A GENETIC ALGORITHM-BASED PHARMACEUTICAL PORTFOLIO SELECTION AND SCHEDULING FRAMEWORK

Gary E. Blau, Karthik Rajan, Joseph F. Pekny and Vishal A. Varma
Department of Chemical Engineering, Purdue University
And
Paul M. Bunch
Eli Lilly & Company

Abstract

One of the greatest challenges facing life sciences companies is the ability to discover and develop new products, which will sustain their long-term economic growth. The problem is difficult because of intense competition, the availability of a large number of new product ideas and limited human and capital resources. The planning problem is exacerbated by the presence of significant uncertainties in development times, development costs, resource requirements and anticipated product sales. In addition, the presence of dependencies between products, both in the market place as well as in their development adds further complexity. In this paper, a discrete event simulation for the drug development process is combined with a Genetic Algorithm (GA) to select the best sequence of projects in the presence of uncertainties and dependencies. A graphical tool ,the risk-reward bubble plot is used to arrive at heuristics that prune the GA search. The GA then captures the optimal structure of the highest expected reward/risk sequences, despite an enormously large combinatorial search space. A risk-reward frontier, which illustrates the trade-offs between risks and expected rewards is presented as an output of the optimization exercise. For a nine drug example case study, the optimal portfolio is expected to give a 50 % higher return than the portfolio suggested by the bubble diagram at the same level of risk. A sensitivity analysis demonstrates the robustness of the sequence to changes in the levels of uncertainties and dependency conditions.

Keywords

Net Present Value, Discrete event simulation, Genetic Algorithm, Risk-Reward Frontier, Bubble Plot, Optimal Sequence

Introduction

New Product Development (NPD) in phamaceutical and biotechnology industries is associated with systematic (or non-diversifiable) risks due to significant product attrition rates (product failures at clinical trial stages) and development time, development cost and marketplace reward uncertainties (Blau et al, 2000, 2001, Pisano, 1997). In the fortunate situation, where a large number of new drug candidates are available from internal discovery or in-licensing opportunities, the problem of selecting which drugs to develop and in what order, is a challenge because of finite human resources and capital constraints (Cooper, 1985, 1998). Effective portfolio management aims at:

- Deciding the Number of Products to develop
- Selecting the "best" projects for development
- Arriving at a robust schedule, which ensures a high level of expected returns at a reasonable risk.

Pharmaceutical New Product Development consists of the following main development activities:

- First Human Dose (Pre-Clinical) Activities: This activity is aimed at assessing the effects of the drug on healthy human volunteers. Various phamaco-kinetic studies such as absorption, distribution and metobolism in the body, are carried out.

- Clinical Trial Stages: There are 3 clinical trial stages, also called Phase 1, Phase 2 and Phase 3. The scale of these trials in terms of the number of tests increases from Phase 1 to Phase 3.
- Launch and Product Supply Chain Activities: It takes about 7-13 years for a compound to be successfully launched. Once, the FDA approval is granted, there are 3 stages of ramp-Up sales, which are followed by mature sales, until the end of the patent horizon.

These development stages are also associated with activities such as process design, construction of pilot as well as large-scale plants (in parallel with phase 3 activities).

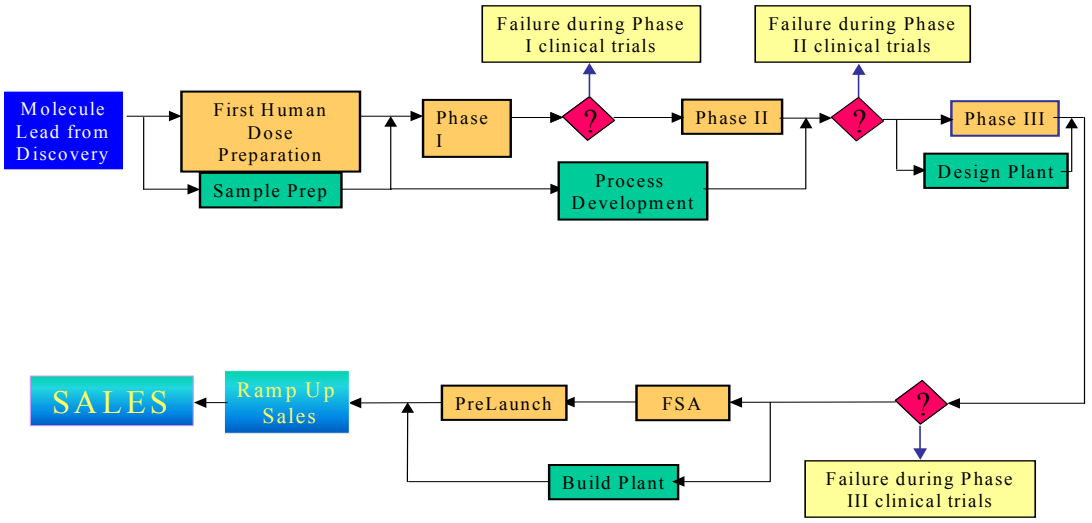

Figure 1. Flow Diagram of steps involved in development and commercialization of a new drug

The Discrete Event Simulator (DES)

A model of the above probabilistic network was simulated using a discrete event simulation engine, CSIM 18 (Mesquite Software). Every activity is associated with probability distributions for duration, costs and rewards. For the present case study, all the distributions were assumed to be triangular. A portfolio of 9 drugs was used to demonstrate the utlity of the GA-DES framework. Due to the complex uncertainties and comparatively large input variances, 10,000 Monte Carlo trials were used for each portfolio of drugs. Each trial is a simulated walk of the potential product along the probabilistic network. The sequence in which projects are developed remains the same for all Monte Carlo replicates of a given simulation. The output of the simulation is a Net Present Value (NPV) distribution. The Expected Return (i.e. rewards) for the sequence is approximated by the mean of positive values under the NPV probability distribution, while the risk is defined as the Probability of Losing Money.

Drug Dependency Considerations

A dependency between two or more drug molecules is manifested by way of effects on anticipated sales, development costs, durations and success probabilities of a drug due to events such as launch, failure associated with other drugs. For instance, when two drugs targeting the same disease are launched, the sales of each are lower than the sales, if only one of them was launched. This dependency is called sales dependency. Technical dependency involves changes in the success probabilities associated with sequential development of drugs from the same families of chemistries. Resource dependencies involve reduction in capital requirement when more than one drug can be made by the same manufacturing process, so that multi-product facilities can be used. All these dependencies, have been incorporated into the DES model, in order to reflect the variations that charateterize a real new product development process.

Bubble Plot (BP) Heuristics

An optimum portfolio must not only contain optimum number of drugs (too many drugs lead to significant queuing for limited resources, while too few projects lead to less than optimal capacity utilization), but also, a development sequence that ensures high rewards and low risk. A graphical tool, called as a bubble plot serves as an aid in determining the sequence. It plots the return/risk ratio against the probability of success. The size of the bubble indicates the capital requirements, while the color indicates the disease type. Intuitively, small bubbles with high reward/risk and high technical success probabilities

are preferred. From the bubble plot in Figure 2, it is easy to see that for the 9-drug portfolio, a preferred development sequence is (5,9,3,7,1,4,8,2,6). Several other sequences (scheduling drugs by trading-off risk and reward e.g. two 5-drug sequences could have been (8,5,4,9,1) or (5,8,9,4,1)) were inferred from the BP. Since, the simulation of each of these sequences established a high reward/risk ratio, the BP was used to generate portfolios of

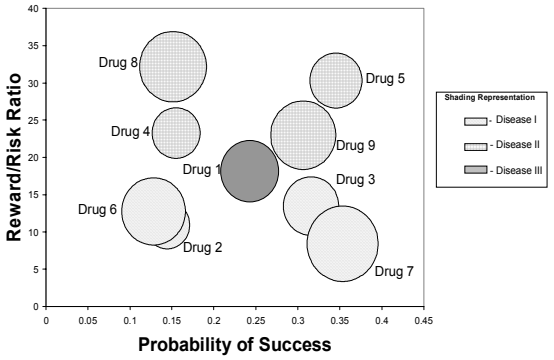

Figure 2. Bubble Plot for 9 New Drug Candidates

drugs for the starting population of the Genetic algorithm. It was observed that a GA initiated by a majority of high reward/risk portfolios leads to faster convergence. The algorithm is discussed in the next section.

The Genetic Algorithm (GA) Search

The large combinatorial decision space can be illustrated by the fact that the total number of possible portfolios equal 9!+8!+7!+….+1!. For instance, a 9-Drug sequence can have as many as 9! different development sequences. Evolutionary search strategies such as GA's (Holland, 1975, Goldberg, 1989) allow robustness with respect to the problem definition. A rank ordering of projects serves as the encoding for the GA. Two-Point crossover operators such as, Order Crossover (OX), Position-based Crossover (PBX) and mutation operators, such as Insertion Mutation (IM) and Swap Mutation (SM) were applied as part of the search procedure (Gen, 1997). A crossover probability of 0.8-0.86 and mutation probability of 0.08-0.1 were found to yield considerably fast convergence to the desired optimal solution. The Genetic Algorithm procedure first evaluates all members of the population. (for the present case study, population size =10). Each member sequence is evaluated by calling the discrete event simulation (10,000 replications/simulation). A statistical procedure called Remainder Stochastic Sampling (RSS) is used to select the members with the highest normalized fitness. The fitness of a member is correlated with a weigted risk-return criteria given by:

$$Z_k = \alpha\left(\frac{NPV_k - NPV_{\min}}{NPV_{\max} - NPV_{\min} + \gamma}\right) + (1-\alpha)\left(\frac{risk_{\max} - risk_k}{risk_{\max} - risk_{\min} + \gamma}\right)$$

The weight α=0.8 was used for the present case study. RSS allows the member sequences with the highest normalized fitness to be represented most strongly in the intermediate population. The intermediate population is subjected to the two-point crossover and mutation operators. A new population of sequences is created and the procedure is repeated until the search space has been explored to the desired level. For the current portfolio study, two sets of GA's were run in order to partition the search space: a GA for 8,9-drug sequences and a GA for 5,6,7-drug sequences. The GA's were run on 1000 MHz, 256 MB RAM, Intel Pentium Processors. Each GA required about 50 generations and consumed a CPU time of ~28 hours.

Computational Results and Analysis

Figure 3 shows the risk-reward plane consisting of 1000 data points, obtained from the two GA runs of 50 generations each. The figure also shows the risk-reward frontier. A point on this frontier represents the maximum expected returns for a given level of risk. It is remarkable to note that the frontier is U-shaped as against the purely increasing frontier of Markowitz's (Markowitz, 1991) mean-variance model frontier. However, the frontier seems to be less than intuitive. Further, the same trends are observed for the individual GA runs (50 generations) also. The Markowitz trend is observed until the point that represents an optimal sequence. For this part of the curve, as the risk increases, the capacity utilization increases, yielding higher returns. A plateau region indicates that sequences (5,9,8,3,1), (5,8,9,4,2,3) and (8,5,9,6,3,2) yield almost identical returns at widely differing risk levels. The reason for this is as follows: A 5-drug sequence leads to a substantially high capacity utilization. Hence, the lead time for a sixth drug is so large that in a majority of replications, the sixth drug is not launched within the 20 year planning horizon. Beyond, the plateau region, the returns reduce drastically as the risk increases. In this region drugs with low reward/risk ratio are prioritized. It is important to note that for all three optimal sequences, determined by the GA, the highest reward/risk drugs 5,8 and 9 were developed first. Inclusion of a low reward/risk drug such as drug 2 increases the risk. The optimal 5-drug sequence (5,9,8,3,1) yields an expected return that is ~50 % higher than the sequence determined from the bubble plot. This exercise proves that in the presence of resource constraints, a high reward/risk prioritized portfolio yields significantly higher returns, even if the work-in-progress (or portfolio size) is small. The effect of dependencies is manisfested by the fact that the optimal sequence contains drugs 5,8 and 9, which target the same disease, thereby reducing capital costs and increasing success probabilities.

Conclusion

Pharmaceutical New Product Portfolio Management is a complex inter-play of achieving an optimal level of capacity utilization, risk diversification and economic returns. This is accomplished by controlling the portfolio size and schedule, by way of assigning an optimal development sequence. The complexities of decision-making in the presence of development uncertainties and dependencies have been addressed in this paper and a Genetic Algorithm based framwork that incorporates these features is presented. From the 9-drug case study runs, it has been proved that the GA is able to capture the optimal structure of the sub-sequences that lead to solutions, whose performance measures far exceed those of the best heuristically generated sequences. In addition, a graphical tool, the bubble plot, has been shown to prune the GA search, thereby reducing time to convergence. Future work in this area includes development of simulation and optimization models to incorporate a continuous product-insertion portfolio.

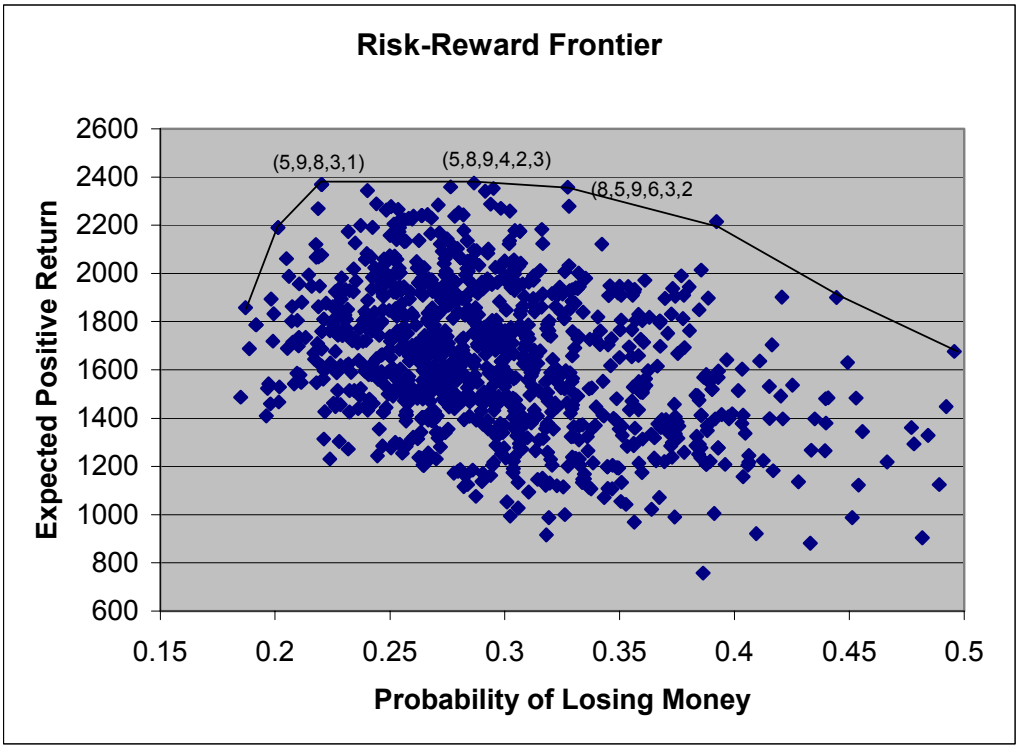

Figure 3. Risk-Return Frontier

References

Blau, G.E., Mehta, B., Bose, S., Pekny, J.F., Sinclair G., Kuenker, K. and Bunch, P. (2000). Risk Management in the Development of New Products in Highly Regulated Industries. *Computers and Chemical Engineering*, 24, 659-664.

Blau, G.E. and Sinclair, G. (2001). Dealing with Uncertainty in New Product Development. *Chemical Engineering Progress*, 21, 2.

Cooper, R.G., Edgett, S.J., Kleinschmidt, E.J. (1998). Best Practices for Managing R&D Portfolios. Research Technology Management, July-August, 20-33.

Cooper, R.G. (1985). Selecting Winning New Product Projects: Using the NewProd System. Journal of Product Innovation Management, 2, 34-44.

Holland, J.H. (1975). Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor.

Goldberg, D.E. (1989). Genetic Algorithms in search, optimization and machine learning, Wiley, New York, NY.

Gen, M. and Cheng, R. (1997). Genetic Algorithms and Engineering Design, Wiley, New York.

CSIM 18 Simulation Engine (C++ Version, 2000), User's Guide, Mesquite Software, Inc., http://www.mesquite.com.

Markowitz, H.M., (1991). Portfolio Selection, 2nd Edition, Blackwell Publishers.

Pisano, G.P. (1997). The Development Factory: Unlocking the Potential of Process Innovation, Harvard Business School Press.