

USING SIMULATION TO UNDERSTAND THE EFFECT OF UNCERTAINTIES IN THE DRUG DEVELOPMENT PROCESS

Sujoy Dey, Stevens Institute of Technology
Jack M. Kloeber, Jr., Ph.D., Johnson & Johnson
John V. Farr, Ph.D., Stevens Institute of Technology

Abstract

The purpose of this paper is to discuss the uncertainties introduced during the process of drug development, define quantified discrete event simulation and their cumulative effect upon R&D productivity after the process of drug development was modeled. The four key input variables modeled were process time for preclinical and clinical trials, attrition rate as the drug attempts to pass scientific and medical hurdles, human resources required represented by full time equivalents (FTEs) and the direct cost for each phase of development. Cumulative and dependent variability is not easily captured in traditional spreadsheet models, especially when modeling processes whose states change at random points in time. We use discrete event simulation in order to handle random times, resources, and dependencies. This paper should increase the visibility of the complexity and variability of the ethical drug development process, and the model described should enhance long-term system control, productivity, stability, and decision support.

Introduction

In 2003, pharmaceutical research companies spent \$33 billion on research to develop new and better medicines, a 7% increase from 2002 (Pharmaceutical Research and Manufacturers of America (PhRMA)). Many major pharmaceutical companies believe they must continue to increase their R&D expenditure every year in order to boost their product pipelines and keep pace with their rivals in a race for market share. It has been suggested that, to simply maintain a modest industry growth rate of 8% (Centre for Medicines Research International), companies would need to triple the number of New Molecular Entities (NMEs) launched annually. However, an emerging picture is that purely increasing R&D investment does not guarantee an increase in NME output. Given the fact that the average cost of launching a new drug is around \$1.7 billion (Kermani, and Grittins, 2004) and rising, the industry faces an immense challenge.

The challenges that face the pharmaceutical industry to gain utmost profitability in developing a new drug are many and complex. It has become very difficult for managers to actually see into the development pipeline and account for the various

uncertainties that exist within each phase of development. Total development times, from discovery to market launch (or from concept to market), are around 10-15 years (The Wellcome Trust, 2004). The development time to first market for each NME is a crucial time period for the marketing company since its duration determines the period of marketing exclusivity (patent cover) available to the company to attempt to recover its R&D expenditure. Additionally, a blockbuster drug/product can lose up to US\$2.7 million in sales every day that launch is delayed. Therefore, reducing overheads and speeding up time to market is a major concern for pharmaceutical companies.

Regulatory tracking and reporting across R&D is highly scrutinized within the pharmaceutical industry. These departments are now being required to continually provide management with metrics that compare resources, time and costs involved in the various tasks performed in the regulatory submission lifecycle. The never changing efforts of reducing expenses by the company add to these pressures. Recent research shows that performance in critical areas is inconsistent and that managers in pharmaceutical companies cannot predict with certainty that their next product will be available according to plan (The Wall Street Journal, 2000).

Problem Statement

This paper will investigate the drug development cycle. Areas of investigation include time line and drug development resource utilization issues. Variables of interest in the process of modeling include process time for drug development, probability of technical and regulatory success of projects per phase of development, associated costs for various phases of drug development, and finally the full time equivalents that are needed for the mission of drug development to be a success.

A Simulation Approach

Industries like the pharmaceutical industry typically use static tools such as spreadsheets, or project management software to develop schedules and allocate resources. These tools provide limited abilities to project time and resources within the scope of project risk (attrition rate) and cost. These are variables

within each phase of development that feed off the variable patterns of arrival rates, full time equivalents (FTEs) or resources, and regulatory risk within each of those phases. Drug development consists of various business and clinical processes that are cross functional and characterized by complex structures and variability, thus increasing uncertainties and dependencies across processes. Those familiar with using spreadsheets would realize that trying to achieve all this using spreadsheets becomes bulky and too cumbersome. We are looking at innumerable spreadsheets with similar innumerable columns to denote elements for multiple processes occurring within each phase of development. The complexity increases if modeling multiple therapeutic areas, with each therapeutic area having multiple disease areas. A simulation method representing such variability would be a powerful approach for analysis and quantitative evaluation of the processes (e.g. such variability can cause unbalanced capacity utilization over time). Powerful hardware and software tools have made simulation effective and a popular technique for improving process performance, design, and efficiency. Recent simulation software has included optimization with the simulation. Optimization would automate the search for the best values given the input factors. This enables the decision maker to identify the critical factors to enable the selection of the best approach for a particular process (Laguna and Marklund, 2005).

The Modeling Approach

The drug development process is a process where the state variables change at discrete points in time. Discrete event simulation describes the modeling of a system or process in which the flow of units is discrete over time. A discrete event simulation model examines the dynamics of the system when a new event occurs after the completion of the previous event as state variables change only when an event takes place. The key variables identified in this process were process time, attrition rate, cost (of individual phases), and FTEs (resources). These variables change as a result of an event (a discrete event) occurring in the process. Models using this type of simulation technique focus on the time instances when these discrete events occur. This feature allows for significant time compression because it makes it possible to skip through all the time segments between events when the state of the system remains unchanged. Therefore, in a short period of time, the computer can simulate a large number of

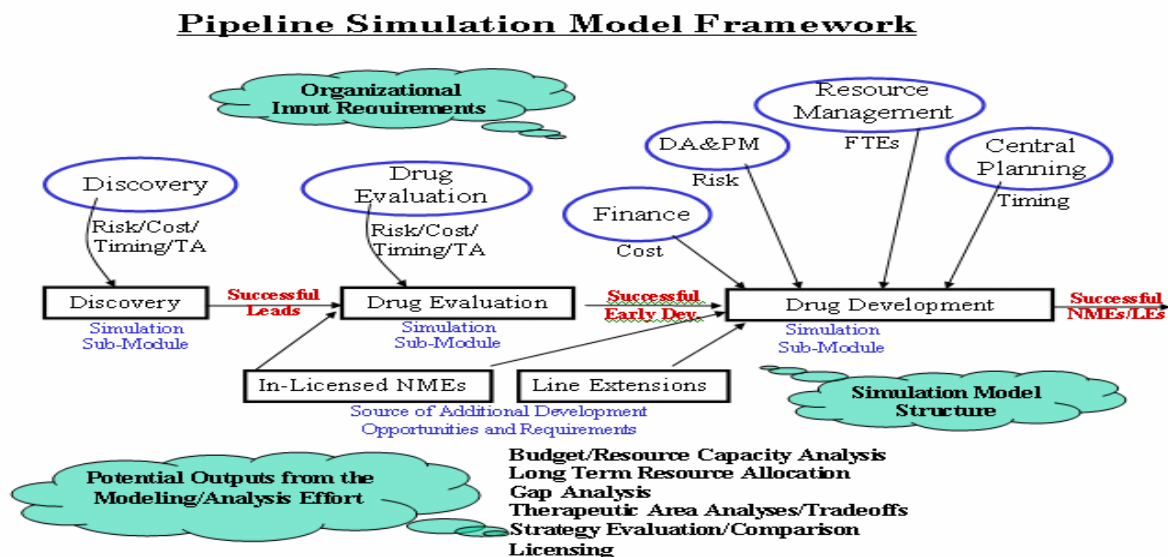
events (characteristic of processes occurring within each phase of drug development) corresponding to a long real-time span (Laguna and Marklund, 2005).

Framework of the study

This study used commercially available general simulation software package called Extend from the company Imagine That, Inc. to model the drug discovery and development pipeline. We selected Extend due to its ability to build hierarchical models to address system- subsystem interactions, the built in database utility, and the assurances that a maintenance and support capability would be available in the future (e.g. for a even higher resolution, lower level modeling within each phase of a therapeutic area) and throughout the planned life cycle of the system. The scope of this effort was to model development efforts from late discovery (lead identification) and early development efforts, through full development and launch. The pipeline model developed was flexible enough to handle multiple therapeutic areas, multiple indications within any particular therapeutic area, and expandable to allow it to evolve into a company-wide model for organizing multidisciplinary efforts.

Our emphasis was on the process of designing and building a pharmaceutical pipeline simulation system. A high level modeling of the pipeline discovery phase followed by a lower level modeling of elements that were identified to be modeled in the drug development phases was performed. These elements were selected through a series of meetings with the stakeholders (Drug Discovery and Drug Development executives as well as resource managers and finance executives) who needed the ability to change these elements in order for them to have the added flexibility in controlling the process. Through the iterative steps of analysis, evaluation, feedback, modification and control, this system was improved in its effectiveness, output quality, ownership cost, and user satisfaction. This study explored the above stated issues in depth, collaborating with individuals who play major roles in drug development in the pharmaceutical industry. Focused interviews with a wide variety of industry participants allowed us to understand the process flow and generate the primary source of data. These data were supplemented with a wide variety of secondary data sources, such as resource management databases from the pharmaceutical company and pharmaceutical market research and consulting firms.

Exhibit 1. Pipeline Simulation Model Framework.



In Exhibit 1, the following abbreviations were used: New Molecular Entity (NME), Line Extension (LE), Therapeutic Area (TA), Full Time Equivalent (FTE), Decision Analysis (DA), and, Portfolio Management (PM)

Modeling the Pipeline

Discovery (the more scientific portion of R&D) was modeled using a discrete event generator modeling the arrival of an optimized lead (a good candidate molecule for eventually becoming a drug). This phase was modeled at a high level since the focus was mainly on the early and full development phases. Multiple therapeutic areas were introduced within this stage which would continue to the other phases. Random time intervals were used for lead arrivals. A Generator block which starts a discrete event simulation was used and the appropriate distribution (Exponential in this case) was applied to indicate inter-arrival times. This model used a mean inter-arrival time of 2 weeks between each lead going into the PreClinical phase. (Note that this model is using dummy data and different companies would have different data for their mean times). Number of resource types (e.g. chemists, biologists, clinical operations, etc.) was accounted for and a resource “pool” for each resource type was created with the number or available resources indicated for each resource pool. The flexibility to change, add, or delete the number of available resources or a resource pool is easily handled.

Each phase received all successful projects (or leads) from the previous phase for processing. These

leads were sent in through a first-in-first-out (FIFO) queue and a random node accounted for risk (probability of technical success) within that phase. The lead was also routed through a queue handling resources which pulled available resources from the resource pool. The added value here is that we also modeled the randomness for the processing times that were used for utilization of the resources for completion of a task (e.g. a clinical test within that particular phase). Resources were released once the task was completed and sent back to the resource pool making it available for the next event.

An important feature was the calculation of waiting times and waiting costs for each resource type while waiting for the queue to free up. This is important in the light of the fact that each day of delay of a potential blockbuster drug could cost the company over \$2 million in sales.

Other important features that added value to the model and to the analysis were the ability to model initial conditions, NMEs, LEs (Line Extensions), and in-licensing. The large amount of data needed for the initial conditions made the internal database control critical. Initial conditions (within a phase) are the existing values (or conditions) that were obtained earlier due to earlier analyses and would be the starting conditions for any of the phases. For example, data for a successful lead from Phase IIa (Proof of Principle) which is ready to enter Phase IIb trials (establishing the best dose) could be modeled as starting from Phase IIb. In other words, arrival times, resource requirements, probabilities of technical and regulatory success, and

costs of successful Phase IIa leads are known and used as starting data for Phase IIb. Since the organization knows a great deal about the characteristics of this drug, it can better estimate how the drug may perform in the coming trials. This information is more reliable than benchmark information (time, cost, resources, and attrition rate) that is used for the generic successful Phase IIa leads that had originated from Discovery.

Modeling of potential LEs that could be developed once a successful NME has launched was also included in this study. Successful NMEs or LEs were routed back for clinical trials in Phase IIb with new “conditions” for time (depending on patent expiry), attrition rate, cost, and resource needs. The same conditions were accounted for when modeling in-licensing a compound from another pharmaceutical or bio-tech company for a particular phase of development.

While modeling initial conditions and NMEs/LEs, the need for a database was solved. The advantage was that a database was embedded within the model. The database management system uses table creation wizards and built-in table editors, so that databases can easily be built and managed entirely from within the model in Extend. This enhanced flexibility allowed for creation, viewing, and manipulation of all model data through the data management system within the model.

Inputs to the model were:

1. A Poisson arrival process to model the arrival time of an optimized lead (a potential drug molecule candidate) from the discovery phase. Historical data was used to select the appropriate Poisson parameter (λ)
2. Probability of technical and regulatory success for the lead within each phase of development.
3. The number of full time equivalents (FTEs) or resources by resource type that are available for each phase for the duration of the phase (as long as it takes for the required tests to be completed for a particular phase).
4. A Uniform distribution for completion time for a test within a phase.

Outputs from the model were:

1. The total number of successes, failures, and in-process leads within each phase of development.
2. Time taken for the arrival of a successful project in a phase.
3. Number of FTEs in each skill type utilized in each phase (utilization rate) due to both successful and failed projects.
4. The total cost spent on development for each phase, by year.
5. Total productivity of the system.

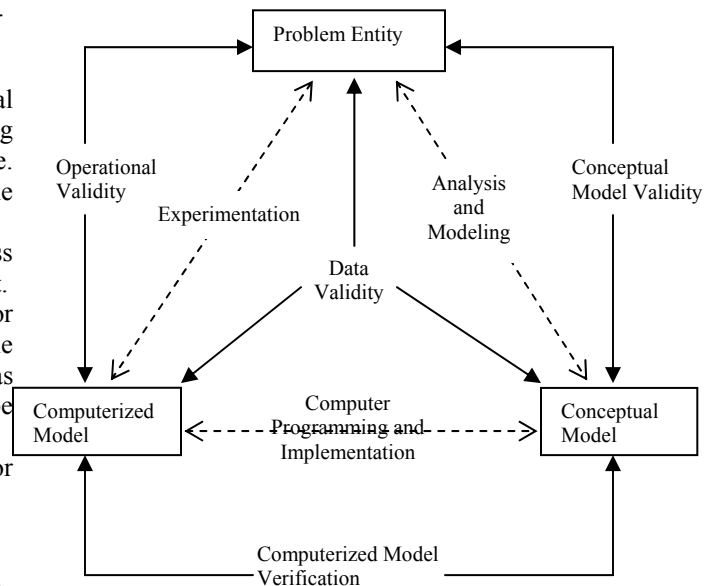
6. Maximum waiting time.
7. Maximum number in queue.
8. Average number in queue.
9. Capacity of the total system.

Verification and Validation

Model verification and validation as related to the development of the model was explained in Exhibit 2, adapted from Sargent (2000).

As shown in Exhibit 2, the problem entity was the system (real or proposed), idea, situation, policy, or phenomena to be modeled; the conceptual model was the mathematical/logical/verbal representation of the problem entity developed for a particular study; and the computerized model was the conceptual model implemented on the computer. The conceptual model was developed through an analysis and modeling phase, the computerized model was developed through a computer programming and implementation phase, and inferences about the problem entity were obtained by conducting computer experiments on the computerized model in the experimentation phase (Sargent 2000).

Exhibit 2. Simplified Version of the Modeling Process (Sargent 2000)



For *verification* of the system, a flow diagram was developed which explained each system action (logic) whenever an event took place. This was repeated for each phase of drug development. This computerized conceptual model was checked by a person (other than the modeler) intimately acquainted with the drug development process regularly as the model was being built. Input settings for various parameters were examined by the model output for reasonableness. At

the end of the run, the input settings were examined to make sure that they did not change for any reason. Animation was used to verify visually whether the system imitated the actual system. Animation was also used for *validation* purposes to display the model's operational behavior and the process flow, and displayed graphically as the successful lead from Discovery moves through the drug development process through time. Other tests, both subjective and objective were used using standard established procedures. To ensure high face validity, discussions were held with people very knowledgeable about the system (system experts). Demonstrations of the model inputs and outputs were performed for industry experts within resource management, finance, portfolio management, decision analysis, and process analysis. Data was compared to validated existing systems that had comparable data outputs. In short, structured walk-throughs were performed before an audience of all key people to ensure high validity and credibility.

A Capacity Analysis for the Early Development Stages of the Drug Pipeline.

This model was used to conduct a capacity analysis to report the insights gained from using a discrete event simulation model. This example analysis will cover the early development stages of drug development (Preclinical through Phase IIa). Assumptions for budget, resources (headcount), attrition, and time were stated. Note that we did not use initial conditions in this test. Therefore the beginning of the simulation required transition time until the system reached steady state.

Four types of resources were used and levels of each set.

1. Chemists
2. Biologists
3. Clinical Operations
4. Clinicians

Two therapeutic areas, the Central Nervous System (CNS) and Oncology (anti-cancer) are shown in Exhibit 3.

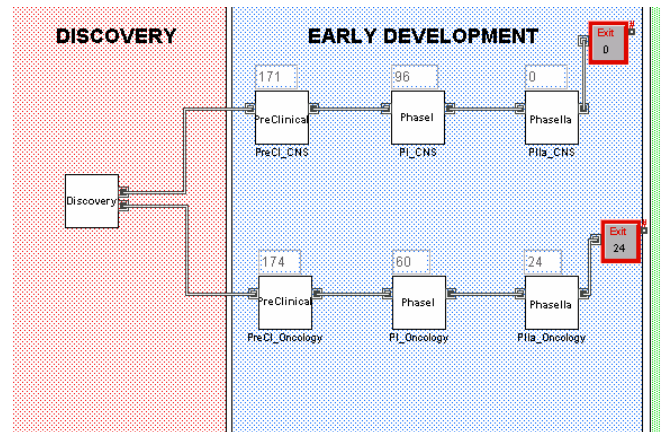
Inputs to the two therapeutic areas were:

1. Demand (resource capacity)
2. Processing (activity) time
3. Probability of technical success (risk)
4. Cost (waiting time and activity based)

Outputs from the two therapeutic areas were:

1. Number of success and failures per phase
2. Resource utilization
3. Average and maximum length of queues
4. Average and maximum time in queues
5. Cost statistics

Exhibit 3. Simulation Pipeline View.



Note: Confidence interval used for outputs was 95%

Exhibit 4. Per-Project Per-Phase Requirements for Therapeutic Area CNS.

CNS	Chemist	Biologist	Clinical Operations	Clinician
Preclinical	2	4	2	0
Phase I	1	3	3	9
Phase IIa	0	1	5	10

Exhibit 5. Per-Project Per-Phase Requirements for Therapeutic Area Oncology.

Oncology	Chemist	Biologist	Clinical Operations	Clinician
Preclinical	1	3	2	0
Phase I	1	3	3	7
Phase IIa	0	1	4	11

The above were numbers specific resource skill type required by the two therapeutic areas. We will check the number of resources *actually* available for use.

Exhibit 6. Total Number of Resource Types Actually Available.

Resource Type	Total No. Available (assumption)
Chemists	10
Biologists	20
Clinical operations	30
Clinicians	30

From Exhibits 4, 5, and 6, it was seen that 37 clinicians were required from an available pool of 30 clinicians. It is very probable that there would be a bottleneck as far as the need for clinicians is concerned. The probability of bottlenecks for the other resources, i.e. chemists, biologists, and clinical operations was lower due the availability of a larger pool for these resources. Hence, we expect to see longer processing times for clinicians and also a longer queue length for the resource clinicians. Similarly, processing times and queue lengths for chemists, biologists, and clinical operations were expected to be lower than those of clinicians. Other data used for CNS as inputs were:

1. Probability of Technical Success (PTS) for Preclinical = 0.69
2. PTS for Phase I = 0.56
3. PTS for Phase IIa = 0.32
4. Activity Delay = Refer to Exhibit 7

Exhibit 7. Process Time of Trials for CNS.

Phases	Time (Activity Delay)	Cost per week [in million]
Preclinical	Test 1 = 9-15 weeks Test 2 = 9-15 weeks	Test 1 = \$0.02 Test 2 = \$0.02
Phase I	Test 1 = 9-18 weeks Test 2 = 10-17 weeks	Test 1 = \$0.25 Test 2 = \$0.25
Phase IIa	45-60 weeks	\$0.69

Other data used for Oncology as inputs were:

1. Probability of Technical Success (PTS) for Preclinical = 0.69
2. PTS for Phase I = 0.80
3. PTS for Phase IIa = 0.60
4. Activity Delay = Refer to Exhibit 8

Exhibit 8. Process Time of Trials for Oncology.

Phases	Time (Activity delay)	Cost per week [in million]
Preclinical	Test 1 = 9-15 weeks Test 2 = 9-15 weeks	Test 1 = \$0.02 Test 2 = \$0.02
Phase I	20-32 weeks	\$0.25
Phase IIa	45-60 weeks	\$0.69

Trial duration's randomness for both therapeutic areas was accounted for using a Uniform distribution. The simulation was run for 1040 weeks or 20 years. The results (at the end of Phase IIa) obtained were:

1. 4 CNS Phase IIa successes were obtained.
2. 8 Oncology Phase IIa successes were obtained.
3. Resource utilization for Chemists = 0.00
4. Resource utilization for Biologists = 0.28
5. Resource utilization for Clinical operations = 0.71
6. Resource utilization for Clinicians = 0.91
7. Average length of queue = High (due to unavailability of adequate number of Clinicians) See Exhibit 10.
8. Waiting Times in queue = High (longer processing times due to unavailability of adequate number of clinicians.) See Exhibit 10.

Exhibit 9 shows the utilization rates of the different resource types as the simulation runs from 0 to 1040 weeks. As mentioned earlier, we did not use initial conditions in this test. Thus, the beginning of the simulation represented transition time until the system reached steady state.

Exhibit 10 shows the various queues (blocks) used in each of the phases. Under the column "Block", the letters help identify the block. These are manually assigned by the modeler. The column "Block Name" tells the type of block that was used. FIFO denotes "first-in-first-out".

Exhibit 11 shows a snapshot of cost statistics as obtained from the model. This tracks the cost per item, cost per time unit, and total cost.

Exhibit 9. Graph of utilization rates for resource types

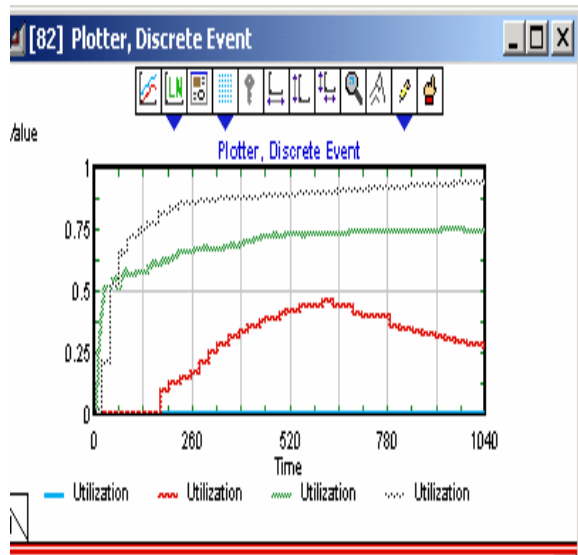
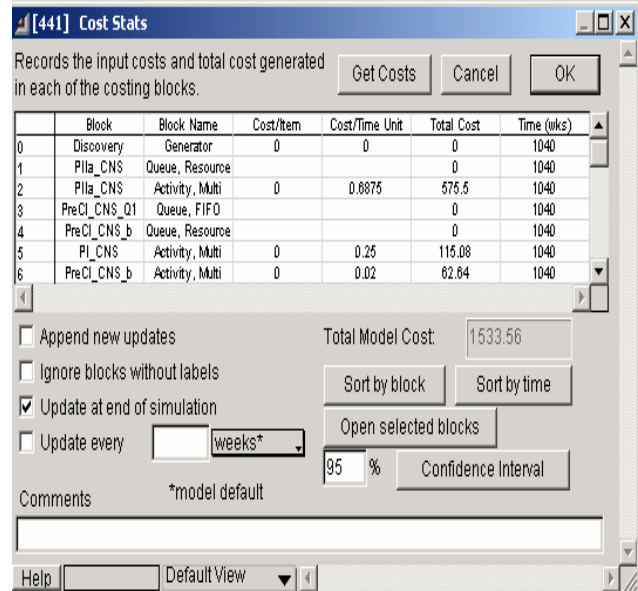


Exhibit 10. Queue Statistics

Block	Block Name	Ave Length	Max Length	Ave Wait	Max Wait
PIIa_CNS	Queue, Resource	8.20	22	250.8	336.40
PreCl_CNS_Q1	Queue, FIFO	0	1	0	0
PreCl_CNS_b	Queue, Resource	0.05	3	0.20	5.85
PreCl_CNS_Q2	Queue, FIFO	0	1	0	0
PI_CNS_Q1	Queue, FIFO	0	1	0	0
PreCl_CNS_a	Queue, Resource	0.03	3	0.12	5.27
PI_CNS	Queue, Resource	68.79	136	396.5	847.06
PI_Onc_Q1	Queue, FIFO	0.22	11	0.73	22.35
PI_Onc	Queue, Resource	61.35	100	499.3	611.41
PI_Onc_Q2	Queue, FIFO	0	1	0	0
PIIa_Onc	Queue, Resource	23.53	62	106.4	135.84

Exhibit 11. Cost Statistics



Hence, we observe that resource type clinicians certainly caused a bottleneck with more than a 90% utilization rate. Thus, it becomes easier for the manager to work the resource type clinicians to remove the bottleneck. Exhibit 9 shows in detail the length of the queue and the waiting times for each of the queues in each phase of development. This helps in easy identification of bottlenecks in the process. Increasing the number of clinicians will definitely improve the utilization rate of this resource and clear the bottlenecks that were identified. On the other hand there would be cost implications since increasing clinicians would increase cost. Note also that utilization rate for Clinical operations was at 71%. This resource type should be monitored because of the fact that this resource type was most likely to cause a future bottleneck.

Conclusions

Summing up the salient features of using this method that would translate into benefits for the reader and his/her pharmaceutical company are:

1. Better utilization of resources and resource capacity
2. Identification of bottlenecks
3. Analytical study of short/long-term effects of various “what-if” scenarios
4. Allowance for variable duration of projects and sub-projects
5. Simulation for long-term planning, analyses, and strategy evaluation
6. Handling of multiple therapeutic areas

7. Valuable inputs for cost-benefit analyses for support/justification of expenditures to management
8. Low cost and responsive demonstration of ideas dynamically
9. Ultimately, better resource management, improved productivity, and lower cost.

The outcome of this study was a very good understanding of the impact of selected key variables on productivity and cost of a very complex drug development process. This study would enhance the capability of managers to understand and account for the critical variations within the process for all the given inputs and convert them to realistic and more importantly, meaningful outputs for analysis. Out of pocket (OOP) costs, resource requirements, timing, and project failures for long term (5-10 ten years) analysis, planning, and strategy evaluation would be possible. Ultimately, this simulation model would enable organizations to increase research productivity by streamlining data management, and increasing information accessibility across the organization. The major beneficiaries of this research would be resource, product and portfolio managers who would have a more robust planning capability to deal with the complexity of the drug development process. Each component, when analyzed separately, could be used to improve decision making which would ultimately give a company a competitive advantage.

References

- Kermani, Faiz, and Gittins, Rebecca, “*Clinical Remains Key*”, May 25, (2004), www.inpharm.com
- Laguna, Manuel, and Marklund, Johan, “*Business Process Modeling, Simulation, and Design*”, Pearson Prentice Hall, (2005).
- Pharmaceutical Research and Manufacturers of America (PhRMA), “*Pharmaceutical Industry Profile*”, (Washington D.C.: PhRMA, 2004).
- Sargent, G. Robert, “*Simulation Model Verification and Validation*”, Proceedings of the 2000 Winter Simulation Conference.
- The Wall Street Journal, “*Drug price program notes*”, Dow Jones & Company, Inc., August 10, (2000).
- The Wellcome Trust, An “*Overview of Drug Development*”, May 25, (2004), www.wellcome.ac.uk

About the Authors

Sujoy Dey is a Research Assistant and PhD student in the Systems Engineering and Engineering Management Department at Stevens Institute of Technology. He teaches courses in systems engineering and engineering management. His research interests include systems modeling and simulation, and decision analysis under

risk and uncertainty. He is a student member of the International Council of Systems Engineering (INCOSE), and the Project Management Institute (PMI). Sujoy holds MS degrees in Technology Management, Organic Chemistry, and Biochemistry, and is currently working on completing his PhD this year in Systems Engineering. He can be reached at sdey1@stevens.edu and phone: (201) 216-8025.

Jack M. Kloeber Jr. is a retired US Army Lieutenant Colonel with interests in military decision-making, modeling and simulation, technology selection, R&D portfolio analysis and decision analysis. He received his PhD in Economic Decision Analysis from the Engineering School at the Georgia Institute of Technology. Having taught graduate-level courses in decision analysis, modeling, and simulation at the Air Force Institute of Technology for six years, Jack retired from the US Army in 2000, and is now Head, Portfolio Management Group, Johnson & Johnson Pharmaceuticals Group, in Raritan, NJ. His Email: jkloeber@prdus.jnj.com, and phone: (908) 927-5075.

John V. Farr is a Professor and Department Director, Department of Systems Engineering and Engineering Management, Stevens Institute of Technology. Before coming to Stevens in 2000, he was a Professor of Engineering Management (EM) at the United States Military Academy at West Point. He served as program director of their EM program from 1997 to 2000. He is a former past president and Fellow of ASEM. He can be contacted at jfarr@stevens.edu, phone: (201) 216-8103, or fax: (201) 216-5541.