Three healthcare programmes have been identified from the 6th and 7th wave of appraisals considered by NICE that were published between 2002 and 2003[9]. The healthcare programmes included in this study, their prevalent and incident populations,

the treatments and the time horizon over which they are appraised by the assessment reports are shown in Table 1. Briefly, we have selected a set of three independent healthcare programmes[DO1] (indexed by $k = 1..3$), each of which is targeted at between 2 and 4 patient populations (indexed by i). Within each programme there are between 2 and 4 treatments (indexed by j). There are 24 decision variables (x_{ijk}) in total, which are the proportion of population group i which is to be allocated treatment j in healthcare programme.

x_{ijk} varies between zero and unity ($0 \le x_{ijk} \le 1$) where

- $x_{ijk} = 0$ means that no proportion of population group i is allocated treatment j in healthcare programme k, and
- $x_{ijk} = 1$ means that all members of population group i are allocated treatment j in healthcare programme k

Note that for each healthcare programme k, the proportions of a population group i should add to unity across treatments j. This means that the following set of equalities should always hold:

$$\sum_j x_{ijk} = 1 \qquad (1.1)$$

Denote the time horizon of the evaluation by n years and denote the time index variable by t, where $t = 1....n$. The above healthcare programmes are evaluated over 15 years (i.e. $n = 15$). Denote by $c_{ijk}(t)$ the incremental cost at year t of treatment j in healthcare programme k if the treatment is given to all members of population group i. The cost $c_{ijk}(t)$ in any year t includes the costs of the prevalent population plus we assume that the same treatment decision will also be applied to patients who are newly diagnosed during the next 5 years and their costs are also taken into account. After 5 years it is assumed the decision will be reviewed and the costs of no further incident patients are considered. The cost for each treatment j where $j > 1$, c_{ijk}, is relative to a comparator treatment ($j = 1$) for which costs are defined to be zero. The comparator is usually current care. The total incremental cost in year t of all

healthcare programmes (relative to current care) across all treatments is therefore given by

$$C_t = \sum_i \sum_j \sum_k x_{ijk}\, c_{ijk}(t) \qquad t = 1...n \qquad (1.2)$$

Equation (1.2) assumes constant return to scale. Non-constant returns to scale will be addressed in a separate problem.

For gross benefits, only the cumulative incremental QALYs (relative to the comparator treatment) over the time horizon of the model are known and these would be denoted by $\{b_{ijk}\}$, where b_{ijk} is the gross benefit of treatment j in healthcare programme k if the treatment is applied to all members of population group i. The total gross incremental benefit relative to current care is therefore

$$B = \sum_i \sum_j \sum_k b_{ijk}\, x_{ijk} \qquad (1.3)$$

As in the case of the Eqn. (1.2) , Eqn. (1.3) also assumes constant return to scale. The coefficients $\{b_{ijk}\}$ are discounted.

The formulation of the basic mathematical programme

The basic Mathematical Programming (MP) is defined as follows. The objective is to determine the optimal values of $\{x_{ijk}\}$ so as to maximise the gross benefit B subject to an overall budgetary constraint δ, and constraints that ensure all members of each healthcare programme k and population group i receive one and only one treatment, that is:

$$\max_{\{x_{ijk}\}}(B)$$

$$\sum_t C_t \leq \delta$$

$$0 \leq x_{ijk} \leq 1 \qquad (1.4)$$

$$\sum_j x_{ijk} = 1$$

Problem (1.4) is a Linear Programming (LP) problem.

Alternative budget rules

Problem (1.4) imposes no constraints about when the budget can be spent, and the solution shows us the optimal allocation of resources over time. Conventional methods of cost-effectiveness analysis assume that the decision maker has complete flexibility about when the budget can be spent. However, a budget in the context of the National Heath Service is usually a sum of money allocated for a particular purpose for a given period of time, usually one year.

It is possible to explore alternative budget rules. Firstly, we consider the case where the total allocation δ is divided into equally sized maximum annual budgets over the time horizon of the analysis, that is, 15 years. The LP is given by Problem (1.5) and the budget rule given by Equation (1.6).

$$\max_{\{x_{ijk}\}}(B)$$

$$C_t \leq \delta_t \qquad t = 1..n$$

$$0 \leq x_{ijk} \leq 1 \qquad (1.5)$$

$$\sum_j x_{ijk} = 1$$

Budget allocation rule number 1:

$$\delta_t = \delta / 15 \qquad t = 1...15 \tag{1.6}$$

Secondly, we consider the case where the total allocation has to be spent within the first 5 years. The budget rule is given by Equation (1.7). In this formulation, treatments that have a cost beyond 5 years are permitted only if their costs are offset by other programmes which are cost saving in these time periods.

Budget allocation rule number 2:

$$\delta_t = \begin{cases} \delta / 5 & t = 1..5 \\ 0 & t = 6..15 \end{cases} \tag{1.7}$$

Efficiency versus Equity considerations

In Problem (1.4), the decision variable $\{x_{ijk}\}$ is permitted to take fractional values. This is the assumption of perfect divisibility. Only a proportion of the population would be allocated to receive each of the treatments under consideration (although all members of the population would receive some treatment). This might be thought of as the 'efficient' solution in the absence of any equity concerns. Its implementation, were it possible, would require some arbitrary allocation mechanism (since the members of the population are assumed to be of equal need), such as first-come first-served.

A requirement that equity considerations should be incorporated can be thought of as imposing additional constraints on the mathematical programme. The 'horizontal equity' consideration that people with equal need should receive equal access to treatment imposes the constraint that the $\{x_{ijk}\}$ are binary for some population group i and healthcare programme k (Equation (1.8)). For example, we might require that all patients with type 1 diabetes are treated in the same way, or that patients aged less than 60 years with non-Hodgkins lymphoma are treated in the same way, or both.

$$x_{ijk} \in 0,1 \qquad (1.8)$$

Further equity considerations can be incorporated into the mathematical programme. Any characteristic which is known to predict life expectancy could be used to differentiate patients with respect to the treatments they are offered. It may not be considered equitable to differentiate between patient groups in certain ways. In some cases it may be acceptable to differentiate between patient groups on the basis of age. However, other characteristics may be more controversial such as gender or social class. This can be expressed as the requirement that patients within the same health care programme have the same probability of receiving a given treatment, regardless of other characteristics. The examples available in this stylised scenario are rather artificial, but are used to illustrate the technique. For example, it might be considered unfair to use patients' age to differentiate with respect to the treatment offered for lymphoma (which can be written as a constraint in the form $x_{122} = x_{222}$), or unfair to allow patients with type 1 and type 2 diabetes different probabilities of receiving the more effective long acting insulin treatment (written as a constraint in the form $x_{123} = x_{223}$). These concerns can be written as equity constraints which can be imposed either separately or together.

The formulation of equity concerns as constraints on a mathematical programme allows us to evaluate their opportunity cost in terms of QALYs forgone, compared to the scenario where there are no equity constraints. Furthermore, we can evaluate the opportunity cost of equity separately for each patient population. In principle, this might be used to test whether equity is 'affordable' for each patient population.

Input parameters for the mathematical programme

The inputs to the mathematical programme model are :

a) the incremental costs for each programme, population and treatment for each year $c_{ijk}(t)$ and the incremental cumulative QALYs for each programme,

population and treatment b_{ijk}. These are calculated from the information provided by the assessment team reports. An example of estimation is shown in the Appendix (http://www.york.ac.uk/inst/che/teehta.htm).

b) the value of the budget constraint δ. This is determined by the decision maker. We illustrate the solution for a range of budgets.

Computational methods

The basic linear programme is solved using the *Mathematica Version 5* command "DualLinearProgramming"[10]. This returns the optimal values of the decision variables and the shadow prices of the constraints. Horizontal equity concerns (Problem (1.4) with the additional constraint (1.8) imposed on one or more population groups) are a 0-1 Mixed Integer Linear Program (0-1 MILP). Problems of this type can be difficult to solve. In this paper, we use a computationally-intensive method of forcing the selected treatment options to take all possible permutations of values of 0 or 1, evaluating the linear programme for each permutation, and choosing the permutation that maximises the objective function. This method is only feasible when a very limited number of decision variables take binary values. Further work will focus on more general analytic methods.

Results

Figure 1a shows the proportion receiving each intervention in one patient population – adults with influenza. Four mutually exclusive treatment options are considered here, but treatments 3 and 4 are never implemented at all at any budget. At a budget between 20 and 50 million, two treatments are mixed in the sense that a proportion of the population receive each. At a budget of say, 180 million, the whole patient population will be offered the same treatment. Figure 1b shows the optimal level of implementation of the 24 decision variables for all patient populations. The solution, and whether an intervention is mixed or pure, changes with the budget.

Figure 2 shows the shadow price of the budget constraint. The shadow price is the gain in QALYs if the budget were to grow by £1 million. The reciprocal of the shadow price is proportional to the threshold willingness to pay for additional QALYs. The shadow price falls (or the threshold willingness to pay increases) with the size of the budget. In this example, at a budget of £338million, all the most effective treatments in all patient groups are funded in full and further budget will have no additional health benefits.

Table 2 shows the health gain achieved at an (arbitrarily chosen) budget of £180m with no equity constraints and the opportunity loss incurred under alternative budget rules. Under these alternative rules, not all the budget will be spent.

Table 3 shows the opportunity loss if horizontal equity is required in two patient population, and if the constraints are imposed in both populations. In the large population with type 1 diabetes, the opportunity loss is 520 QALYs. The effect (compared with the base case) is to decrease the health of the type 1 diabetes population, who now all receive the less effective treatment where previously only a proportion would have done, but to increase health in other populations more of whom now receive more effective treatments. The requirement for horizontal equity only in the small population with non-Hodgkins lymphoma costs 19 QALYs. The requirement for horizontal equity in both populations does not impose further opportunity loss because the second constraint is not binding once the first has been imposed.

Table 4 shows the results of an equity concern that it is unfair to use certain characteristics to differentiate with respect to the treatment offered, firstly for lymphoma and secondly for long acting insulins. As before, the opportunity loss of holding these equity concerns is not the same for all patient-programmes. There is an additional opportunity loss if we wish to hold these equity constraints in both populations.

Discussion

This study has used linear programming to assist making a policy decision. In doing so, we have incorporated a number of features into the problem that are unavailable if decisions are made by using a threshold willingness to pay or 'league tables' of ICERs. Firstly, making decisions subject to a total or annual budget constraint allows us to consider the affordability and opportunity cost of the choices that are made. We show that the profile of costs over time is important and evaluate alternative budgetary rules. Secondly, we are able to incorporate equity concerns as constraints and find that different equity concerns have different implications for efficiency and the effect will vary from patient population to population.

These methods may be useful to assist priority setting in a context where there are a finite number of clear alternative options and a fixed budget. One such context may be the decisions made each year by an agency such as NICE. The application of these techniques to a decision in this context has revealed several interesting methodological challenges that should be addressed if these techniques are to progress. We discuss a few of these and mention others.

Firstly, the current position of NICE is that its role is to decide whether a treatment is cost-effective. It is up to the government to decide whether it is affordable based on the macroeconomic position and competing demands for public funds (Sir Michael Rawlins, SMDM Conference Oct 2004).

Secondly, the reversibility or irreversibility of decisions is a difficult modelling problem. If decisions are considered irreversible, then if we wish to model annual future costs (and benefits) we should take into account the incidence of the disease, that is, the numbers of patients requiring the treatment in future years as well as current prevalence. For example, the prevalence of type 2 diabetes has doubled in the past 10 years [8 11]and any fixed budget health care system should take account of this projected growth. Each years costs will include both the cost in the first year of treating newly diagnosed patients plus the continuing costs of managing patients diagnosed and treated in earlier periods. In this case, the question arises as to how many years into the future will costs be considered for – in principle the time horizon could be infinite? In order to compare treatment programmes on a like-for-like basis we are really interested in modelling the long term equilibrium situation, in which all

currently prevalent patients treated on the 'old' technology have died or been successfully treated. It is not clear how to combine the 'short run' and 'long run' into a single model.

If the decisions are considered reversible, then the question arises as to when the decision will be reviewed. As in the previous case, we are really interested in the long-run equilibrium, so perhaps the prevalent population should be modelled separately.

Thirdly, there is a large set of treatment programmes which require a more dynamic modelling approach, either because the prevalent population affects future incidence of the disease (for example, infectious diseases) or because the treatment decisions and level of implementation taken now affect future incidence (e.g. disease prevention or health promotion programmes). Approaches to these problems have been suggested in the literature[3 12].

Fourthly, the parameters of the model are not known with certainty. There is uncertainty in the costs and benefits per patient, and the epidemiological parameters of current prevalence and future incidence. A probabilistic approach to handling uncertainty would need to incorporate the coefficients of the model as random variables with a priori specified probability distributions, including the correlations between them (particularly between costs and effects). The formulation of the mathematical program where the coefficients are random variables requires careful thought. One proposed formulation is that the budget constraint is satisfied in probability only, but this probability coefficient can be set to be arbitrarily close to 1[13]. This solution is unlikely to be the same as that for the deterministic case, (or the case where the coefficients are replaced with their expected values). Sendi 2003 suggests an approach to this problem where there are two portfolios to compare[14]. However, it is not clear how this should be solved in the general case.

There are several other challenges that would need to be addressed, among them numerically efficient methods for solving 0-1 MILP problems, the issue of uncertainty in parameter values (formulated as stochastic mathematical programming problems),

and non-linear cost and benefit functions (formulated as nonlinear mathematical programming problems).

References

1. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence it? A binary choice analysis. *Health Economics* 2004;13:437-452.
2. Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. *Journal of Health Economics* 1996;15(5):641-653.
3. Brandeau ML, Zaric GS, Richter A. Resource allocation for control of infectious diseases in multiple independent populations: beyond cost-effectiveness analysis. *Journal of Health Economics* 2003;22(4):575-598.
4. Wang LY, Haddix AC, Teutsch SM, Caldwell B. The role of resource allocation models in selecting clinical preventive services. *American Journal of Managed Care* 1999;5(4):445-454.
5. Zaric GS, Brandeau ML. Optimal investment in a portfolio of HIV prevention programs. *Medical Decision Making* 2001;21(5):391-408.
6. Earnshaw SR, Richter A, Sorensen SW, Hoerger TJ, Hicks KA, Engelgau M, et al. Optimal allocation of resources across four interventions for type 2 diabetes. *Medical Decision Making* 2002;22(5):S80-S91.
7. Maynard A, Bloor K, Freemantle N. Challenges for the National Institute for Clinical Excellence. *BMJ* 2004;329:227-229.
8. Efficiency of treatment of diabetes mellitus in general internal medicine departments in Germany: A DEA bootstrap approach. Fifth European Conference on Health Economics, London School of Economics, 8-11 Sept 2004; 2004; London.
9. NICE. http:\\www.nice.org.uk, 2004.
10. Wolfram S. *The Mathematica Book*. 4 ed. Champaign IL: Wolfram Media/ Cambridge University Press, 2003.
11. SMDM, editor. Symposium: Economic analysis informing health policy decision making. Society for Medical Decision Making 26th Annual Meeting, October 2004; 2004; Atlanta. SMDM.
12. Zaric GS, Brandeau ML. Dynamic resource allocation for epidemic control in multiple populations. *Ima Journal of Mathematics Applied in Medicine and Biology* 2002;19(4):235-255.
13. Chalabi Z, Dowie J. Cost effectiveness analyses: a stochastic programming approach for optimal decision making. *Health Economics Study Group Jan 2003*. Leeds, 2003.
14. Sendi P, Al MJ, Gafni A, Birch S. Optimizing a portfolio of health care programs in the presence of uncertainty and constrained resources. *Social Science & Medicine* 2003;57(11):2207-2215.

Table 1: Programmes, treatments, populations and time horizon of assessment reports. The first three healthcare programmes are used in the analysis.

Healthcare programme (k) and population groups (i)	*Population Prevalent*	*Size: Incident /yr*	*Treatments (j)*	*Time horizon of assessment report*
k=1: Influenza			Four treatments available	Less than 1 year
1. Adults	4.805m	4.805m	1. Current care (no medication)	
2. Elderly	1.075m	1.075m	2. Drug A	
3. Residential elderly	0.025m	0.025m	3. Drug O	
4. Children	4.047m	4.047m	4. Drug Z	
k=2: Rituximab			Two treatments available	15 years
1. Over 60s	236	11	1. Current care (CHOP)	
2. Under 60s	1243	59	2. R+CHOP	
k=3: Long acting insulins			Two treatments available	9 years
1. Type 1	117000	4056	1. Current care (NHP)	
2. Type 2	39000	2790	2. Insulin glargine	
k=4: Dialysis for patients suitable for home dialysis	1300	236	Three treatments available 1. Current care * 2. Satellite dialysis 3. Home haemodialysis	10 years
k=5: Revascularisation			Two treatments available	6 years
1. High risk	7359	7359	1. PTCA + bare metal stents	
2. Low risk	22076	22076	2. PTCA + drug alluding stents 3. CABG	

* current care is the current mix of services : 30% have hospital dialysis, 30% satellite and 40% home dialysis. Hospital dialysis is not considered as a treatment option since it is less effective and more costly than alternatives in every year. However, future analyses may consider switch costs which are not included here.

** Healthcare programmes are indexed with letter k, treatments with letter j and population groups with letter i

Table 2 : Opportunity loss (in QALYs) of alternative budget rules

Budget rule	*Health gain (QALY)*	*Opp Loss (QALY)*	*Budget spent*
No constraint	7317	0	£180m
Equal phasing	3586	3731	£103m
All in 1st 5 years	4879	2438	£75m

Table 3 : Opportunity loss of horizontal equity concerns

	Health gain (QALY)	Opp. Loss (QALY)
No equity constraint †	3586	0
Equity popn. 1 (type 1 diabetes)	3066	520
Equity popn. 2 (age<60, lymphoma)	3547	19
Equity popn 1 and popn 2	3066	520

†The ‘base case’ is chosen for a budget that is ‘equally phased’.

Table 4: Opportunity loss of equity between populations

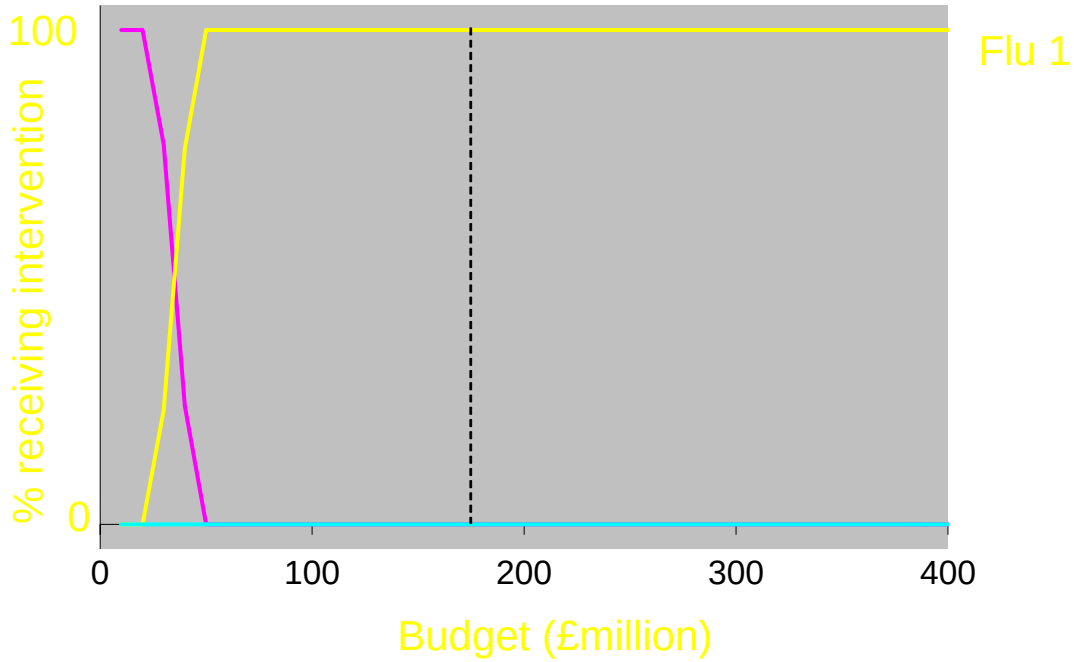
	Health gain (QALY)	Opp. Loss (QALY)
No equity constraint	3586	0
Equity: programme 1 (lymphoma: older = younger)	3579	7
Equity: programme 2 (diabetes: type 1 = type 2)	3126	460
Equity prog 1 and prog 2	3122	464

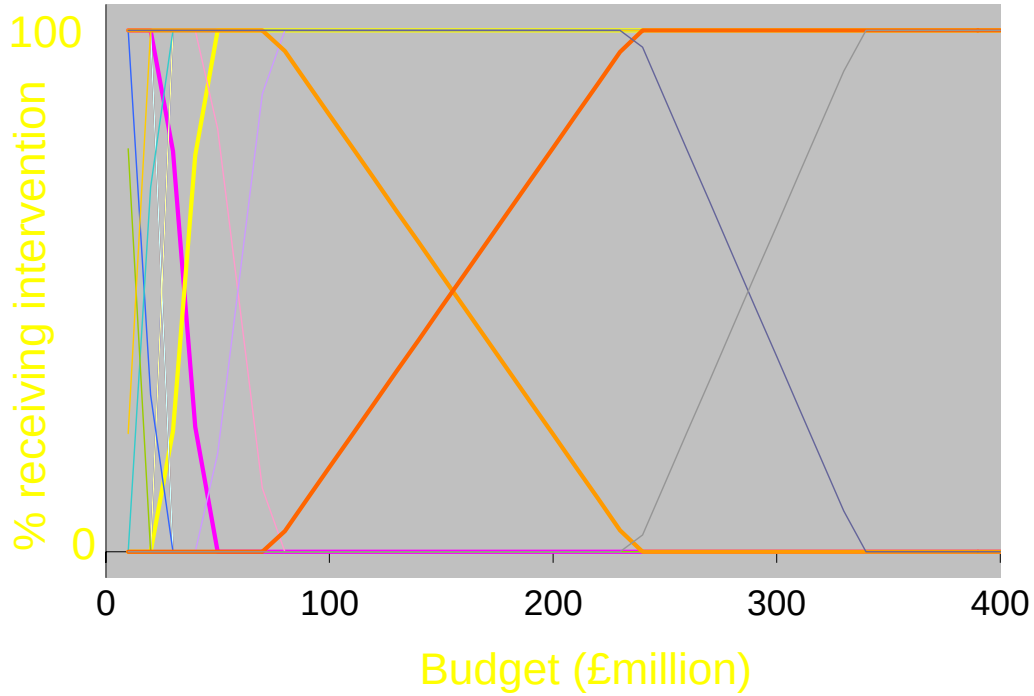
†The ‘base case’ is chosen for a budget that is ‘equally phased’.

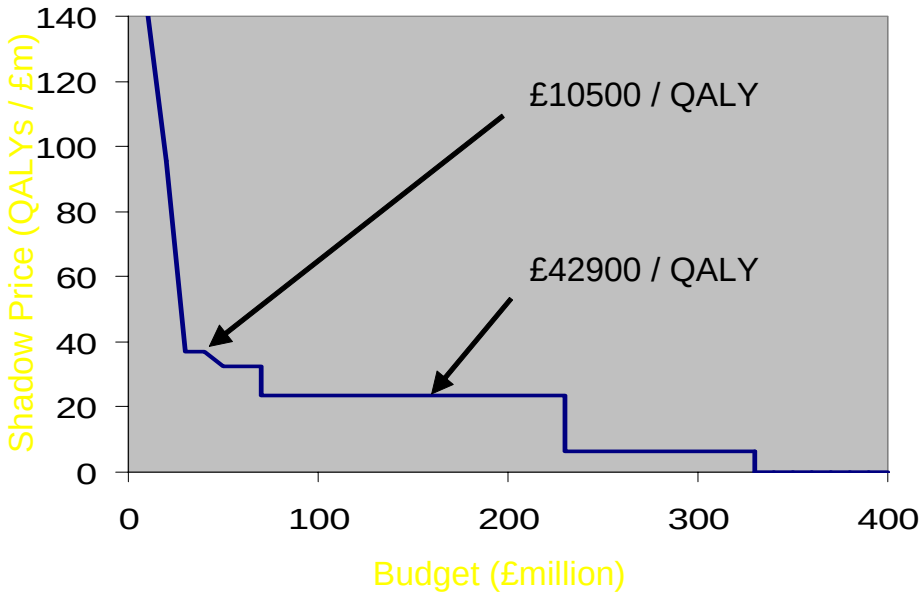
Figure 1a: The optimal solution for the treatment of influenza in the adult population at different values of the overall budget

Figure 1b: The optimal solution for all treatments and populations at different values of the overall budget

Figure 2: The shadow price of the overall budget constraint at different values of the overall budget, and corresponding threshold cost per QALY


100
% receiving intervention
0
0
100
200
300
400
Budget (£million)
Flu 1


100
% receiving intervention
0
0
100
200
300
400
Budget (£million)


140
120
100
80
60
40
20
0
Shadow Price (QALYs / £m)
£10500 / QALY
£42900 / QALY
0
100
200
300
400
Budget (£million)