

RESEARCH

Open Access



Pharmaceutical portfolio optimization under cost uncertainty via chance constrained-type method

Maboubeh Farid^{1*} , Abraar Chaudhry², Magnus Ytterstad¹ and Stig Johan Wiklund¹

*Correspondence:

maboubeh.farid@Captario.com

¹Captario AB, Gothenburg, Sweden

Full list of author information is available at the end of the article

Abstract

Project selection for a portfolio is a pivotal decision in the pharmaceutical industry. In this paper, we study a portfolio optimization problem for pharmaceutical companies considering the uncertainty of the cost of each phase of drug development and the specific value of the annual budget. The presented optimization model is suitable to make investment decisions for multi-phase drug development projects and a stochastic approach is applied to handle the uncertainty in the model. Post-optimality analysis for annual budget is studied. An illustrative example is included to demonstrate the presented approach.

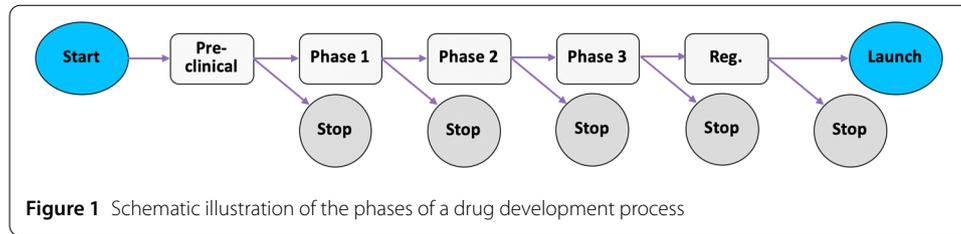
Keywords: Portfolio optimization; Stochastic approach; Chance constraint; Pharmaceutical industry

1 Introduction

Selecting projects for a portfolio is one of the major decision-making problems in portfolio management [1] and decision-making is a core function of drug development in the pharmaceutical industry. Drug development in a pharmaceutical pipeline is a well-defined and long-term process which must proceed through several crucial phases and is connected with timing. Drug development can be described as a sequence of phases (Fig. 1).

In project management, decisions must be made at key points throughout the development process, and whether to continue or terminate a particular project is a pivotal decision. To state the obvious, the cost of development will not be refunded if the project fails to reach market, and all phases must be sequentially accomplished before revenue is realized. Therefore, allocating funds to projects with greater potential revenue and higher probabilities of success in each phase is a major priority for portfolio managers. According to a study from 2013, the cost of new medicine was estimated at \$ 2.5 billion [2], and on average only 5%–10% of potential drugs will successfully pass through all phases of drug development [3]. Given that, any strategy for successfully achieving portfolio optimization in the pharmaceutical industry should consider long-term survival the fundamental driving force [4]. There are, however, other factors to consider when selecting a drug can-

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



didate for a portfolio, for example success risk, expected revenue, development costs, and budget allocation.

Since the execution of a drug development project is exceptionally costly, the company may not have sufficient R&D resources to pursue all available projects. Consequently, a selection needs to be made of which projects to fund. Optimization techniques have been widely used to improve decision-making for portfolios and an appropriate optimization model can efficiently improve the procedure of selecting projects. Several mathematical models have been developed in the form of linear, nonlinear, multi-objective optimization, etc. for solving portfolio optimization problems [5, 6]. A binary multi-objective linear programming was presented by Ghasemzadeh et al. [7], and different approaches have been presented which take into an account the uncertainty of data [8–10].

A key to achieving successful portfolio optimization is to capture the uncertainty of important parameters and also the factors that affect them. As such, mathematical models built on deterministic optimization are not suitable approaches for portfolio optimization since they cannot account for uncertainty. An obvious source of uncertainty is the cost of running the projects through the development process. To account for this uncertainty, stochastic programming may be applied as an approach for modeling optimization problems that involve uncertainty with random parameters. The stochastic framework provides a roadmap for decisions when information about the future may be subjected to significant uncertainty. Stochastic programming is among the most difficult problems in the field of mathematical modeling and finding an exact solution to these problems is quite challenging. Chance constrained (CC) programming was presented by Charnes et al. [11]. It is a quintessential technique that can handle the uncertainty inherent in portfolio optimization, in which the constraints in the model cannot be violated with more than a predefined level of probability.

In this study, our aim is to develop an approach in the framework of chance constrained programming that enables us to select projects for a portfolio that maximizes expected revenue and addresses uncertainty while remaining computationally efficient. We will formulate the problem as an abstract chance constrained program and present an approach to convert it into a form that can be tractably handled by modern optimization solvers. Furthermore, we will employ parametric programming for our mathematical modeling to investigate how the optimal value of the expected revenue and optimal project selection depend on the annual budget.

The rest of the paper is organized as follows: Sect. 2 presents the background of mathematical optimization and in Sect. 3, we outline the structure of the data underlying the optimization problem. In subsequent sections we introduce the mathematical formulation of the problem and the approaches employed for solving the problem. An illustrative example of the proposed approach is included, and a discussion section concludes the paper.

2 Mathematical background

The general mathematical model of optimization problem is:

$$\begin{aligned} & \text{maximize} && f(x) \\ & \text{subject to} && G_i(x) \leq 0, \quad i = 1, \dots, m, \end{aligned} \quad (1)$$

where f is the objective function, G is the inequality constraints and x is the vector of decision variables.

Stochastic programming has been developed to model optimization problems where coefficients of decision variables, in the objective function and constraints, are under uncertainty. To handle uncertainty in the optimization problem, the stochastic information is combined with the model of the optimization problem (1). The general mathematical model of the optimization problem with uncertain coefficients can be written as follows:

$$\begin{aligned} & \text{maximize} && f(x, \zeta) \\ & \text{subject to} && G_i(x, \zeta) \leq 0, \quad i = 1, \dots, m, \end{aligned} \quad (2)$$

where ζ is the vector of coefficients' uncertainties.

Several approaches have been proposed for solving (2). In 1959, Charnes and Cooper [11] proposed chance constrained programming and it has received a great deal of attention for solving optimization problems with uncertain coefficients in constraints G . In chance constrained programming, the inequality of constraints G in (2) must be satisfied with the predefined confidence level and these constraints are called chance constraints. Furthermore, an optimization problem which employs chance constrained programming is called a chance constrained optimization problem. In the following, we describe the general mathematical model of a chance constrained optimization problem:

$$\begin{aligned} & \text{maximize} && f(x) \\ & \text{subject to} && \text{Prob}\{G_i(x, \zeta) \leq 0, i = 1, \dots, m\} \geq (1 - \alpha), \end{aligned} \quad (3)$$

where $\alpha \in (0, 1)$. The model (3) searches to find the optimal decision vector that maximizes the objective function f subject to the constraints being satisfied with probability at least $(1 - \alpha)$.

In chance constrained programming, inequalities $G_i(x)$ ($i = 1, \dots, m$) can be handled in two ways: individually and jointly. In individual chance constrained programming, it would be guaranteed that each constraint is satisfied with its own confidence level. In joint chance constrained programming, all constraints will be satisfied with a certain confidence level. Joint chance constrained programming is more robust than the individual version but it is significantly more difficult to solve as the problem may become non-convex (see [12–16]). Generally, chance constrained programming is powerful when modeling optimization problems under uncertainty but solving the problems can be challenging in real applications. Chance constrained programming is widely used in practice, including for water management [17], chemical processes [18], energy management [19], and supply chain planning [20].

If we assume the sample space is finite and that G is bounded, we can change (3) into an integer program [21]. By employing binary control variables z_k , ($k = 1, 2, \dots, N$) and

“big- M ” coefficients, the joint chance constrained program can be reformulated as a set of linear constraints as follows:

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && G_i(x, \zeta) - Mz_k \leq 0, \quad i = 1, \dots, m, k = 1, 2, \dots, N; \\ & && \sum_{k=1}^N z_k \pi_k \leq \alpha, \end{aligned} \tag{4}$$

where $M > 0$ is a large constant and $\alpha \in (0, 1)$. There are N scenarios and π_k , ($k = 1, 2, \dots, N$) is the probability of each scenario. When $z_k = 0$, the related constraint must be satisfied, otherwise when $z_k = 1$, the constraint can be violated and big value of M ensures that it would be an inactive constraint. The constraints in (4) are called Big- M constraints. Even when we do not have full knowledge of the sample space, (4) can still be a useful proxy for (3). When combined with Monte Carlo sampling, this is known as the sample average approximation (SAA) and its statistical properties are well studied [22, 23].

For simplicity, we call (4), CC-MBP. The CC-MBP (4) with binary decision variables is an integer programming problem that can be solved by integer programming solvers, as will be further described in later parts of this article.

3 Input data structure

The drug development process was described schematically in Fig. 1. There are many parameters that impact the outcome of such a process. For the specific decision problem at hand, i.e. the selection of projects to form an optimal portfolio, the following parameters are of particular importance:

- The cost of a project in each phase,
- The expected revenue of a project,
- The probability that a project proceeds to the next phase of development.

We will for the optimization problem presume that some assumptions have been made regarding these parameters for each of the available projects. We assume that a cost of running a phase is distributed over the duration of that phase, and let G_{ij}^* denote the cost associated with running project i in year j . Further, let R_i^* denote the revenue from project i if it is successfully launched to the market.

In the problem described in this paper, the values of G_{ij}^* and R_i^* are outcomes from a Monte Carlo simulation. The simulations are made to represent the drug development process as illustrated in Fig. 1, and they follow the framework for modelling of drug development processes outlined by [24]. The simulations account for the fact that many drug projects are terminated before completion, as illustrated in Fig. 1 by the possibility that the project is stopped after each phase. The terminated projects will consequently induce zero costs for the phases after termination, and they will also in these cases generate zero revenue.

Following an often-used terminology in the optimization literature, we will refer to the iterations of the Monte Carlo simulations as scenarios. Assume that simulations have generated K scenarios. The actually realized cost of project i in year j in scenario k is then

$$C_{ijk} = \begin{cases} C_{ijk}^* & \text{if } j \leq \tau, \\ 0 & \text{if } j > \tau, \end{cases}$$

where τ is the time of termination of the project. Similarly, the revenue is

$$R_{ik} = \begin{cases} R_{ik}^* & \text{if } \tau \geq L, \\ 0 & \text{if } \tau < L, \end{cases}$$

where L is the time at which the project might be launched to the market and start generating revenue. From the simulated scenarios we then estimate the expected revenue, $\bar{R}_i = (\sum_k R_{ik})/K$. The values of C_{ijk} and \bar{R}_i are key inputs to the optimization problem to be described in the subsequent sections. A numerical illustration of the optimization problem, using these variables as input, is given in a later section.

4 Mathematical modeling of the problem

In this section we will present the mathematical model of the project selection problem by adopting the general mathematical model of the optimization problem under uncertainty (2). The aim is here to maximize the expected revenue, \bar{R}_i , subject to the constraint that the cost under uncertainty, C_{ij} , may not exceed a defined annual budget, B_j . Let x_i denote the binary decision variables, defined as

$$x_i = \begin{cases} 1 & \text{if project } i \text{ is selected,} \\ 0 & \text{otherwise.} \end{cases}$$

We then have the problem as:

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^n x_i \bar{R}_i \\ &\text{subject to} && \sum_{i=1}^n x_i C_{ij}(\zeta) \leq B_j, \quad j = 1, \dots, T, \\ &&& x_i \in \{0, 1\} \quad i = 1, \dots, n. \end{aligned} \tag{5}$$

The objective function vector, coefficient matrix and right-hand side values are defined as \bar{R}_i , C_{ij} and B_j , respectively. Since the cost of each project in each year is a random variable and we are interested to ensure that the probability of being within the annual budget B_j is above a defined confidence level, we employ chance constrained programming. We replace the constraint in (5) with a chance constraint and present the mathematical model as follows:

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^n x_i \bar{R}_i \\ &\text{subject to} && P\left(\sum_{i=1}^n x_i C_{ij}(\zeta) \leq B_j, \forall j\right) \geq 1 - \alpha \quad j = 1, \dots, T, \\ &&& x_i \in \{0, 1\} \quad i = 1, \dots, n, \end{aligned} \tag{6}$$

where $\alpha \in (0, 1)$ and will be decided by decision makers. The chance constraint in (6) guarantees that the probability of exceeding the annual budget should be less than a predefined risk level α .

5 Approach to solution

5.1 The big-M constraint

We next adapt the general form of the CC-MBP (4) to our project selection problem, as defined in (6). In our problem, we sample the cost K times in the Monte Carlo simulation, and since K is typically a large number, this results in a large number of stochastic scenarios for cost. All scenarios from the simulation have the same probability, $p_k = 1/K$. Let z_k be a binary variable for each scenario controlling the enforcement of constraints of scenario k . If $z_k = 0$, the budgetary constraint is enforced, otherwise it can be violated. Model (6) can then be reformulated with a set of linear constraints as follows:

$$\begin{aligned}
 &\text{maximize} && \sum_{i=1}^n x_i \bar{R}_i \\
 &\text{subject to} && \sum_{i=1}^n x_i C_{ij}(\zeta) - Mz_k \leq B_j, \quad j = 1, \dots, T, k = 1, \dots, K, \\
 &&& \sum_{k=1}^K z_k p_k \leq \alpha, \\
 &&& x_i, z_k \in \{0, 1\} \quad i = 1, \dots, n,
 \end{aligned} \tag{7}$$

where $\alpha \in (0, 1)$, $\sum_{k=1}^K p_k = 1$ and M is a sufficiently large number which handles violation when $z_k = 1$. If the budgetary constraints are satisfied in a scenario, it is referred to as a responsive scenario, otherwise it is a non-responsive scenario and its constraints are ignored. Now our chance constrained optimization problem is converted into an integer programming problem that can be solved by integer programming solvers, as will be further described in later parts of this article.

5.2 Related approaches

We have in the sections above presented the CC-MBP for solving the optimization problem under uncertainty of coefficient at hand. While this is a computationally challenging approach, there are other simpler approaches that could also have been applied. Two such approaches are briefly described below, i.e. the Deterministic optimization and the Quantile chance constraint. In the Numerical Illustrations section, we will also give some numerical results comparing the performance of these two approaches with the proposed CC-MBP.

5.2.1 Deterministic optimization

In deterministic optimization, the stochastic nature of the cost would be ignored. Instead, the cost of a project would be represented with a fixed number. In our application, this would correspond to replacing the distribution of costs generated from the Monte Carlo simulation and instead use the expected cost of each project in each year. Let the mean, $\bar{C}_{ij} = \sum_k C_{ijk}/K$, be the estimate of the expected cost, we have the following model for the

deterministic optimization problem

$$\begin{aligned}
 & \text{maximize} && \sum_{i=1}^n x_i \bar{R}_i \\
 & \text{subject to} && \sum_{i=1}^n x_i \tilde{C}_{ij} \leq B_j, \quad j = 1, \dots, T, \\
 & && x_i \in \{0, 1\} \quad i = 1, \dots, n.
 \end{aligned} \tag{8}$$

5.2.2 Quantile chance constraint

With the Quantile chance constraint, Q-CC, the stochastic nature of the cost is handled by replacing the actual distributions with the $(1 - \alpha)$ percentile of the underlying distribution. Let \tilde{C}_{ij} be the $(1 - \alpha)$ percentile, for each project in each year, from the empirical distribution of the K scenarios from the Monte Carlo simulation. We then have the following deterministic formulation of the optimization problem.

$$\begin{aligned}
 & \text{maximize} && \sum_{i=1}^n x_i \bar{R}_i \\
 & \text{subject to} && \sum_{i=1}^n x_i \tilde{C}_{ij} \leq B_j, \quad j = 1, \dots, T, \\
 & && x_i \in \{0, 1\} \quad i = 1, \dots, n.
 \end{aligned} \tag{9}$$

It can be noted that the simplification of the Deterministic optimization and Q-CC reduces the number of constraints of the problem. While there is $(K \cdot T + n + 1)$ constraints in the CC-MBP, the other approaches only involve $(T + n)$ constraints. Despite the large number of constraints in CC-MBP, it is possible to solve the portfolio problem in a very reasonable amount of time.

5.3 Perturbation analysis of the annual budget

In the previous section, we developed the CC-MBP to deal with the presence of cost uncertainty and to enable us to solve the optimization problem at hand. In this section, we employ a parametric linear programming approach for investigating how the optimal portfolio selection changes as the annual budget limitation varies.

The parametric approach was developed in parallel with sensitivity analysis [25]. Both approaches use the information of post-optimality analysis and investigate how the optimal solution changes with uncertain parameters, but they are different. By employing the sensitivity analysis approach we can just investigate the current situation and what amount of changes in the current annual budget are allowed, such that the current optimal value of expected revenue as well as optimal project selection remain optimal. On the contrary, parametric programming can handle global uncertainty and by applying a parametric approach, the optimization problem is solved as a function of one parameter. Information can be provided about the dependence of the optimal solution on the uncertain parameter for the entire range from the minimum to the maximum value of expected revenue without the need to resolve the model.

Let \bar{B} be the same size of B and consider the parametrized linear program $z^*(\lambda) = \text{maximize}\{E^T x : Cx = B + \lambda \bar{B}, x \in \{0, 1\}\}$. We try to answer the question of over what range

of λ an optimal project selection and expected revenue remain optimal for the current budget. By replacing the budget constraint in (7) as follows, we describe the parametric program for our model:

$$\begin{aligned}
 & \text{maximize} && \sum_{i=1}^n x_i \bar{R}_i \\
 & \text{subject to} && \sum_{i=1}^n x_i C_{ij}(\zeta) - Mz_k \leq B_j + \lambda \bar{B}_i, \quad j = 1, \dots, T, k = 1, \dots, K, \\
 & && \sum_{k=1}^K z_k p_k \leq \alpha, \\
 & && x_i, z_k \in \{0, 1\}, \quad u_1 \leq \lambda \leq u_1 \quad i = 1, \dots, n,
 \end{aligned} \tag{10}$$

where $\alpha \in (0, 1)$, $\sum_{k=1}^K p_k = 1$, λ is a scalar and \bar{B}_j is the perturbation of the annual budget. First the mathematical problem (10) is solved based on the current budgetary constraint. Then parametric linear programming is employed to investigate the effect of budget variation, for more details, see [26]. This provides information about the critical region as a subset of all the parameters λ for which the optimal project selection set remains unchanged and the break point of λ which is the point when the optimally selected projects and expected revenue change.

6 Software

We have in the previous sections described the mathematical model of the optimization problem at hand (6), as well as the approach that we propose for solving the model (7), i.e. the CC-MBP. The final step in the analysis is to employ appropriate software to perform the actual numerical optimization procedure. Since our variables are restricted to be binary, we use the integer optimizer in the packages PULP and GLPK, and the code was written in Python 3 for solving models (7),(8) and (9). The calculations were performed on an Intel(R) i5 CPU, 2.5 GHz, with 8 GB of RAM.

7 Numerical illustration

In this section, we analyze the performance of the proposed approaches by applying each to a sample portfolio in the pharmaceutical industry. The sample portfolio contains a total of 15 independent candidate drugs in different stages of their development process. The number of simulation K is equal to 1000. At the time of the modelled decision-making, five projects are in the pre-clinical phase (PC), two projects are in phase 1 (PH1), six projects are in phase 2 (PH2), and two projects are in phase 3 (PH3). Each phase of drug development may take several years and the planning horizon, T , is set to 23 years. The timelines for the projects in the portfolio are illustrated in Fig. 2. There is a limitation for the annual budget and therefore all projects in the portfolio cannot be developed. Decision makers must decide which projects should be selected to maximize the expected revenue, while also considering the uncertainty in the related cost.

In Table 1, we summarize the assumptions regarding average costs to begin subsequent phases of each drug development project and expected revenue of each project. These costs are then distributed over the duration of the actual phase, to provide a cost per year

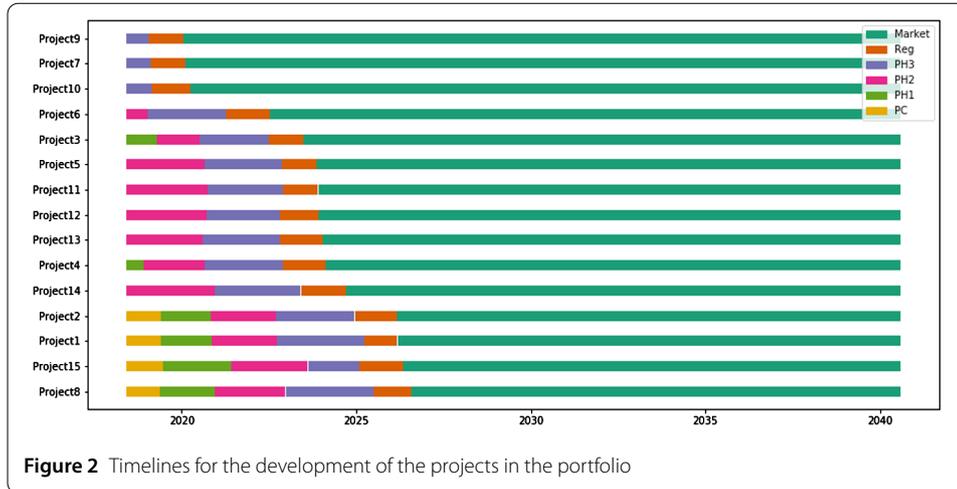
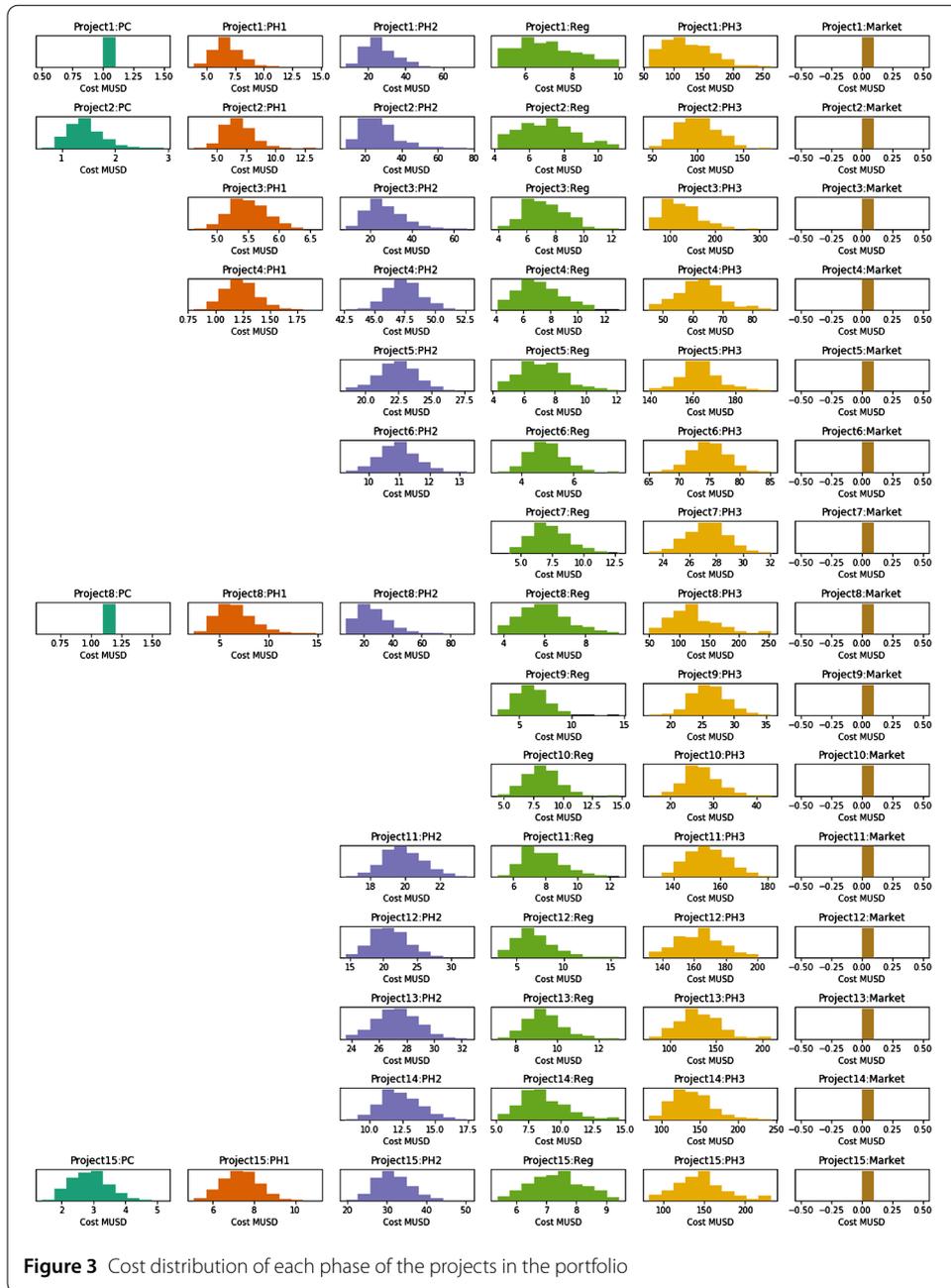


Table 1 Expected cost and expected revenue of each project (\$ million)

Project	Phase				Reg	Expected revenue
	PC	PH1	PH2	PH3		
Project 1	1	7	27	126	7	191
Project 2	1	7	27	101	7	235
Project 3	–	5	27	127	7	14,205
Project 4	–	1	47	63	7	8100
Project 5	–	–	22	164	7	955
Project 6	–	–	11	75	5	744
Project 7	–	–	–	27	7	6043
Project 8	1	7	27	123	6	179
Project 9	–	–	–	26	6	5677
Project 10	–	–	–	27	8	13,530
Project 11	–	–	20	155	8	668
Project 12	–	–	21	164	7	1036
Project 13	–	–	27	132	9	1347
Project 14	–	–	12	140	9	1285
Project 15	3	7	32	146	7	259

as described in the Input Data Structure Section. While the expected costs are summarized in Table 1, the underlying distributions are illustrated in Fig. 3. The costs presented in Table 1 and Fig. 3 are the costs conditional on the actual phase of development being conducted, not adjusted for the early termination of projects, hence being the basis for C^* as defined for the input data. It may be noted from Table 1 that both the cost and the expected revenue vary substantially between the project. Each of the three approaches, CC-MBP, Deterministic optimization and Q-CC, were employed for portfolio optimization. The results are shown in Table 2, where the expected revenue of selected portfolios, the risk of exceeding the specified annual budget and computation time are reported. The risk level was set to $\alpha = 0.1$ and the budgetary constraint was $B_j = 210$ MUSD. The results in Table 2 shows that the Deterministic optimization leads to a portfolio selection with a high expected revenue, but with a very large risk of exceeding the budget. Q-CC is conservative in that it has a much lower risk than the nominal 10%, and consequently selects a portfolio with lower expected revenue. The CC-MBP is successful in selected an optimal portfolio close to the specified nominal risk level. It may also be noted that despite the large number of constraints in CC-MBP, the applied approach is able to solve



the portfolio problem in a very reasonable time. In Table 3, we report the result of employing the CC-MBP with different risk levels and the annual budget set to 220 MUSD. Corresponding results are also visualized in Fig. 4. From these results it can be seen that the expected revenue changes significantly between $\alpha = 0.0$ and $\alpha = 0.01$, and after that the expected revenue is monotonically increasing. This type of data allows the decision maker to strike a balance between higher expected revenue and an increased risk of exceeding the allocated budget.

Table 4 shows the numerical result of post-optimality analysis for a ranges of annual budgetary restrictions with $\alpha = 0.2$ risk. The effect of budget perturbation on expected revenue and optimal project selection is studied globally. In this study, the budget of all

Table 2 Optimal project selection

Approach	Selected projects	E(Revenue) (\$ million)	Probability of over budget	CPU time (seconds)
Deterministic optimization	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15	54,456	29.5%	8e-3
CC-MBP	3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 15	53,182	9.4%	9.5
Q-CC	3, 4, 6, 7, 9, 10, 13, 15	49,906	1.3%	2.13e-2

Table 3 Optimal project selection

Risk level	Selected projects	Expected revenue (\$ million)
0%	2, 3, 7, 9, 10, 14, 15	41,235
1%	1, 2, 3, 4, 6, 7, 9, 10, 14, 15	50,270
5%	2, 3, 4, 5, 6, 7, 9, 10, 13, 14, 15	52,381
10%	2, 3, 4, 6, 7, 9, 10, 12, 13, 14, 15	53,417
15%	1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 15	53,787
20%	2, 3, 4, 5, 6, 7, 9, 10, 11, 13, 14, 15	54,085

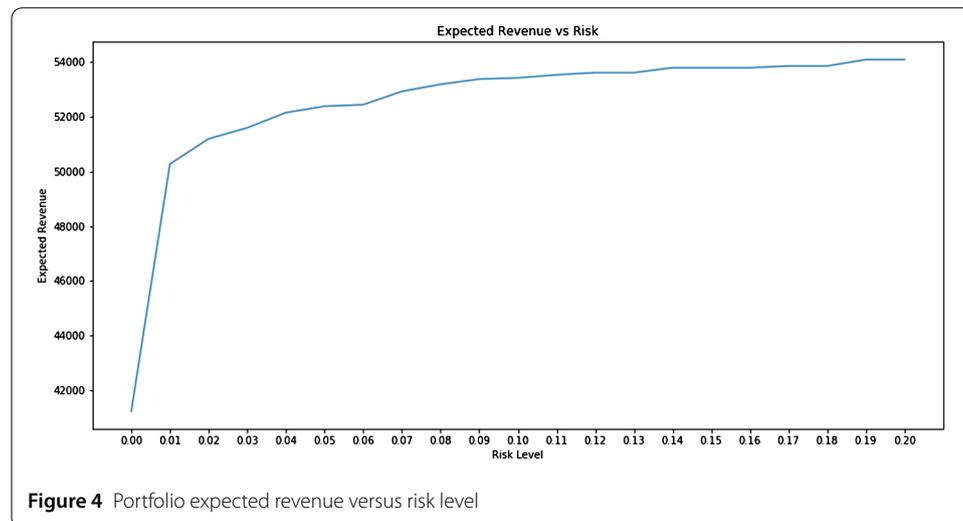


Figure 4 Portfolio expected revenue versus risk level

Table 4 Optimal project selection

Annual budget (\$ million)	Expected revenue (\$ million)	Selected projects
224	54,276	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15
222	54,265	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
217	54,085	2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15
210	53,850	3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15
200	53,787	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15
192	53,608	1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 15
190	53,552	1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15
182	53,417	1, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 15
177	53,182	3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 15
173	52,923	3, 4, 5, 6, 7, 9, 10, 12, 13, 14

years is perturbed equivalently and perturbation vector \vec{B} is considered as vector of ones, $\vec{1}$. The result of λ is converted into budget in which the decision maker is able to find break points on the annual budget. The optimal selected projects, and corresponding expected revenue of each break point of annual budgetary constraint, are listed in the Table 4. Critical regions can be found by the difference of two successive budget break points im-

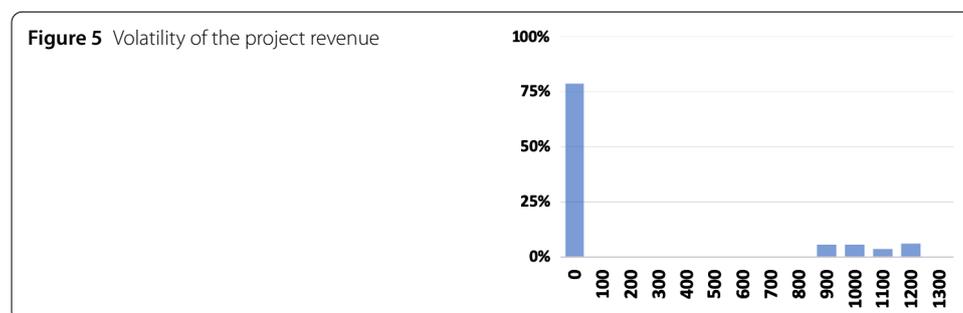
plying that increasing annual budget in these regions does not change the optimal selected projects and maximum expected revenue. As an example, Table 4 demonstrates that by allocating 200 million dollars each year, all projects except project 11 can be in the portfolio, yielding an expected revenue of 53,787 MUSD. An increase of budget to 210 MUSD is required to change the selected portfolio, hence increasing the expected revenue.

8 Discussion

The aim of this paper was to develop a model for stochastic portfolio optimization in the pharmaceutical industry. We have developed a chance constraint with Big-M coefficients and binary decision variables, CC-MBP, accounting for the uncertainty in the cost of drug development phases which leads to maximizing the expected revenue of the portfolio, while satisfying an annual budgetary limitation. A parametric programming technique was applied to systematically study the effect of annual budget variation on the optimal solution. The presented model can be a very valuable tool for decision makers, where adjustments can be made to cost distributions, expected revenues, risk level and budgetary limitations to intelligently select projects for a portfolio without dealing with mathematical formulae.

Even though the representation of pharmaceutical assets in these models is simplified, the results will provide valuable insights for portfolio decision makers not in the sense that the optimal solution is the one that a decision maker should necessary implement, but in the sense that the solutions provide more knowledge about the portfolio and its dynamics and constraints. The perturbation analysis provides insights about the relative value of assets in the portfolio in a very straight-forward way; it shows the thresholds at which the optimal solution changes. Furthermore, it prioritizes the assets based on risk, cost and return from a portfolio perspective, and will then be a well-rounded evaluation of the relative value of each asset. An addition to the perturbation analysis will be to add sensitivity on the risk appetite of the decision maker. If a greater appetite for exceeding budget is tolerated, will the composition of the optimal portfolio change? The answer will give yet another level of insight to the decision maker and will enable optimality across several portfolios that share budget constraints.

This study addressed the strategic objective of maximizing expected revenue subject to budgetary limitation. The cost for each project has uncertainty that we manage in the suggested models. Further additions to the model would be to introduce uncertainty in the target function (revenue), i.e., just as we are using uncertainty for project costs, we should do the same for project revenue. Rather than using a risk-adjusted expected revenue, it would be interesting to see if taking into account the volatility of the project revenue (see Fig. 5 for an example) may influence the selection of the optimal set of projects.



We have throughout the methodology section given a brief illustration of the general version of the methodology, alongside the adaptation of the methods to the current situation of our application. This serves to indicate that the developed approaches should be generally applicable to a wide range of stochastic optimization problems, not confined to the current pharmaceutical portfolio selection problem that was the focus of our investigation. As an example, the current data did not take into account the uncertainties and complexities involved in the pharmaceutical portfolio such as the dependency among drug development projects, the uncertainty in the lengths of phases, the budget limitations varying over time, as well as non-mathematical aspects such as the decision maker's preferences. However, the general approach should also allow the handling of these types of situations.

Acknowledgements

We thank Jonathan Trombley for his editing and useful suggestions.

Funding

This research has not received any funding.

Abbreviations

Q-CC, Quantile chance constraint; CC-MBP, chance constrained programming with big-M coefficients and binary variables.

Availability of data and materials

Simulated data are available upon reasonable request.

Competing interests

MF, MY and SJW are employees of Captario AB, marketing the software Captario SUM which was used to perform the modeling and simulations to create data for the empirical illustration. AC was an intern at Captario AB while doing this research. MY and SJW are shareholders of Captario AB.

Authors' contributions

The main idea of this paper was proposed by MY and MF. MF developed the model and wrote manuscripts. AC mainly developed the code for the model research. SJW helped with several pieces of useful theoretical advice. All authors read and approved the final manuscript.

Authors' information

MF has a PhD in Mathematical Optimization and a post doctorate in Computational Optimization and Applications. Her interest is in applications of optimization techniques to real industry problems. She held an academic position for several years and she currently works as a senior optimization researcher in the analytic group at Captario AB. AC is a PhD student in the department of Operations Research and Financial Engineering at Princeton University. He does research in optimization, learning and computational aspects of dynamics and control. MY has M.Sc. degree in Computational Mathematics and has a long experience working in the pharmaceutical industry. He has a background working with product strategy, portfolio management and decision support and currently heads up the analytics group at Captario AB. SJW has a PhD in Statistics and many years of experience from the pharmaceutical industry. He has a specific interest in the application of modeling and simulation to support strategic decisions in pharmaceutical drug development. He is currently the Chief Scientific Officer at Captario AB.

Author details

¹Captario AB, Gothenburg, Sweden. ²Department of Operations Research and Financial Engineering, Princeton University, Princeton, USA.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 29 November 2019 Accepted: 20 January 2021 Published online: 02 February 2021

References

1. Lean Y, Shouyang W, Fenghua W. Genetic algorithm-based multi-criteria project portfolio selection. *Ann Oper Res.* 2012;197(1):71–86.
2. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of r&d costs. *J Health Econ.* 2016;47:20–33.
3. Pharmaceutical research and manufacturers of America: annual report. Washington. 2016.
4. Munos B. Lessons from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov.* 2009;8(12):959–68.

5. Heidenberger K. Dynamic project selection and funding under risk: a decision tree based MILP approach. *Eur J Oper Res.* 1996;95(2):284–98.
6. Mokhtar M, Shuib A, Mohamad D. Mathematical programming models for portfolio optimization problem: a review. *Int J Math Comput Sci.* 2014;8(2):411–20.
7. Ghasemzadeh F, Archer N, Iyogun P. A zero-one model for project portfolio selection and scheduling. *J Oper Res Soc.* 1999;50(7):745–55.
8. Hassanzadeh F, Modarres M, Nemati HR, Amoako-Gyampah K. A robust r&d project portfolio optimization model for pharmaceutical contract research organizations. *Int J Prod Econ.* 2014;158(c):18–27.
9. Rogers MJ, Maranas CD. Valuation and design of pharmaceutical r&d licensing deals. *AIChE J.* 2005;51(1):198–209.
10. Sun B, Liu Y, Yang GQ. A robust pharmaceutical r&d project portfolio optimization problem under cost and resource uncertainty. *J Uncertain Syst.* 2017;11(3):205–20.
11. Charnes A, Cooper W. Chance-constrained programming. *Manag Sci.* 1959;6(1):73–9.
12. Miller B, Wagner H. Chance constrained programming with joint constraints. *Oper Res.* 1965;13(6):930–45.
13. Prekopa A. On probabilistic constrained programming. In: Kuhn H, editor. *Proceedings of the Princeton symposium on mathematical programming*, Princeton, 1967. Princeton: Princeton University Press; 1970. p. 113–38.
14. Song Y, Luedtke J, Küçükyavuz S. Chance-constrained binary packing problems. *INFORMS J Comput.* 2014;26(4):735–47.
15. Vielma JP, Ahmed S, Nemhauser GL. Mixed integer linear programming formulations for probabilistic constraints. *Oper Res Lett.* 2012;40:153–8.
16. Zeng B, An Y, Ludwig K. Chance constrained mixed integer program: bilinear and linear formulations benders decomposition. *Optimization Online. Technical report.* 1–30. 2014.
17. Dupačová J, Gaivoronski A, Kos Z, Szántai T. Stochastic programming in water management: a case study and a comparison of solution techniques. *Eur J Oper Res.* 1991;52(1):28–44.
18. Henrion R, Li P, Möller A, Steinbach MC, Wendt M, Wozny G. In: Grötschel M, Krumke SO, Rambau J, editors. *Stochastic optimization for operating chemical processes under uncertainty*. Berlin: Springer; 2001. p. 457–78.
19. van Ackooij W, Zorgati R, Henrion R, Möller A. Chance constrained programming and its applications to energy management. In: Drietas I, editor. *Stochastic optimization*. Rijeka: IntechOpen; 2011. Chap. 13.
20. Gupta A, Maranas C, McDonald C. Mid-term supply chain planning under demand uncertainty: customer demand satisfaction and inventory management. *Comput Chem Eng.* 2000;24:2613–21.
21. Luedtke J, Ahmed S, Nemhauser GL. An integer programming approach for linear programs with probabilistic constraints. *Math Program.* 2010;122(2):247–72.
22. Shapiro A. Monte Carlo sampling methods. In: *Stochastic programming. Handbooks in operations research and management science.* vol. 10. 2003. p. 353–425.
23. Luedtke J, Ahmed S. A sample approximation approach for optimization with probabilistic constraints. *SIAM J Optim.* 2008;19:674–99.
24. Wiklund SJ. A modelling framework for improved design and decision-making in drug development. *PLoS ONE.* 2019;14(8):e0220812. <https://doi.org/10.1371/journal.pone.0220812>.
25. Gal T, Greenberg HJ. *Advances in sensitivity analysis and parametric programming*. Boston: Kluwer Academic; 1997.
26. Marsten RE, Morin TL. Parametric integer programming: the right-hand side case. *Ann Discrete Math.* 1977;1:375–90.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)