A Stochastic Programming Approach for Clinical Trials Planning

Matthew Colvin, Christos T. Maravelias

Department of Chemical & Biological Engineering, University of Wisconsin-Madison, 1415 Engineering Dr, Madison, WI 53706, USA

Abstract

We present a multi-stage stochastic programming formulation for the planning of clinical trials in the pharmaceutical research and development (R&D) pipeline. Using a scenario-based approach the discrete nature of the uncertainty in clinical trials can be modeled without loss of information. Given a portfolio of potential drugs and limited resources, the model finds the optimal timing to maximize the expected net present value. Ideas are presented to reduce the size of the formulation, focusing on the non-anticipativity constraints required to model indistinguishable scenarios.

Keywords: stochastic programming; pharmaceutical research and development; optimization under uncertainty.

1. Introduction and Background

The pharmaceutical industry has been undergoing change due to rapid growth and changes in the managed-health care environment. The productivity of research and development (R&D) pipelines, in terms of new entities registered per dollar of investment, is in decline, while effective patent lives are shorter and barriers to entry during active patents are lowered [1]. Therefore, it is imperative for pharmaceutical companies to manage their R&D pipelines to lower costs and improve throughput [2, 3]. This is a challenging task due to the stochastic nature of the R&D process: if a drug fails any clinical trial, its development is stopped and all prior investment is lost; if it passes all trials, it enters the marketplace where profits are usually significantly larger than development costs. Methods applied to this problem include deterministic models using expected values, simulation-based approaches and dynamic programming methods [4, 5, 6]. The goal of this paper is to develop a computationally tractable multi-period stochastic programming (SP) formulation that accounts for uncertainty in clinical trial outcomes.

1.1. Pharmaceutical Research and Development

For each drug that enters the marketplace, thousands of compounds are tested and over $900 million are typically spent on R&D over a period of 10-15 years [3]. To maintain a steady stream of new drugs to market, it is therefore necessary to have a number of candidate products in the pipeline at all times. However, the number of drugs in testing at any given time is limited by the availability of key resources. Given a portfolio of potential drugs, one then has to prioritize them and decide how to allocate scarce resources among them.

While uncertainty exists in cost, duration, resource requirements and revenue from sales, outcome of the clinical trials is the most significant source of uncertainty for the development process, in large part due to the *all or nothing* nature of the approval process. The goal of a planning method is to optimize an economic metric such as the

expected net present value of the candidate portfolio. Minimization of weighted time to market, resource utilization and metrics of volatility and risk are also considered [2].
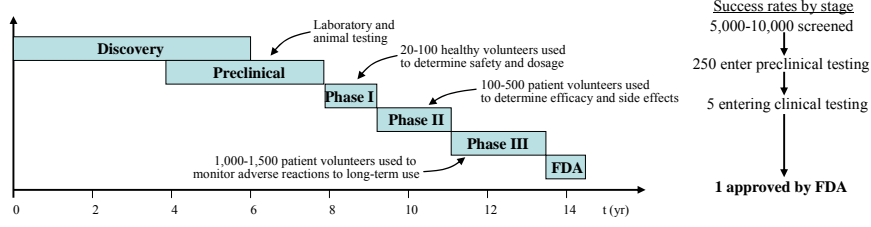

Figure 1: Stages of Pharmaceutical Product Development [3].

1.2. Multistage Stochastic Programming

The two basic elements of scenario-based multi-stage stochastic programming (SP) formulations are: scenarios representing the possible realizations of uncertainty and the definition of stages. The stages typically represent the discretization of the planning horizon into time periods, $t \in T$, while for discrete uncertainty (such as explored in this problem) each scenario, $s \in S$, represents a unique combination of the realizations of all uncertain events.

At $t = 1$, the decision-maker has no information about uncertainty, which means that all scenarios are indistinguishable and all decisions made must be identical. As trials are performed, uncertainty is revealed, scenarios become distinguishable and the decision-maker can take recourse actions to account for this. If we assume that scenarios s and s' are indistinguishable before time $t^{s,s'}$, it is necessary to force the optimization decisions x in scenarios s and s' to be identical at all times prior to $t^{s,s'}$:

$$\left\{ t < t^{s,s'} \right\} \Rightarrow \left\{ x_{ts} = x_{ts'} \right\} \ \forall s, s', t \tag{1}$$

To develop a mathematical programming formulation, we need to convert the logic condition in eqn (1) into mixed-integer constraints. If stage $t^{s,s'}$ is known in advance, this is fairly simple to do, but in the R&D pipeline, this depends on choices made by the decision-maker. These constraints are usually enforced through the introduction of a binary variable $y_{tss'}$ for each pair (s,s') and stage t. They are referred to as non-ancticipativity constraints and thorough discussion can be found in [7].

2. Proposed Approach

2.1. Problem Statement

The formal problem statement of the problem we consider in this paper is as follows:
Given,

  i)   a fixed time horizon divided into uniform time periods $t \in T = \{1,2,\dots,|T|\}$
  ii)   a set of candidate drugs $i \in I$ with known potential revenues
  iii) a set of resource types $r \in R$ with fixed availability ρ_r^{max}
  iv)   a set of clinical trials $j \in J = \{PI, PII, PIII\}$ for each drug $i \in I$, with probability of success $\hat{p}_{ij}$, deterministic cost c_{ij}, duration τ_{ij}, and resource requirements ρ_{ijr}

determine which clinical trials should be carried out and when in order to maximize the expected net present value of the R&D pipeline.

We assume that the revenue Rv_i from sales of drug i is a linearly decreasing function of the delay, D_i, in the development and completion time, L_i, of PIII (shown in Figure 2) and that there are no interdependencies regarding the success probabilities of drugs.
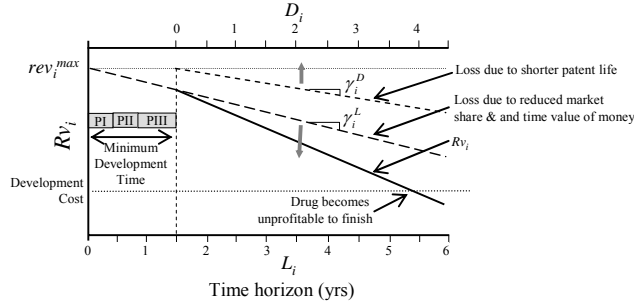
Figure 2: Revenue as a function of time

Since positive cash flows occur only at the end of PIII and we consider a finite time horizon, solutions with empty pipelines towards the end of the horizon will appear which would not be optimal in a real situation. To address this problem, we assign a weighted expected revenue to drugs that have not failed any tests, but still are not completed by the end of the time horizon.

2.2. Scenario Representation
The uncertainty considered in this work is discrete: a drug either passes or fails a clinical trial. In other words, the outcome of clinical trial j of drug i, can be viewed as a discrete random variable with sample space Ω_{ij} = {pass, fail}. A naïve approach would be to define eight outcomes for each drug, based on passing or failing each trial. However, a number of these scenarios would never be distinguished as after the failure of trial, no other trials are performed. Therefore it is possible to aggregate these outcomes into four events per drug based upon when the drug first fails a trial or successfully navigates all trials. Thus, uncertainty is represented via a single uncertain parameter per drug $i \in I$ with revised sample space Ω_i = {I-F,II-F,III-F,III-P}.

2.3. Model Reductions
Our approach exploits the following five ideas to reduce the number of non-anticipativity constraints and binary variables (see [8] for details):
1) The number of scenarios necessary to represent uncertainty in the outcome of clinical trials of |I| drugs can be reduced from $8^{|I|}$ to $4^{|I|}$.
2) By definition, if pair (*s,s'*) is indistinguishable at stage *t*, so is (*s',s*).
3) <u>Property 1</u>: It is sufficient to express non-anticipativity constraints only for pairs of scenarios that differ in the outcome of only one drug.
4) <u>Property 2</u>: It is sufficient to express non-anticipativity constraints only for scenarios that differ in the outcome of a single clinical trial, i.e. (*s,s'*) represent consecutive elements of Ω_i.
5) The logic condition that links scenarios (*s,s'*) can be expressed in terms of an existing decisions binary variable; i.e. no new binary variables $y_{ss't}$ need to be introduced.

Property 1 is based on the work of Goel and Grossmann [7]. Property 2 is applicable only in problems where the sequence of uncertainty realization for some parameters is known. In this case all clinical trials must be carried out in order (PI > PII > PIII).

We next define set Ψ to be the set of scenario pairs over which non-anticipativity constraints must be enforced. For a scenario pair $(s,s') \in \Psi$, there exists a single trial $(i^{s,s'}, j^{s,s})$ at which they become distinguishable. If X_{ijts} indicates whether trial *j* of drug *i* is being started at time *t* in scenario s, it is possible to express the logic condition for the non-anticipativity constraints in eq (1) using variable X_{ijts} for $i=i^{s,s'}$, $j=j^{s,s'}$. Implementing

the above reductions to a 3-drug 12-period problem, we can reduce the cardinality of Ψ from 261,632 to 141, and the number of non-anticipativy constraints from approximately 37 million to 31,104, while not introducing additional binary variables.

3. Mathematical Formulation

3.1. Optimization Variables
To express non-anticipativity constraints, we have to keep track of when the differentiating trial $(i^{s,s'},j^{s,s'})$ is finished. To facilitate this, we introduce variable Y_{ijts} that is equal to 1 if task (i,j) is finished by the beginning of period t in scenario s:

$$Y_{ijts} = Y_{i,j,t-1,s} + X_{i,j,t-\tau_{ij},s} \quad \forall i,j,t,s \tag{2}$$

As mentioned in section 2.1, it is necessary to assign revenue to drugs that have not gone through all trials, but also have not failed any trials by the end of the planning horizon. We define variable $Z_{ijts} = 1$ if trial $(i,j-1)$ is finished by time t and trial (i,j) has not started at or before time t in scenario s:

$$Z_{i,PI,t,s} = 1 - \sum_{t' \leq t} X_{i,PI,t',s} \quad \forall i,s,t \tag{3}$$

$$Z_{ijts} = Z_{i,j,t-1,s} + X_{i,j-1,t-\tau_{ij-1},s} - X_{ijts} \quad \forall i,j \in \{PII,PIII\},t,s \tag{4}$$

Note that both Y_{ijts} and Z_{ijts} can be calculated in terms of variables X_{ijts} and thus do not have to be defined as binaries.

3.2. Non-anticipativity Constraints
As mentioned in sections 2.3 and 3.3, non-anticipativity constraints for $(s,s') \in \Psi$ must be active until the differentiating trial $(i^{s,s'},j^{s,s'})$ has finished, which can be expressed as:

$$\{Y_{i^{s,s'},j^{s,s'},t,s} = 0\} \Rightarrow \{X_{ijts} = X_{ijts'} \quad \forall i,j\} \quad \forall t,(s,s') \in \Psi$$

This can be converted into the following constraints:

$$-Y_{i^{s,s'},j^{s,s'},t,s} \leq X_{ijts} - X_{ijts'} \leq Y_{i^{s,s'},j^{s,s'},t,s} \quad \forall i,j,(s,s') \in \Psi,t>1 \tag{5}$$

$$X_{i,PI,1,s} = X_{i,PI,1,1} \quad \forall i,s \tag{6}$$

3.3. Sequencing and Resource Constraints
Equation (7) enforces that profit generating trials are not run multiple times:

$$\sum_t X_{ijts} \leq 1 \quad \forall i,j,s \tag{7}$$

Equation (8) keeps trial (i,j) from starting until trial $(i,j-1)$ is completed:

$$\sum_{t' \leq t} X_{ijt's} \leq Y_{i,j-1,t,s} \quad \forall i,j \in \{PII,PIII\},t,s \tag{8}$$

Resource constraints are expressed as follows:

$$\sum_i \sum_j \sum_{t-\tau_{ij}<t' \leq t} \rho_{ijr} X_{ijt's} \leq \rho_r^{max} \quad \forall r,t,s \tag{9}$$

In a given scenario s, there are clinical trials that cannot be performed because a previous trial fails. Thus, if $F(s){\subset}IxJ$ is the subset of trials (i,j) that cannot be carried out in scenario s, we can add:

$$X_{ijts} = 0 \quad \forall t,s,(i,j) \in F(s) \tag{10}$$

3.4. Objective Function

The total development cost Cst_s in scenario s is calculated in eqn (12), where cd_t is a time discounting factor:

$$Cst_s = \sum_{i,j,t} cd_t c_{ij} X_{ijts} \quad \forall s \tag{11}$$

If we define $S^I(s){\subset}I$ as the subset of drugs that can successfully pass PIII in scenario s, we calculate the revenue Rv_s from the completion of all trials:

$$Rv_s = \sum_{i \in S^I(s)} \sum_t \left\{ rev_i^{max} X_{iPIIIts} - \gamma_i^D (Z_{iPIIts} + Z_{iPIIIts}) - \gamma_i^L (t + \tau_{iPIII}) X_{iPIIIts} \right\} \forall s \tag{12}$$

The revenue FRv_s that can be materialized in the future is calculated as:

$$FRv_s = \sum_{i \in S^I(s)} \sum_j rev_{ij}^{open} f_{ij} Z_{ij|T|s} + \sum_{i \in S^I(s)} \sum_{j \in \{PI,PII\}} \sum_{t>|T|-\tau_{ij}} rev_{ijt}^{run} f_{ij+1} X_{ijts} \quad \forall s \tag{13}$$

where rev_{ij}^{open} and rev_{ijt}^{run} are estimates of the revenue that would be realized if open drug i were completed as quickly as possible after the time horizon and f_{ij} is a discounting factor to encourage the completion of clinical trials.

If p_s is the probability of scenario s occurring, the objective function is:

$$\max ENPV = \sum_s p_s (Rv_s + FRv_s - Cst_s) \tag{14}$$

with $X_{ijts} \in \{0,1\}$, Y_{ijts}, $Z_{ijts} \in [0,1]$ (15)

The proposed SP formulation **M1** consists of eqs (2)-(15).

3.5. Tightening Constraints

Even with the dramatic reduction in constraints and binary variables, **M1** is still a large MIP formulation that grows exponentially in the number of drugs. By exploiting the structure of the problem, we can express a subset of non-anticipativity constraints as equalities:

$$X_{i^{s,s'}jts} = X_{i^{s,s'}jts} \quad \forall j \le j^{s,s'}, t > 1, (s,s') \in \Psi \tag{16}$$

$$X_{ijts} = X_{ijts'} \quad \forall i \ne i^{s,s'}, j, t < \sum_{j'<j} \tau_{ij'}, (s,s') \in \Psi \tag{17}$$

$$-Y_{i^{s,s'},j^{s,s'},t,s} \le X_{ijts} - X_{ijts'} \le Y_{i^{s,s'},j^{s,s'},t,s}, \forall i \ne i^{s,s'}, j, t > \max\{\sum_{j'<j} \tau_{ij'}, \sum_{j' \le j^{s,s'}} \tau_{i^{s,s'}j'}\}, (s,s') \in \Psi \tag{18}$$

The reformulated model **M2** with eqs (16)-(18) instead of eq (5) has fewer constraints and smaller integrality gap (see Table 1). Formulation **M2** can be used to solve problems with up to five drugs entering clinical trials on a standard desktop computer. Of the non-anticipativity constraints that remain as inequalities in eq (18), only a small fraction are active. We are currently developing a branch-and-cut algorithm that exploits this, potentially allowing more challenging problems to be solved.

Table1: Model statistics for **M1** and **M2** and solution statistics for **M2**.

Drugs	Variables	Equations 2-4, 6-15	Non-anticipativity		Integr. Gap (%)		Solution statistics	
			M1	**M2**	**M1**	**M2**	Nodes	CPU s
3	14,365	14,209	23,470	18,284	7.5	4.2	6	62.4
4	51,201	51,500	111,612	93,180	5.6	2.2	6	70.7
5	255,489	261,533	687,200	595,040	5.3	2.2	16	2,215.5

4. Example

Three products are to be tested using two resource types. The planning horizon is 36 months divided into 12 3-month periods. The optimal solution has an ENPV of $840.1M. The probability density functions of NPV (64 scenarios) and a timeline of where scenario decisions differ are shown in Fig 3. The proposed formulation was modeled in GAMS 22.4 and solved using CPLEX 10.0 on a 2.8 GHz Pentium4 with 512 MB of RAM. The example was generated in 17 sec and solved in 25 sec.

Table2: Data for example.

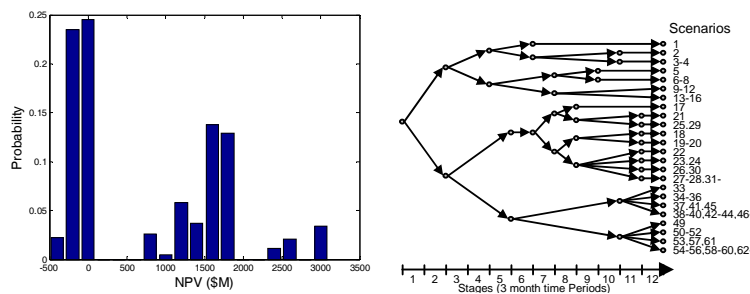| Drug	Duration (months)			Probability			Cost ($1M)			Resource 1 max=4			Resource 2 max=4			rev^{max} ($1M)
	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII	Rmax
D1	6	12	12	0.50	0.60	0.70	10	100	220	1	2	3	2	2	3	2,500
D2	6	9	15	0.45	0.65	0.70	10	80	200	2	2	3	1	2	3	2,400
D3	6	9	12	0.50	0.50	0.80	10	70	210	1	2	3	1	3	3	2,650

Figure 3. Probability density functions of NPV and decision tree for example.

References

1. C. Robins-Roth, 2001, *From Alchemy to IPO: The Business of Biotechnology*, Perseus.
2. N. Shah, 2004, Pharmaceutical Supply Chains: Key Issues and Strategies for Optimization, *Comput. Chem. Eng.*, 28, 929-941.
3. Tufts Center for the Study of Drug Development, 2001.
4. C. T. Maravelias and I. E. Grossmann, 2004, Optimal Resource Investment and Scheduling of Tests for New Product Development. *Comput. Chem. Eng.*, 28, 1021-1038.
5. G. E. Blau, J.F. Pekny, V.A. Varma and P.R. Bunch, 2004, Managing a portfolio of interdependent new product candidates in the pharmaceutical industry. *Journal of Product Innovation Management*, 21(4), 227-245.
6. J. Choi, M.J. Realff and J.H. Lee, 2004, Dynamic Programming in a Heuristically Confined State Space: A Stochastic Resource-constrained Project Scheduling Application. *Comput. Chem. Eng.*, 28 (6-7), 1039-1058.
7. V. Goel and I.E. Grossmann, 2006, A class of stochastic programs with decision dependent uncertainty, *Math. Program.*, 108 (2-3): 355-394.
8. M. Colvin and C. T. Maravelias, 2007, A Stochastic Programming Approach for Clinical Trial Planning in New Drug Development, *Comput. Chem. Eng.*, in press (2007).