

OPTIMAL SELECTION POLICIES FOR A SEQUENCE OF CANDIDATE DRUGS

C. CHARALAMBOUS* ** AND

J. C. GITTINS,* *** *University of Oxford*

Abstract

Pharmaceutical companies have to face huge risks and enormous costs of production before they can produce a drug. Efficient allocation of resources is essential to help in maximizing profits. Yu and Gittins (2007) described a model and associated software for determining efficient allocations for a preclinical research project. This is the starting point for this paper. We provide explicit optimal policies for the selection of successive candidate drugs for two restricted versions of the Yu and Gittins (2007) model. To some extent these policies are likely to be applicable to the unrestricted model.

Keywords: Pharmaceutical research; resource allocation; semi-Markov decision process; optimization

2000 Mathematics Subject Classification: Primary 90B36

Secondary 90C40; 91B32

1. Introduction

The pharmaceutical industry has been expanding rapidly in the past 40 years. Its growth is currently estimated to be 9–15% per year. A very large initial investment is needed to produce a new drug. The combined cost of research and development and market introduction for a significant product exceeds £700 million, extending over a period of around 15–30 years (see [8]). At most 20% of new chemical entities (otherwise known as candidate drugs or development compounds) survive the clinical trials and become marketable. Thus, the whole process of producing a drug is characterized by great uncertainty and the risks involved are huge.

Pharmaceutical companies strive to increase the probability of success of their research projects, and above all to maximize profitability. In order to achieve this, the resource allocation in the research and development process leading to a drug may be modeled, and optimality criteria used so that resources are allocated to the most profitable projects.

An important stage in a pharmaceutical research project is the discovery of a *lead compound*. This is a compound which has sufficiently interesting characteristics to be used as the starting point of a *lead series (LS)* of similar compounds for testing in the search for a compound which is worthy of submission for clinical trials as a *development compound (DC)*, or candidate drug. LS are series of compounds with similar chemical structure. Thus, a lead compound is a lead in the investigation, like a piece of evidence in police work. Most DCs fail at some point during clinical trials, so typically further DCs are in due course identified. These may be from one or more LS.

Received 2 February 2007; revision received 4 March 2008.

* Postal address: Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK.

** Email address: charalam@stats.ox.ac.uk

*** Email address: gittins@stats.ox.ac.uk

The work described in this paper is part of a research program developing models and software for resource allocation in pre-clinical pharmaceutical research, which Gittins and co-workers in Oxford have been working on for the past twelve years. Gittins [6], [7] set up a stochastic model to investigate the relationships between profitability and the number of scientists allocated to the different stages of a pharmaceutical research project. Chen [5] extended this model by incorporating internal rate of return and the option valuation approach as additional project evaluation criteria. An important general conclusion is that larger project teams than those which are typical would be much more profitable in some cases. Yu and Gittins [15] extended the earlier models to take specific account of the number of distinct LS from which the DCs in a project should be selected. An important conclusion is that it is often worth incurring the additional cost of optimizing more than one LS in the search for backup DCs, so as to reduce the risk of a string of DCs which all fail for similar reasons.

Some good sources describing models as aids in the determination of plans for pharmaceutical research and development expenditure are the journals *Pharmacoeconomics* and *R&D Management*. Good review papers have been written by Miller [11] and Poh *et al.* [12]. Bergman and Gittins [1] described the research and development resource allocation methods which at that time had been applied in the pharmaceutical industry. Halliday *et al.* [9] invited the world's top 45 pharmaceutical companies in terms of sales revenue and research and development expenditure to complete a questionnaire concerning their research and development activity, including their strategies employed to speed up the process of drug development. The survey divided the companies into top ten companies by research and development expenditure, other western companies, and Japanese companies. An important finding was that the top ten companies in research and development to sales ratio send more DCs into clinical trials. Jacob and Kwak [10] reviewed the recent changes in the pharmaceutical industry regarding costs, technology, risks, and environmental changes and described how real option theory, based on the Black and Scholes [2] methodology for valuing financial options, has been used in pharmaceutical research and development in the last few years. Major improvements in technology management can be achieved by using options, such as the option to wait and the option to abandon (see also [5]).

In this paper we derive explicit optimal policies for the selection of successive candidate drugs in the pre-clinical stages of research. This is done by means of detailed stochastic modeling. We are not aware of any other similar work. The comment by Miller [11], for example, that there is currently very little use of pharmacoeconomics in the early phases of drug development and that much more can be done, supports this impression. (Pharmacoeconomics is the science of relating the costs and benefits, both to individuals and to society, of therapeutic regimes, including drugs (see [15]). Analysis along these lines is becoming routine in the planning of clinical trials.)

2. Aims

This paper introduces Models 2 and 3 which are restricted and simplified versions of the more realistic model considered by Yu and Gittins [15], which we will call Model 1. The three models differ in their criteria for optimality and in other ways. For Model 1, there are two alternative optimality criteria for assessing the profitability of an allocation plan for a project. These are the project's *profitability index* and its *internal rate of return*, both of which are standard criteria for the evaluation of capital projects (see, e.g. [3]). The model for a research project includes the features described in this section for Model 3, together with more detailed modeling of the successive stages of the research process, of the rate at which effort is applied at each stage, and of the effect of obsolescence.

In Section 3 we set out some of the general theory which is needed to work out the expected value generated by different policies for the selection of DCs. An expression for the expected value of the next DC chosen, conditional on the LS from which previous DCs were chosen, is derived in Section 3.2.2.

In Section 4 we explore suitable policies for selecting DCs from more than one LS using Models 2 and 3. For these models, the standard theory of semi-Markov decision processes, as described, for example, in [13], is applicable. As we will see, for Models 2 and 3, the optimal policy takes a simple form. For Model 2, it is optimal to select DCs repeatedly from each LS in turn, and, for Model 3, it is optimal to select a fixed number of DCs from one LS after another, never coming back to a previously sampled LS. For these policies, it is straightforward to derive expressions for the *total expected reward (TER)* and, hence, to examine its dependence on the various project parameters.

In Section 5 the general conclusions of this report and possible future extensions are discussed.

For Model 2, our aim is to maximize the TER from a number n of DCs selected from a fixed number m of LS. The state space consists of m -vectors which give the number of DCs which have so far been selected from each of the m LS. The action w which is chosen indicates the LS from which the next DC is selected.

Formally, Model 2 is as follows.

Model 2. A typical state $i = (i_1, i_2, \dots, i_m)$, where i_1, i_2, \dots are nonnegative integers. Action $w \in \{1, 2, \dots, m\}$. Action w causes transition to state $j = (j_1, j_2, \dots, j_m)$ with probability $P(i, j, w)$, where

$$P(i, j, w) = \begin{cases} 1 & \text{if } j_w = i_w + 1; j_v = i_v, v \neq w, \\ 0 & \text{otherwise.} \end{cases}$$

The transition time $t(i, w)$ needed to move from state i to the next state when action w is chosen is 1, for all i and w . There is no discounting. The expected reward $R(i, w)$ when action w is chosen in state i is the expected value of the resulting DC. As we will see, this is given by Theorem 2, below, with $s = i_w$.

For Model 3, our aim is to maximize the *total expected discounted reward (TEDR)* from an infinite number of available LS, from any of which we can choose an infinite number of DCs; we assume an infinite time horizon, that rewards are discounted with time, and that there is a nonzero probability that a given LS will not produce a DC. Model 3 also allows for the fact that looking for a DC in a new LS takes longer, and is more likely to fail, than looking for further DCs from an LS that has already produced a DC. The state is again an m -vector, now representing the m LS from which we have so far selected at least one DC.

Formally, Model 3 is as follows.

Model 3. Initial state $i = 0$. Otherwise state $i = (i_1, i_2, \dots, i_m)$, where i_1, i_2, \dots are positive integers. Action $w \in \{1, 2, \dots, m, \star\}$, where $w = \star$ means we look for a DC from an LS in which no DC has been found and carry on looking in different new LS until a DC is found. Let

$$P(\text{search of an LS finds a DC}) = \begin{cases} p & \text{on the first search,} \\ 1 & \text{on each later search if the first search is successful,} \\ 0 & \text{on each later search if the first search is unsuccessful.} \end{cases}$$

Action w causes transition to state $j = (j_1, j_2, \dots, j_m)$ with probability $P(i, j, w)$, where

$$P(i, j, w) = \begin{cases} 1 & \text{if } n = m + 1; j_v = i_v; v = 1, 2, \dots, m; j_{m+1} = 1; w = \star, \\ 1 & \text{if } n = m; j_w = i_w + 1; j_v = i_v, v \neq w; w \in \{1, 2, \dots, m\}, \\ 0 & \text{otherwise.} \end{cases}$$

The transition time

$$t(i, w) = \begin{cases} Kt_1 & \text{if } w = \star, \\ t_2 & \text{if } w \neq \star. \end{cases}$$

Here t_1 and t_2 are fixed times and $P(K = k) = p(1 - p)^{k-1}$, $k = 1, 2, \dots$. A reward at time t is discounted by the factor $e^{-\gamma t}$. The reward $R(i, w)$ occurs after the transition time $t(i, w)$ and is as for Model 2 unless $w = \star$, in which case it is given by Theorem 2 with $s = 0$.

The action space $A(i)$ is defined by the set of available LS. Thus, the set of available actions is different for different states. Since LS from which no attempt has so far been made to find a DC are assumed to be indistinguishable, this means that the set of available actions is always finite.

Actions are chosen according to some policy. A policy is any rule for choosing actions. At any stage the action chosen could be random and depend on time and on the previous history of the process. Important classes of policies are the classes of *deterministic Markov (DM)* policies, and of *deterministic stationary Markov (DSM)* policies.

- *Deterministic* policies involve no randomization.
- *Stationary* policies involve no explicit time dependence.
- *Markov* policies are those for which each decision taken depends only on the current state and point in time.

Thus, a DSM policy is equivalent to a function from the state space to the action space.

The following theorem gives the results which we need from the standard theory of semi-Markov decision processes.

Theorem 1. *For a semi-Markov decision process with countable state space and finite action space,*

- (i) *with a finite time horizon and optimality criterion TER, there is an optimal policy which is DM; and*
- (ii) *with an infinite time horizon and optimality criterion TEDR, there is an optimal policy which is DSM, provided that the TEDR is bounded and that Condition 1, below, holds.*

Condition 1. *For every state i and action a , there is a positive probability that the transition time is greater than some fixed positive δ .*

Theorem 1(ii) is part of Theorem 7.2 of Ross [13] with different notation. Theorem 1(i) is a simpler result which follows from the discussion given by Ross, and is explicitly stated as Result 2.3 of [14]. Note that the proofs still work if the set of available actions is different for different states, even if the union of all these sets is infinite.

Model 2 is a Markov decision process with optimality criterion TER and Theorem 1(i) applies. Model 3 is a semi-Markov decision process with infinite time horizon and optimality

criterion TEDR and Theorem 1(ii) applies. We will therefore restrict attention to DM policies for Model 2 and to DSM policies for Model 3. One consequence is that, for Model 3, we only consider policies which, after unsuccessfully looking for a first DC from an LS, immediately look for a first DC in another LS.

3. The values of a sequence of candidate drugs

3.1. Assumptions and notation

A *good* LS is an LS from which we can get a successful DC. A DC is *successful* if it passes through clinical trials and becomes marketable. Different LS are good or bad independently and with the same probability conditional on the target being achievable. DCs are successful or unsuccessful independently and with the same probability conditional upon them coming from good LS.

3.1.1. General notation.

- S_j : the j th available LS, $j = 1, 2, \dots$
- C_{ij} : the i th DC from LS S_j , $i, j = 1, 2, \dots$
- U : the therapeutic target.

3.1.2. Events.

- A_{ij} : C_{ij} passes successfully through clinical trials and becomes a marketable drug. In brief, A_{ij} is the event $\{C_{ij} \text{ is successful}\}$.
- B : therapeutic target U is in principle achievable.
- C_i : LS S_i has no systematic defect which makes it impossible to reach the target U . In brief, C_i is the event $\{S_i \text{ is good}\}$.
- \bar{X} : the complement of the event X , for any X .

3.1.3. Probabilities.

- Conditional on B , the events C_i are independent.
- Conditional on B , C_j , and C_n , the events A_{ij} and A_{mn} are independent (including the case in which $j = n$, where $(i, j) \neq (m, n)$).
- $P(C_i | \bar{B}) = 0$.
- $P(A_{ij} | \bar{B}) = P(A_{ij} | \bar{C}_j) = 0$.
- $P(A_{ij}) = a > 0$.
- $P(B) = b > 0$.
- $P(C_i | B) = c$, $0 < c < 1$.

$$P(A_{ij} | B, C_j) = \frac{P(A_{ij} \cap B \cap C_j)}{P(B \cap C_j)} = \frac{P(A_{ij})}{P(C_j | B) P(B)} = \frac{a}{bc} = \alpha.$$

In the above i, j, m , and n are any positive integers.

3.1.4. *Values.*

- V : the value of the first DC if it was available now. In principle, this is the expected net present value of all the cash flows directly attributable to the compound.
- $\lambda^n V$, $0 < \lambda < 1$: the value of the $(n + 1)$ th successful DC in a project, ignoring discounting because of the need to wait until the compound is available. The factor λ represents the diminishing returns associated with a sequence of compounds all competing for the same target.
- $V(r)$: the value of the r th DC. Note that this depends on the success or failure of the $r - 1$ previous DCs.

3.2. Evaluation of probabilities and expected values

3.2.1. *Two LS, S_1 and S_2 .* Assume that the DCs are selected alternately from S_1 and S_2 . Let $A(r, m, n)$ be the event that the r th DC is successful and is preceded by m successful DCs from S_1 and n successful DCs from S_2 (together with a total of $r - 1 - m - n$ unsuccessful DCs). Bearing in mind that the $(2r + 1)$ th compound is chosen from the first LS we find that, for $n > 0$,

$$\begin{aligned}
 P(A(2r + 1, m, n)) &= \binom{r}{m} \underbrace{P(A_{11} | B, C_1)^{m+1}}_{\text{successful compounds from 1st LS}} \underbrace{P(\bar{A}_{11} | B, C_1)^{r-m}}_{\text{unsuccessful compounds from 1st LS}} \cdot \\
 &\quad \times \binom{r}{n} \underbrace{P(A_{12} | B, C_2)^n}_{\text{successful compounds from 2nd LS}} \underbrace{P(\bar{A}_{12} | B, C_2)^{r-n}}_{\text{unsuccessful compounds from 2nd LS}} P(B) P(C_1 | B) P(C_2 | B) \\
 &= \binom{r}{m} \alpha^m (1 - \alpha)^{r-m} \binom{r}{n} \alpha^n (1 - \alpha)^{r-n} ac.
 \end{aligned}$$

The case when the $(2r + 1)$ th DC is successful and is preceded by m successful DCs from S_1 and no successful DCs from S_2 requires separate consideration. It is now uncertain whether LS S_2 is good or bad. We have

$$\begin{aligned}
 P(A(2r + 1, m, 0)) &= P(A(2r + 1, m, 0 | B, C_1, C_2)) P(B) P(C_1 | B) P(C_2 | B) \\
 &\quad + P(A(2r + 1, m, 0 | B, C_1, \bar{C}_2)) P(B) P(C_1 | B) P(\bar{C}_2 | B) \\
 &= \binom{r}{m} \alpha^m (1 - \alpha)^{r-m} (1 - \alpha)^r ac \\
 &\quad + \binom{r}{m} \alpha^m (1 - \alpha)^{r-m} a(1 - c).
 \end{aligned}$$

Thus,

$$\begin{aligned}
 E[V(2r + 1)] &= V \sum_{m=0}^r \sum_{n=0}^r \lambda^{n+m} P(A(2r + 1, m, n)) \\
 &= V \left(\sum_{m=0}^r \sum_{n=0}^r \binom{r}{m} (\lambda\alpha)^m (1 - \alpha)^{r-m} \binom{r}{n} (\lambda\alpha)^n (1 - \alpha)^{r-n} ac \right. \\
 &\quad \left. + \sum_{m=0}^r \binom{r}{m} (\lambda\alpha)^m (1 - \alpha)^{r-m} a(1 - c) \right)
 \end{aligned}$$

$$\begin{aligned}
 &= V(ac(1 - (1 - \lambda)\alpha)^{2r} + a(1 - c)(1 - (1 - \lambda)\alpha)^r) \\
 &= aV(c\eta^{2r} + (1 - c)\eta^r),
 \end{aligned}$$

where $\eta = 1 - (1 - \lambda)\alpha$.

Similarly, for $m > 0$,

$$P(A(2r, m, n)) = \binom{r}{m} \alpha^m (1 - \alpha)^{r-m} \binom{r-1}{n} \alpha^n (1 - \alpha)^{r-1-n} ac,$$

and, when $m = 0$,

$$\begin{aligned}
 P(A(2r, 0, n)) &= \binom{r-1}{n} \alpha^n (1 - \alpha)^{r-1-n} (1 - \alpha)^r ac \\
 &\quad + \binom{r-1}{n} \alpha^n (1 - \alpha)^{r-1-n} a(1 - c).
 \end{aligned}$$

Thus,

$$\begin{aligned}
 E[V(2r)] &= V(ac(1 - (1 - \lambda)\alpha)^{2r-1} + a(1 - c)(1 - (1 - \lambda)\alpha)^{r-1}) \\
 &= aV(c\eta^{2r-1} + (1 - c)\eta^{r-1}).
 \end{aligned}$$

3.2.2. *Arbitrary set of LS, arbitrary selection rule.* The following notation is assuming that we are given the numbers of DCs which have so far been selected from each of the available LS, and that we do not know whether they are going to be successful or not. In particular, the distribution of $V(\mathcal{C})$ is conditional on the current state.

- \mathcal{C} : the next DC chosen at a given point in the sequence.
- $V(\mathcal{C})$: the value of the next DC.
- S : the LS from which we will choose the next DC.
- C : the event $\{S \text{ is good}\}$.
- Θ : the set of all LS from which DCs have been selected prior to DC \mathcal{C} , excluding LS S , where $\mathcal{C} \in S$.
- θ : any subset of Θ and the event $\{\theta \text{ is the set of all good LS in } \Theta\}$.
- $p_\theta = P(\theta \mid B)$.
- r_θ : number of DCs selected from any LS in θ before \mathcal{C} is selected.
- s : number of DCs selected from LS S before \mathcal{C} is selected.
- $\mathbf{1}(i)$: the indicator random variable for the event that the i th DC from $\{\theta\} \cup S$ is successful.
- $\eta = 1 - (1 - \lambda)\alpha$.

Theorem 2. For any $\mathcal{C} \in S$,

$$E[V(\mathcal{C})] = aV\eta^s \sum_{\theta \subset \Theta} p_\theta \eta^{r_\theta}.$$

Proof. We have

$$E[V(\mathcal{C})] = \sum_{\theta \subset \Theta} E[V(\mathcal{C}) \mid B, C, \theta] P(B, C, \theta). \tag{1}$$

Now $E[V(\mathcal{C}) \mid B, C, \theta] = E[\lambda^g \mathbf{1}(r+1)V \mid B, C, \theta]$, where $g = \sum_{i=1}^r \mathbf{1}(i)$ and $r = r_\theta + s$. Since the random variables $\mathbf{1}(i)$ are independent, conditional on B, C , and θ , it follows that

$$E[V(\mathcal{C})] = V E[\mathbf{1}(r+1)] E[\lambda^{\mathbf{1}(i)}]^r,$$

all expectations being conditional on B, C , and θ . We have $E[\mathbf{1}(r+1) \mid B, C, \theta] = \alpha$ and $E[\lambda^{\mathbf{1}(i)} \mid B, C, \theta] = \alpha\lambda + 1 - \alpha = \eta$, so that $E[V(\mathcal{C}) \mid B, C, \theta] = V\alpha\eta^{r_\theta+s}$. Also, $P(B, C, \theta) = P(B)P(C \mid B)P(\theta \mid B, C) = bc\rho_\theta$.

Substituting into (1), it follows that

$$E[V(\mathcal{C})] = aV\eta^s \sum_{\theta \subset \Theta} p_\theta \eta^{r_\theta},$$

where now the expectation is unconditional, as required.

Alternatively, we may write $E[V(\mathcal{C})] = aV E[\eta^\rho \mid B]$, where ρ denotes the number of DCs selected before \mathcal{C} which belong either to S or to a good LS.

Corollary 1. *We have*

$$E[V(\mathcal{C}) \mid B, C] = \alpha V \eta^s \sum_{\theta \subset \Theta} p_\theta \eta^{r_\theta}.$$

Proof. We have

$$E[V(\mathcal{C}) \mid B, C] = \sum_{\theta \subset \Theta} E[V(\mathcal{C}) \mid B, C, \theta] p_\theta.$$

Using the proof for Theorem 2, we know that $E[V(\mathcal{C}) \mid B, C, \theta] = V\alpha\eta^{r_\theta+s}$. Therefore,

$$E[V(\mathcal{C}) \mid B, C] = \sum_{\theta \subset \Theta} E[V(\mathcal{C}) \mid B, C, \theta] p_\theta = \sum_{\theta \subset \Theta} V\alpha\eta^{r_\theta+s} p_\theta = \alpha V \eta^s \sum_{\theta \subset \Theta} p_\theta \eta^{r_\theta},$$

as required.

4. Optimal selection policies

4.1. Model 2 (undiscounted)

4.1.1. *Background.* Consider a finite number of LS m from which we are going to take a finite number n of DCs. There are m^n possible policy schedules (or, equivalently, DM policies), each consisting of a string of n integers specifying the order in which the m LS are to be used to produce successive DCs. We look for a schedule that maximizes the total reward from the n DCs. As the number of possible schedules is finite, this maximum must be attained.

4.1.2. *Optimal policy derivation.* We will show that the optimal policy is to divide the n DCs selected as nearly equally as possible between the m LS. To prove this, we use the following four lemmas.

Lemma 1. *If DCs are to be selected from m LS, with r_i DCs to be selected from LS S_i , $i = 1, 2, \dots, m$, the total number of successful DCs does not depend on the order of selection.*

Proof. This is an immediate consequence of the fact that none of the probabilities described in Section 3.1.3 depends on the order of selection.

Lemma 2. *The TER from fixed numbers of DCs from each LS does not change with the order of selection.*

Proof. From Lemma 1, it follows that all functions of the number of successful DCs, such as the total reward, do not depend on the order of selection, and the lemma follows.

Lemma 3. *At any stage the expected value of the next DC is maximized by choosing it from an LS from which the minimum number of compounds have so far been selected.*

Proof. We will show that, for any two LS, we will get a higher expected value if we choose the one from which the lower number of DCs have so far been selected. Recall that Theorem 2 gives $E[V(\mathcal{C})] = aV E[\eta^\rho \mid B]$, where ρ denotes the number of DCs selected before \mathcal{C} which belong either to S or to a good LS, S being the LS from which the DC \mathcal{C} is chosen.

We will consider selection from two different LS, S_i and S_j . Assume that we have so far selected r_i DCs from S_i and r_j DCs from S_j .

Let $\rho = \rho_i$ if we select from S_i and let $\rho = \rho_j$ if we select from S_j .

Let v_{ij} denote the number of DCs which have been selected from good LS other than S_i and S_j . Thus,

$$\begin{aligned} \rho_i &= v_{ij} + r_j \mathbf{1}_j + r_i, \\ \rho_j &= v_{ij} + r_j + r_i \mathbf{1}_i, \end{aligned}$$

where $\mathbf{1}_k$ is the indicator random variable for the event C_k , $k = 1, 2, \dots, m$.

Let \mathcal{C}_k denote the next DC to be chosen if this is from S_k , $k = 1, 2, \dots, m$. Thus,

$$\begin{aligned} \frac{E[V(\mathcal{C}_i)] - E[V(\mathcal{C}_j)]}{aV} &= E[\eta^{\rho_i}] - E[\eta^{\rho_j}] \\ &= E[\eta^{\rho_i} - \eta^{\rho_j}] \\ &= E[\eta^{v_{ij} + r_j \mathbf{1}_j + r_i} - \eta^{v_{ij} + r_j + r_i \mathbf{1}_i}] \\ &= E[\eta^{v_{ij}} (\eta^{r_j \mathbf{1}_j + r_i} - \eta^{r_j + r_i \mathbf{1}_i})], \end{aligned}$$

where all expectations on the right-hand side are conditional on B , the event that the target is in principle achievable. Now, noticing that, conditional on B , v_{ij} is independent of \mathcal{C}_i and \mathcal{C}_j , we have

$$E[\eta^{v_{ij}} (\eta^{r_j \mathbf{1}_j + r_i} - \eta^{r_j + r_i \mathbf{1}_i})] = E[\eta^{v_{ij}}] E[\eta^{r_j \mathbf{1}_j + r_i} - \eta^{r_j + r_i \mathbf{1}_i}]$$

and

$$\begin{aligned} E[\eta^{r_j \mathbf{1}_j + r_i} - \eta^{r_j + r_i \mathbf{1}_i}] &= c\eta^{r_i + r_j} + (1 - c)\eta^{r_i} - c\eta^{r_i + r_j} - (1 - c)\eta^{r_j} \\ &= (1 - c)(\eta^{r_i} - \eta^{r_j}), \end{aligned}$$

and this is positive if $r_i < r_j$, since $c < 1$ and $0 < \eta < 1$. Hence, $E[V(\mathcal{C}_i)] - E[V(\mathcal{C}_j)] > 0$ if $r_i < r_j$. Thus, the expected value of the next DC is maximized by choosing it from an LS from which the minimum number of compounds have so far been selected, as required.

Lemma 4. *Given any selection of DCs which includes two LS for which the numbers of DCs selected differ by two or more, there is an alternative selection for which the expected reward is larger.*

Proof. Suppose that the two LS are S_i and S_j . Without loss of generality, it therefore follows from Lemma 2 that after $n - 1$ of the n selections have been made we may assume, in the notation of the proof of Lemma 3, that $r_i < r_j$, and that the last DC is selected from S_j . From Lemma 3, it follows that the expected reward would be greater if this last selection was switched from S_j to S_i , completing the proof.

4.1.3. Optimal policy.

Theorem 3. *The optimal selection policy is to choose compounds from the available LS in as nearly equal numbers as possible.*

Proof. From Lemma 4, it follows that any optimal policy must have the property stated. From Lemma 2, and because of the exchangeability of LS labels, every policy with that property leads to the same TER, completing the proof.

From Theorems 2 and 3, it follows that the maximum TER when n/m is an integer is

$$aV \left(\sum_{j=0}^{n/m-1} \sum_{k=0}^{m-1} (c\eta^{n/m} + 1 - c)^k \eta^j \right).$$

4.2. Model 3 (discounted)

4.2.1. *Background.* For this model, our aim is to choose an infinite sequence of DCs so as to maximize the TEDR. There are potentially an infinite number of LS from which we will choose an infinite number of DCs. The time t_1 taken in looking for the first DC from each LS is longer than the time t_2 taken for each subsequent (*backup*) DC. The probability of success in the search for a first DC from any particular LS is p , $0 < p < 1$. For all searches other than for the first DC from an LS, the probability that we will find a DC is 1. We know from Theorem 1(ii) that there is an optimal policy which is DSM. For such a policy, the next LS from which we select is a deterministic function of the current state. We restrict attention to policies of this form. It follows that a DSM policy is defined by a predetermined infinitely long list of symbols defining the sequence of successive choices. Each of these symbols tells us from which LS to look for the next DC, identifying an LS by the number of DCs which it has so far produced. (There may be more than one LS from which we have selected the same number of DCs. This is a benign ambiguity, as our model behaves in the same way whichever LS is chosen from this equivalent group.) When an LS is selected for the first time it may not yield a DC. As noted in Section 2, in that case a DSM policy selects another LS for the first time, and it continues to do so until one of them does yield a DC. Thus, in this case too our infinite list of symbols defines every choice. We call this list a *policy schedule*.

4.2.2. *Optimal policy derivation.* Together with the notation in Sections 3.1 and 3.2.2, we will use some new notation.

- \star : look for a DC in a new LS and continue doing so with a succession of new LS until we have found a DC.
- i : select the next DC from an LS from which we have so far selected i (>0) DCs.

- (Shorthand notation.) $L(i) = \star 1 2 \dots i - 1$ ($i = \infty$ is possible) and $L(1) = \star$. Thus, every policy schedule starts with $L(i)$ for some i .
- \succ : ‘is preferred to’ or ‘is better than’ in terms of the expected reward. This can be for any two policy schedules, or partial schedules (\succ^{PS}), or in the sense of adjacent interchange (\succ^{AI}), where $i \succ^{\text{AI}} j$ means that the order of selection ij is preferred to ji .
- $X \succ^{\text{PS}} Y$ means that the partial schedule X is preferred to the partial schedule Y . Thus, $i \succ^{\text{AI}} j$ is equivalent to $ij \succ^{\text{PS}} ji$.
- \sim : as ‘ \succ ’, but now referring to no preference between the two policies.
- R denotes the TEDR.
- $R(i)$ denotes the total additional expected discounted reward starting from a point when we have already selected i DCs all from the same LS. Thus, $R = R(0)$.
- p denotes the probability of finding a first DC from an LS, $q = 1 - p$.
- t_1 denotes the time taken to look for a first DC from any LS.
- t_2 denotes the time taken to choose a DC other than the first one from an LS (a backup compound).
- T denotes the time taken from starting to optimize a new LS until finding the first DC in that LS or another new LS under the partial schedule \star .
- G denotes the time which elapses up to the point in a schedule at which we are considering alternative partial schedules.
- M denotes the number of successful DCs prior to the same point in the schedule from LS which are not included in any of the alternative partial schedules.
- γ denotes the discount rate in real terms. (In general, a sum of money has different values on different dates. Using continuous compounding (see [3]), for some γ , £1 now is worth the same as $\text{£}e^{\gamma t}$ after t years. We assume that both these sums of money are measured with the effect of inflation removed, thus, representing actual purchasing power.)
- $V(X)$ denotes the value of policy schedule X .
- $V(ij)$ denotes the value of choosing the i th DC from an LS S_i followed by choosing the j th DC from an LS S_j at some stage in a policy schedule.

Convention. If $L(i)j$ occurs in any policy and $L(i) \sim_{\text{AI}} j$ and $jL(i)$ is a feasible alternative policy then we always choose the $jL(i)$ alternative. The sequence $jL(i)$ will not be feasible if, for example, $L(i)$ is at the start of a policy schedule. Thus, if $L(i)L(i+1) = L(i)L(i)i$ occurs and $L(i) \sim_{\text{AI}} i$ then we choose instead $L(i)iL(i) = L(i+1)L(i)$, leaving the total expected value unchanged.

To derive the optimal policy, we will use the following lemmas.

Lemma 5. *If $i < j$ then $i \succ^{\text{AI}} j$, or, equivalently, $ij \succ^{\text{PS}} ji$.*

Proof. Using Theorem 2,

$$\begin{aligned} \frac{E[V(ij)]}{aV} &= E[e^{-\gamma G}]E[\eta^M](c\eta^{i+j} + (1-c)\eta^i + \exp(-\gamma t_2)(c\eta^{i+j+1} + (1-c)\eta^j)), \\ \frac{E[V(ji)]}{aV} &= E[e^{-\gamma G}]E[\eta^M](c\eta^{i+j} + (1-c)\eta^j + \exp(-\gamma t_2)(c\eta^{i+j+1} + (1-c)\eta^i)), \\ \frac{E[V(ij)] - E[V(ji)]}{aV E[e^{-\gamma G}]E[\eta^M]} &= c\eta^{i+j} + (1-c)\eta^i + \exp(-\gamma t_2)(c\eta^{i+j+1} + (1-c)\eta^j) \\ &\quad - (c\eta^{i+j} + (1-c)\eta^j + \exp(-\gamma t_2)(c\eta^{i+j+1} + (1-c)\eta^i)) \\ &= (1-c)(\eta^i - \eta^j) + (1-c)\exp(-\gamma t_2)(\eta^j - \eta^i) \\ &= (1-c)(\eta^i - \eta^j)(1 - \exp(-\gamma t_2)) \\ &> 0, \end{aligned}$$

as required, since $c < 1$, $i < j$, and $0 < \eta < 1$.

Lemma 6. *If XY is a complete optimal schedule (so that Y is an infinite list), and no LS forms part of both X and Y , then Y is also a complete optimal schedule.*

Proof. We start by introducing some additional notation.

- Ω denotes the set of all policy schedules.
- $\Omega_{\bar{X}}$ denotes the set of all policy schedules which involve none of the LS included in partial schedule X .
- G_X denotes the time taken to complete X .
- r_X denotes the number of successful DCs in X .

Thus, using Theorem 2,

$$E[V(XY)] = E[V(X)] + E[e^{-\gamma G_X}]E[\eta^{r_X}]E[V(Y)].$$

Since XY is optimal, it follows that

$$E[V(Y)] = \sup_{\{H \in \Omega_{\bar{X}}\}} E[V(H)].$$

Since an infinite number of exchangeable LS is available, for any schedule in Ω , there is a schedule in $\Omega_{\bar{X}}$ with the same TEDR. Thus, Y is an optimal schedule, as required.

Lemma 7. *If $L(i) \succ^{AI} j$ then $L(i) \succ^{AI} k$, $k > j$.*

Proof. Applying Theorem 2 to each of the DCs listed in $L(i)j$ we have

$$\begin{aligned} \frac{E[V(L(i)j)]}{aV} &= E[e^{-\gamma G}]E[e^{-\gamma T}]E[\eta^M] \\ &\quad \times ((1 + \eta \exp(-\gamma t_2) + \eta^2 \exp(-2\gamma t_2) \\ &\quad + \dots + \eta^{i-1} \exp(-(i-1)\gamma t_2))(c\eta^j + 1 - c) \end{aligned}$$

$$\begin{aligned}
 &+ \eta^j \exp(-i\gamma t_2)(c\eta^i + 1 - c) \\
 = &E[e^{-\gamma G}]E[e^{-\gamma T}]E[\eta^M] \\
 &\times \left((c\eta^j + 1 - c) \left(\frac{1 - \eta^i \exp(-i\gamma t_2)}{1 - \eta \exp(-\gamma t_2)} \right) + (c\eta^i + 1 - c)\eta^j \exp(-i\gamma t_2) \right).
 \end{aligned}$$

Also, applying Theorem 2 to each of the DCs listed in $jL(i)$ we have

$$\begin{aligned}
 \frac{E[V(jL(i))]}{aV} = &E[e^{-\gamma G}]E[\eta^M] \left(\eta^j \exp(-\gamma t_2) + E[e^{-\gamma T}](c\eta^{j+1} + 1 - c) \right. \\
 &\left. \times \left(\frac{1 - \eta^i \exp(-i\gamma t_2)}{1 - \eta \exp(-\gamma t_2)} \right) \exp(-\gamma t_2) \right).
 \end{aligned}$$

Now let

$$\begin{aligned}
 \Delta = &\frac{E[V(L(i)j)] - E[V(jL(i))]}{aV E[e^{-\gamma G}]E[e^{-\gamma T}]E[\eta^M]} \\
 = &\left(\frac{1 - \eta^i \exp(-i\gamma t_2)}{1 - \eta \exp(-\gamma t_2)} \right) (c\eta^j (1 - \eta \exp(-\gamma t_2)) + (1 - c)(1 - \exp(-\gamma t_2))) \\
 &+ \eta^j \exp(-\gamma t_2) ((c\eta^i + 1 - c) \exp(-(i - 1)\gamma t_2) - (E[e^{-\gamma T}])^{-1}).
 \end{aligned}$$

The factor multiplying η^j is

$$\begin{aligned}
 &c(1 - \eta^i \exp(-i\gamma t_2)) + c\eta^i \exp(-i\gamma t_2) + (1 - c) \exp(-i\gamma t_2) - \frac{\exp(-\gamma t_2)}{E[e^{-\gamma T}]} \\
 = &\underbrace{c + (1 - c) \exp(-i\gamma t_2)}_{<1} - \underbrace{\frac{\exp(-\gamma t_2)}{E[e^{-\gamma T}]}}_{>1} \\
 &< 0.
 \end{aligned}$$

Thus, Δ increases with j , and it follows that if $L(i) \succ^{AI} j$ then $L(i) \succ^{AI} k$, $k \geq j + 1$, as required.

Lemma 8. *The policy $L(\infty)$ is not optimal.*

Proof. The policy $L(\infty)$ means that we choose DCs from one LS for ever. To prove that $L(\infty)$ cannot be optimal, we will show that there is a finite N for which

$$E[V(L(N)L(\infty))] > E[V(L(\infty))].$$

The TEDR under any policy may be written as

$$R = R(0) = \exp(-\gamma t_1)(p(aV + R(1)) + qR(0)),$$

since with probability p , we will find a DC from the first LS we choose to optimize, and with probability $q = 1 - p$, we will not find a DC. Thus,

$$R = \left(\frac{p \exp(-\gamma t_1)}{1 - q \exp(-\gamma t_1)} \right) (aV + R(1)).$$

For the policy $L(\infty)$, using Theorem 2,

$$R(1) = aV \sum_{m=1}^{\infty} \eta^m \exp(-m\gamma t_2). \tag{2}$$

For the policy $L(N)L(\infty)$,

$$R(1) = aV \sum_{m=1}^{N-1} \eta^m \exp(-m\gamma t_2) + \exp(-(N-1)\gamma t_2)(c\eta^N + 1 - c) E[V(L(\infty))]. \tag{3}$$

Note that (2) and (3) differ for $m \geq N$, and, hence, we need to show that

$$\eta^{N-1} aV \sum_{r=1}^{\infty} \eta^r \exp(-r\gamma t_2) = \frac{aV \eta^N \exp(-\gamma t_2)}{1 - \eta \exp(-\gamma t_2)} < (c\eta^N + 1 - c) E[V(L(\infty))].$$

This is true for large enough N , as the left-hand side tends to 0 as $N \rightarrow \infty$ and the right-hand side tends to $(1 - c) E[V(L(\infty))]$ as $N \rightarrow \infty$. Thus, for sufficiently large N , $L(N)L(\infty) \succ L(\infty)$, as required.

In the following we use the notation

$$L(m, n, l) = \underbrace{m \ m \ \dots \ m}_n \underbrace{(m+1) \ (m+1) \ \dots \ (m+1)}_n \underbrace{(m+2) \ (m+2) \ \dots \ (m+2)}_n \dots \underbrace{(m+l) \ (m+l) \ \dots \ (m+l)}_n.$$

Lemma 9. *The infinite partial schedule $L(m, n, \infty)$ is not optimal.*

Proof. We measure time from the point when the partial schedule $L(m, n, \infty)$ starts. We will show that, for all sufficiently large values of m ,

$$E[V(L(\infty))] > E[V(L(m, n, \infty))].$$

This proves the lemma since, for any $k \geq m$, $L(m, n, \infty)$ finishes with $L(k, n, \infty)$.

Now

$$\begin{aligned} E[V(L(m, n, \infty))] &\leq E[\text{total additional value if all further selections are} \\ &\quad \text{from just one LS from which } m \text{ DCs have so far} \\ &\quad \text{been selected, and that LS is good}] \\ &= \alpha V \sum_{\theta \in \Theta} p_{\theta} \eta^{r_{\theta}} \sum_{k=1}^{\infty} \eta^{m+k} \exp(-k\gamma t_2) \quad (\text{using Corollary 1}) \\ &= W \eta^m, \quad \text{where } W > 0 \text{ and does not depend on } m. \end{aligned}$$

Also,

$$E[V(L(\infty))] \geq E[V(L(\infty)) \mid Z] P(Z),$$

where Z is the event {no sampled LS is good except possibly the one sampled by $L(\infty)$ }.

Now recall from (2) that

$$E[V(L(\infty)) \mid Z] = \left(\frac{p \exp(-\gamma t_1)}{1 - q \exp(-\gamma t_1)} \right) \left(\frac{aV}{1 - \eta \exp(-\gamma t_2)} \right) > 0,$$

and note that $P(Z) = c^z$, where z is the number of other sampled LS. Hence, it is sufficient to show that $E[V(L(\infty)) \mid Z]P(Z) > E[V(L(m, n, \infty))]$. This holds if

$$\left(\frac{p \exp(-\gamma t_1)}{1 - q \exp(-\gamma t_1)} \right) \left(\frac{aV}{1 - \eta \exp(-\gamma t_2)} \right) c^z > W\eta^m.$$

This is true for all sufficiently large m , as the left-hand side is greater than 0 and does not depend on m and the right-hand side tends to 0 as $m \rightarrow \infty$. This completes the proof.

Lemma 10. *If the infinite partial schedule D does not include \star then D is not optimal.*

Proof. From Lemma 5, it follows that if D is optimal and does not include \star then it must finish as $L(m, n, \infty)$ for some m and n . However, from Lemmas 8 and 9, we know that $L(m, n, \infty)$ is not optimal, completing the proof.

From Lemma 10, it follows that any optimal schedule may be written in the form $L(i_1)L(i_2)D_2L(i_3)D_3 \dots L(i_k)D_k \dots$, where (i_k) is an infinite sequence and $D_k, k = 2, 3, \dots$, is a partial schedule which does not include \star , does not start with i_k , cannot be infinite, and may be \emptyset .

Lemma 11. *If an optimal policy has a partial schedule of the form $L(i)DL(j)$, where D is a partial schedule which does not include \star , does not start with i , and is not \emptyset , then $j < i$.*

Proof. Consider a partial schedule of the form $L(i)DL(j)$, and suppose that $j \geq i$. Now,

$$L(i)DL(j) = L(i) \underbrace{f \dots m}_{=D} L(i) i (i + 1) \dots (j - 1),$$

where $m \geq f$ (by Lemma 5) and $fL(i)$ is a feasible alternative to $L(i)f$ as $f \neq i$. However, it is impossible to have $L(i) \succ^{AI} f$ and $m \succ^{AI} L(i)$ (by Lemma 7). This is a contradiction, and it follows that $j < i$ as required.

In the following we use the notation

$$\underbrace{L(i)L(i) \dots L(i)}_{j \text{ times}} = L(i, j).$$

When $j = \infty$, we have a cyclic policy $L(i, \infty)$.

Lemma 12. *Any optimal schedule which is not cyclic either*

- (a) *has a start of the form $L(i, h)L(i + 1)i$ for some i and h ; or*
- (b) *includes a partial schedule of the form $Q = L(i)L(j)x$, where $j < i$ and $x \neq j$.*

Proof. Assume that the optimal policy starts with $L(i, h)L(j)x$, where $h \geq 1, j \neq i$, and $x \neq j$. It is easy to see that any acyclic policy schedule must start in this way. We will consider the following three cases, one of which must occur.

Case 1: $j < i$. We have a partial schedule $L(i)L(j)x$, which is of the form Q .

Case 2: $j > i + 1$. We have a partial schedule of the form $L(i)L(j)x = L(i) L(i) i (i + 1) \dots (j - 1) x$. The possible values of x are either \star (choose a DC from a new LS) or i . If $x = i$, we have a partial schedule of the form

$$L(i)L(j)x = L(i) L(i) i (i + 1) \dots \underline{(j - 1)} i.$$

However, $i \succ^{AI} j - 1$ (Lemma 5) as $j - 1 > i$, so this is impossible. Hence, $x = \star$ and we have a partial schedule of the form $L(i)L(j)L(k)y, y \neq k$. If $k \geq i$, this may be written as

$$L(i) \underline{L(i)} i (i + 1) \dots \underline{(j - 1)} L(i) i (i + 1) \dots (k - 1) y.$$

The underlined parts of the schedule contradict Lemma 7 as $j - 1 > i$. Hence, we must have $k < i$ which gives a partial schedule $L(j)L(k)y$, where $k < j$, which is of the form Q .

Case 3: $j = i + 1$. We have a partial schedule of the form $L(i)L(j)x$, where x can be equal to either i or \star . If $x = i$, the partial schedule has the form $L(i)L(i + 1)i$ and the optimal schedule is as in (a). If $x = \star$, the schedule starts as $L(i)L(i + 1)L(y)z$, where $y < i$ so as not to contradict Lemma 7 as in case 2. Here $L(i + 1)L(y)z$ is of the form Q .

This completes the proof.

Lemma 13. *It is impossible for the sequential pairs $L(i)L(j)$, where $i > j$, and $L(k)L(m)$, where $m > k$ and $k \leq j$, to both occur in an optimal schedule.*

Proof. Let an optimal policy schedule contain both sequential pairs $L(i)L(j), i > j$, and $L(k)L(m), m > k$ and $k \leq j$. The optimal schedule is of the form

$$\begin{aligned} &\dots L(i)L(j) \dots L(k)L(m) \dots \\ &= \dots \underbrace{L(i - 1) (i - 1)}_{=L(i)} \underbrace{L(k) k \dots (j - 1)}_{L(j)} \dots L(k) \underbrace{L(k) k \dots (m - 1)}_{L(m)} \dots \end{aligned}$$

For this to be optimal, it follows from our convention that $i - 1 \succ^{AI} L(k)$ and $L(k) \succ^{AI} k$. However, $k \leq i - 1$, so this contradicts Lemma 7, completing the proof.

Lemma 14. *If an optimal policy includes the partial schedule $Q = L(i)L(j)x$ with $j < i$ and $x \neq j$, then it must finish with $L(k, \infty)$ for some k .*

Proof. First note that it follows from Lemma 13 that if an optimal policy includes Q then it does not also include $L(k)L(m)$ with $m > k$ and $k \leq j$. This observation together with Lemma 11 implies that the sequence (i_r) is decreasing for $r \geq l$, where $i_l = i$ and $i_{l+1} = j$. Since (i_r) is a sequence of nonnegative integers it follows that, for some k and $N, i_r = k$ for all $r \geq N$. Because of the strict inequality in Lemma 11, it also follows that $D_s = \emptyset, s \geq r$, and the optimal schedule therefore finishes with $L(k, \infty)$, as required.

Lemma 15. *If an optimal policy schedule starts with $L(i, h)L(i + 1)i$ then it must finish with $L(k, \infty)$ for some k .*

Proof. Suppose that an optimal policy schedule starts as $L(i, h)L(i + 1)i$. It must continue as $L(i, h)L(i + 1)DL(k) \dots$, where $k < i + 1$ (Lemma 11) and D is a finite partial schedule starting with i , and not including \star .

We next note that it is impossible for the optimal policy to include $L(m)L(n)$ with $m \leq k$ and $n > m$, because this would mean that we had both $lL(m), l \geq i$, and $L(m)L(m)m$ as

partial schedules. This would contradict our convention (if $m = l$), or Lemma 7 (if $m < l$) as shown underlined below.

$$L(i, h)L(i + 1) \underbrace{i \dots l}_{=D} \underline{L(m)} m \dots k - 1 \dots \underline{L(m)} \underline{L(m)} m \dots (n - 1) \dots$$

This observation, together with Lemma 11, means that the sequence

$$\underbrace{i_{h+1}, i_{h+2}, i_{h+3}, \dots}_{(=i+1)}$$

is decreasing. The rest of the proof is the same as for Lemma 14.

4.2.3. Optimal policy.

Theorem 4. *There is an optimal policy schedule which is cyclic.*

This is an immediate consequence of Lemmas 6, 12, 14, and 15.

Using Theorem 2, it may be shown that the TEDR for a cyclic policy $L(k, \infty)$ is

$$\frac{aVp \exp(-\gamma t_1)(1 - (\eta \exp(-\gamma t_2))^k)}{(1 - \eta \exp(-\gamma t_2))(1 - \exp(-\gamma t_1)(q + p \exp(-(k - 1)\gamma t_2)(c\eta^k + 1 - c))}$$

The derivation is given by Charalambous [4], who also investigated numerically the relationships between the optimal TEDR and k and the various parameters of our model.

5. Conclusions

In this paper we investigated the optimal allocation of resources throughout a pre-clinical pharmaceutical research project. It is based on the model and software described by Yu and Gittins [15] (Model 1). We derived optimal selection schedules for a sequence of DCs using two restricted versions of the basic model. This was done for the undiscounted finite horizon case (Model 2), and for the more realistic discounted infinite horizon case (Model 3). For Model 2, a finite number of LS is available and there is no setup cost which penalizes switches between LS; it turns out that optimal policies select as nearly as possible the same number of DCs from each LS. For Model 3, an infinite number of LS is available and there are additional costs in switching to a new LS; it turns out that optimal policies never switch back to an earlier LS once a switch has been made from that LS.

The setup for Model 1 is more complex, but it seems likely that if we again assume that an infinite number of LS is available, an optimal policy will progress from one series to another with no switching back, as for Model 3. If, on the other hand, there is just a finite number of promising LS available, it seems likely that after choosing DCs from one LS after another with no switching back, as for Model 3, an optimal policy then moves to a phase in which one DC is selected from each LS in turn, as for Model 2. Policies along these lines are being investigated numerically for Model 1, and will be reported separately.

Further plans involve taking account of the feedback that is provided from clinical trials on a DC in the future selection of DCs. As this information comes in, the scientists reevaluate what to do next. They must decide whether to carry on with the selection schedule as was originally planned, stop the project immediately, or work on an alternative selection schedule. For example, a DC might turn out to have unfortunate side effects which would significantly reduce the expected value of the next DC if it were to be taken from the same LS. This might

cause a modification in the original selection schedule. A Bayesian formulation will be needed to carry out this plan.

References

- [1] BERGMAN, S. W. AND GITTINS, J. C. (1985). *Statistical Methods for Pharmaceutical Research Planning*. Marcel Dekker, New York.
- [2] BLACK, F. AND SCHOLES, M. (1973). The pricing of options and corporate liabilities. *J. Political Econom.* **81**, 637–659.
- [3] BREALY, R. A. AND MYERS, S. C. (2000). *Principles of Corporate Finance*, 6th edn. McGraw-Hill, New York.
- [4] CHARALAMBOUS, C. (2006). Optimal selection policy for a sequence of candidate drugs. Doctorate Thesis Transfer of Status Report, Department of Statistics, University of Oxford.
- [5] CHEN, B. P. K. (2004). Prioritization of research projects in the pharmaceutical industry. Doctoral Thesis, Department of Statistics, University of Oxford.
- [6] GITTINS, J. C. (1996). Quantitative methods in the planning of pharmaceutical research. *Drug Inf. J.* **30**, 479–487.
- [7] GITTINS, J. C. (1997). Why crash pharmaceutical research? *R&D Manag.* **27**, 79–85.
- [8] GRAY, N. (2005). *PharmExec 50: Our Sixth Annual Report of the World's Top 50 Pharma Companies*. Available at http://bio.cocoonworks.com/wp-content/uploads/2006/03/2005_Global_Top_50_Pharmaceutical_Companies.pdf.
- [9] HALLIDAY, R. G., DRASDO, A. L., LUMLEY, C. E. AND WALKER, S. R. (1997). The allocation of resources for R&D in the World's leading pharmaceutical companies. *R&D Manag.* **27**, 63–77.
- [10] JACOB, W. AND KWAK, Y. H. (2003). In search of innovative techniques to evaluate pharmaceutical R&D projects. *Technovation* **23**, 291–296.
- [11] MILLER, P. (2005). Role of pharmaeconomic analysis in R&D decision making. *Pharmacoeconomics* **23**, 1–12.
- [12] POH, K. L., ANG, B. W. AND BAI, F. (2001). A comparative analysis of R&D project evaluation methods. *R&D Manag.* **31**, 63–75.
- [13] ROSS, S. M. (1970). *Applied Probability Models with Optimization Applications*. Holden-Day, San Francisco, CA.
- [14] WHITE, D. J. (1993). *Markov Decision Processes*. John Wiley, Chichester.
- [15] YU, J. Y. AND GITTINS, J. C. (2007). Models and software for improving the profitability of pharmaceutical research. *Europ. J. Operat. Res.* **189**, 459–475.