INT. J. PROD. RES., 2001, VOL. 39, NO. 2, 395–412

Queueing network modelling and lead time compression of pharmaceutical drug development

N. VISWANADHAM†* and Y. NARAHARI‡

Designing an optimized pharmaceutical drug development process is an important problem in itself and is of significant practical and research interest. Drug development lead time is a critical performance metric for a pharmaceutical company. In this paper, we develop a multiclass queueing network model to capture the project dynamics in drug development organizations that involve multiple, concurrent projects with contention for human/technical resources. We explore how drug development lead times can be reduced using efficient scheduling and critical mass-based resource management. The model captures important facets of any typical drug development organization, such as concurrent execution of multiple projects, contention for resources, feedback and reworking of project tasks, variability of new project initiations and task execution times, and certain scheduling issues. First, we show, using a class of fluctuation smoothing scheduling policies, that development lead times can be compressed impressively, without having to commit additional resources. Next, we show that critical mass-based project teams can compress lead times further. The model presented, though stylized, is sufficiently generic and conceptual, and will be of much value in new drug development project planning and management.

1. Introduction

The pharmaceutical industry has evoked considerable interest among professionals and researchers alike. Strong disciplinary focus, tight government control, and higher performance in comparison with most other industries in terms of key financial ratios and higher product development times, unparalleled by any other industry, are a few reasons for the interest in this industry (Basa and Allen 1994).

The pharmaceutical industry is highly capital and R&D intensive. The commercial success of any pharmaceutical company is primarily focused on its research equity. Intellectual property rights and patents also play a crucial role, more so than in any other industry, in determining the financial success of the company. The pharmaceutical industry has been one of the most successful industries in the recent past. Its high profit rate has been attributed to the high-risk nature of its drug development, drug approval,and marketing process. The success of the industry, according to Sarantopoulos *et al.* (1995), is due to:

(1) the increasingly ageing population, which creates a demand for more drug products;
(2) international sales, especially in the developing countries;

Revision received July 2000.

† Mechanical and Production Engineering, National University of Singapore, Singapore – 119 260.

‡ Computer Science and Automation, Indian Institute of Science, Bangalore – 560 012, India.

* To whom correspondence should be addressed. e-mail: mpenv@nus.edu.sg

International Journal of Production Research ISSN 0020–7543 print/ISSN 1366–588X online
http://www.tandf.co.uk/journals
DOI:10.1080/00207540010004304

(3) The possible innovation and diversification of products to treat a greater number of diseases; and
(4) enhancement of the industry's productivity.

As the pharmaceutical industry is highly regulated, patents play an important role in determining the success of the company. The success of any pharmaceutical company depends on its ability to successfully develop and market new drugs, faster than the competition. During the period of validity of the patent the company enjoys a monopoly in that particular drug. Therefore, successful drug development is crucial for any pharmaceutical company. In this paper we shall focus on this aspect of the pharmaceutical company,namely the drug development process (DDP) (Yevich 1991).

1.1. *Motivation and objectives*

Designing an optimized DDP is an important problem in itself and is of significant practical and research interest. Effective drug development involves minimizing the time and resources required to deliver an outstanding new drug to the customer. One of the most important performance metrics of a DDP is the end-to-end drug development lead time (DDLT) that a drug takes from the discovery stage to the final marketing stage. Introducing new drugs faster than the competition allows companies several opportunities, such as setting new product standards, being a technical pioneer, being able to respond rapidly to customer feedback, and ultimately realizing higher profit margins (Zirger and Hartley 1996). Lead time compression in a DDP assumes special relevance in the light of patents and the monopoly associated with them. The faster a pharmaceutical company is able to market the drug after its initial discovery, the greater is the time of the monopoly enjoyed and thus the higher are the profits. Furthermore, with the advent of the Internet and direct selling of drugs to customers, there is increased need for accelerating the DDP. Thus product development managers are all the time looking for techniques to shrink the lead times of new drug development projects. This article explores effective means of compressing the DDPT, using the modelling framework of queueing networks.

Recognizing the importance of lead times in a drug development setting, it is our objective in this paper to develop lead time models of drug development organizations (DDOs) using queueing networks, which have been successfully used in factory floor operations modelling. The project dynamics in a DDO or in general in any product development organization is slow compared with that of a production process on a factory floor. In this paper, we explore and demonstrate the validity of queueing network models in capturing the dynamics of DDOs to obtain important insights into lead time reduction.

In particular, our aim is to model the lead time performance of a typical DDO having the features of:

- multiple, concurrent new drug development projects in progress,
- contention for human/technical resources in the organization,
- feedback and reworking of project tasks, and
- randomness in task execution times and arrivals of new projects.

The model created forms the basis for exploring opportunities for compressing the DDLT.

1.2. *Review of relevant work*

Several recent articles and books have addressed the problem of acceleration of the product development process, for example see the articles by Adler *et al.* (1995, 1996), Zirger and Hartley (1996),Hauptman and Hirji (1996), Cohen *et al.* (1997) and Narahari *et al.* (1999) and the books by Smith and Reinertsen (1991) and Wheelwright and Clark (1992). The articles by Adler *et al.* (1994, 1995) formulate a single class queueing network model for a real-world product development organization (PDO) and conduct a rich variety of experiments with the model to bring out several strategies for speeding up the product development process. In these papers, the authors argue that process models are better than project models for reasoning about and improving the performance of (PDOs) with multiple, concurrent, non-unique projects. The authors present a process model from which they create a single class queueing network model which is parameterized using a detailed set of measurements in a real-world organization. The model so created is simulated in a variety of resource allocation and decision-making scenarios and validated against the performance measured. Many what-if type of experiments are conducted with the model to focus on various strategies for accelerating new product development projects. Narahari *et al.* (1999) have looked into the lead time modelling of generic product development and product design organizations using single class and multiclass queueing network models. They have shown using these models that lead times can be reduced in an innovative way by using efficient scheduling, input control, load balancing, and variability reduction. These ideas have been applied in the present article to the specific case of pharmaceutical DDO.

Harrison and Loch (1995) advocate the use of simple stylized queueing network models to study the quantitative impact of input conditions on the performance of any business process, so as to develop broadly applicable intuition about the process performance. They emphasize the effect of variability on system performance. In the manufacturing arena, lead time reduction is an important subject. For example, Hopp *et al.* (1990) emphasize the role of variability reduction as a means of reducing cycle times, using a queueing theoretic framework. The book by Hopp and Spearman (1996) contains several ideas on lead time reduction, again from a queueing theoretic perspective. Suri (1996) has explored the use of queueing models in the design and analysis of quick response manufacturing systems. There are also interesting case studies on lead time reduction, see for example, Bourland and Suri (1992) and Bourland (1994). Many ideas embodied in these works can be used in the context of compressing the end-to-end delay of a DDP.

There are as yet no studies on analytical modelling of the kind attempted here of the DDP. There are however many empirical results available in many recent web sites (Web 1–9).

1.3. *Contributions and outline*

We explore queueing networks (Viswanadham and Narahari 1992, Hopp and Spearman 1996) as the lead time models, motivated by their success in modelling discrete event systems in general and factory floors in particular (Viswanadham and Narahari 1992). However, since there are both similarities and differences between a factory floor and a DDO, the fact that queueing models are successfully used in

factory floor modelling, by no means, implies trivial extension to modelling the dynamics of multi-project DDOs. For example, the project dynamics in a DDO or in a general product development organization is slow compared with that of a production process on a factory floor. Furthermore, resources in a product development context are of a totally different nature, namely R&D teams, scientists, technical resources, engineering workstations, etc. A primary objective of this paper is to show that in spite of these differences,queueing network models can faithfully capture the dynamics of project execution in a DDO at a certain level of abstraction, namely the level of abstraction of a product development manager or drug design scientist. We would like to validate the use of such models in accurately capturing the dynamics of DDOs, leading to important insights into lead time reduction.

In this paper, we show that the dynamics of project execution in a typical DDO can be accurately described by a probabilistic re-entrant line (which is a multiclass queueing network) (Kumar 1993, Narahari and Khan 1997) and such a model can be used to obtain valuable insights into reducing development lead times. The use of probabilistic re-entrant lines makes the models more realistic than the coarse single class queueing network models and enables subtle, internal scheduling issues to be revealed, motivating the use of fluctuation smoothing scheduling policies (Lu *et al.* 1994) and other queueing techniques in achieving lead time reduction.

In § 2 of this paper, we present the various activities and decisions in a typical new DDP. In § 3, we describe a re-entrant line model of any typical DDO. The model is quite conceptual and serves to describe aggregately any multi-project DDO with contention for resources. We also briefly review fluctuation smoothing policies in the context of re-entrant lines and bring out their relevance for multi-project DDOs. In § 4, we bring out how performance analysis of the queueing network model can be used to estimate development lead times under a variety of resource allocation scenarios and to identify and exploit opportunities for reducing the lead times. First, we explore fluctuation smoothing policies and show that by intelligently selecting the next task class to be processed, one can reduce cycle times in a quite remarkable way. Next, we show the positive effect of critical mass-based project teams in compressing development lead times even further.

Since company-specific data is hardly available in the open literature, the present study relies on generic, empirical data available on some web sites, for formulating and validating the model. Company-specific data will enrich the models further. We believe, however, that all critical determinants of lead time performance have been captured by the model presented in this study.

2. New drug development process

Developing a pharmaceutical drug is a complex process and calls for extensive collaboration and coordination between different functional units of the company (Knoop and Worden 1988). The skills required for the different stages of the DDP are very different from one another and require expertise from fields such as chemistry, biology, law, patents, bio-statistics, etc. Further, the strict stipulations laid down by the US Food and Drug Administration (FDA) makes the DDP lengthy and expensive.

In the past three decades, the time required to successfully develop and market a new drug has increased from 8.2 years in the 1960s to 14.8 years in the 1990s (Yevich 1991). Further, the pre-tax capitalized cost of bringing a new drug into the market

has increased from US $ 54 million in1976 to US $ 359 million in 1990 (Yevich 1991). This however, is the cost estimate of the drugs that entered clinical trials in the late 1970s. A precise projection of the cost to develop a new drug from scratch is difficult because of the long time horizons of the drug development process. This means that measurement based validation is not possible. The long development time exposes the DDP to varied sources of risk and uncertainties such as exchange rate and interest rate fluctuations, politics and regulatory policies etc. to name a few.

The drug research process can be divided roughly into the following stages:

(1) biology and chemistry
(2) pharmacology and toxicology
(3) pharmacy and analysis, and
(4) clinical trials.

Besides these, the evaluations conducted by the authorities (Investigational New Drug (IND) and New Drug Application (NDA)) are also an important part of the research process.

Figure 1 shows the various steps in a typical DDP along with approximate time frames (Yevich 1991).

2.1. *Biology and chemistry*

During the first stage of the research process the biologists try to target a particular disease or condition and then try to understand the mechanisms that cause the body to show those symptoms. It then becomes necessary to develop and synthesize a substance which influences a specific mechanism in a desired manner. This stage calls for a close interaction and coordination between chemists and biologists to target and develop a suitable chemical structure.

2.2. *Pharmacology and toxicology*

During this phase the new substance developed is tested. The tests can be performed on isolated tissue preparations, cell structures or with animals and are called pre-clinical trials. These investigate the effects of the new substance. The critical factors that must be determined for any experimental drug are its absorption, distribution, metabolism and elimination (ADME) characteristics. The side effects and/or the genetic effects that the new substance might have are studied at this stage. The effect of this experimental drug on the illness or condition that it is supposed to cure is also studied at this stage. A large number of experiments are carried out to obtain enough knowledge and understanding of the substance so that it can be safely given to humans.

2.3. *Pharmacy and analysis*

The active substance made by the biologist and chemists has to be made suitable for use in patients. This requires developments of an effective drug delivery system for the experimental drug, that is a suitable formulation such as a tablet, capsule or solution. This delivery system should provide the patient with the right amount of the drug at the right time at the right rate. Also, the chemical process needed to manufacture the drug on a large scale needs to be developed. Besides the review of new drugs, the FDA has established the *Good Manufacturing Practices* regulations as a set of rules that govern the manufacturing of the pharmaceutical products (Web 8). Good Manufacturing Practices do not mention how to perform the specific opera-

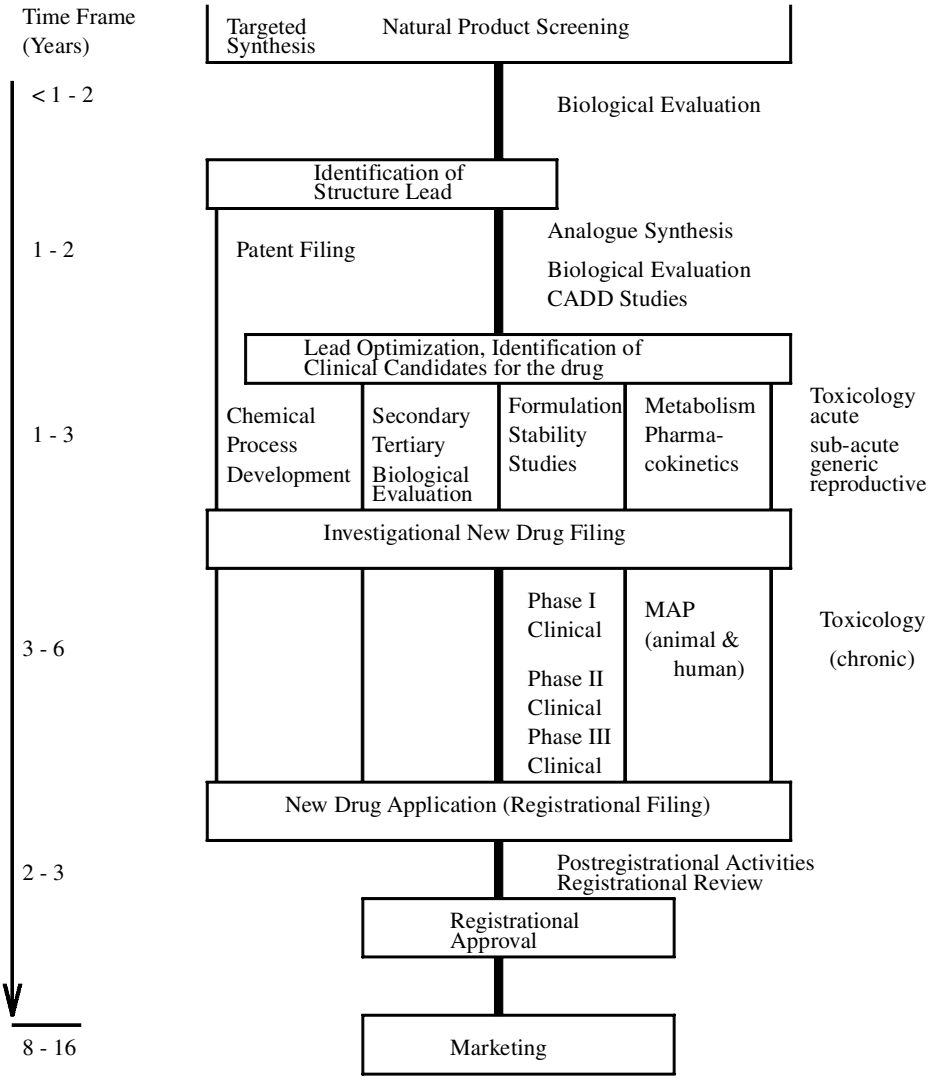

Figure 1. A typical new DDP.

tions but lay down general policies that the manufacturers have to follow for FDA approval of a new drug. A typical example of these policies is scrapping upon failure as well as stopping all the immediate machines while clearing the buffers. Also highly sensitive quality control methods need to be developed to measure and control the quality of the drug before it ever reaches the patients.

2.4. *Investigational new drug*

After extensive toxicology tests and animal studies, if the manufacturer decides to test the drugs on humans then it has to file an IND application with the FDA. Usually the application runs to as many as 2000 pages. It must also include information on the pre-clinical tests (pharmacology and toxicology) and a proposed plan to carry out the clinical studies on human beings. Unless the FDA orders a hold,

clinical trials can begin 30 days after the application is filed. In addition, the application must be approved by a review board of institution(s) where the trials can be conducted.

2.5. *Clinical trials: Phases I–III*

In Phase I of the clinical trials, safety studies are carried out on a small number (usually 20–100) of healthy volunteers. The potential side effects of the drug are identified and a dosage range is determined. The volunteers are given complete information about the trials. Phase II of the trials involves the administering of the drug for the first time to study its effects on the disease it is supposed to cure. The effectiveness of the drug is determined. Approximately100–300 volunteers who have the targeted disease participate in the trials. It is also important to find out the optimum dose as quickly as possible. The Phase III trials are conducted on a very large scale, often at multiple sites, involving around 1000–3000 patients in clinics and hospitals. The purpose of this phase to study the drug's efficacy and safety in long term use. Also comparative studies are done at this stage to compare the drug with the best therapy that is currently available.

2.6. *New drug application*

After all the tests are over and the drug's efficacy and safety have been established, the company compiles all the data relating to the drug and submits a NDA to the FDA. The NDA typically may run to 50 000 pages. The NDA must contain all the information the company has gathered. It contains comprehensive documentation relating to the new drug and is sent to different countries for approval. According to the law, the FDA has 180 days to review an NDA. But the average review time for a NDA in the year 1995 was 19.2 months. Although this is a considerable improvement over previous years (it was 34.1 months in 1984), total drug delivery time, from discovery to final FDA approval, has been increasing considerably.

Apart from the cost and time, the complexity of the DDP has increased significantly over the years. The number of patients per clinical trial, the number of clinical trials and the number of pages in the NDA have all doubled. The growing complexity of the DDP means increased uncertainty, risk, and cost of bringing new drugs to the market. For patients, longer drug development times mean doing without new drugs. Diseases such as AIDS give drug development a special urgency. Longer drug development times are unacceptable. Thus the controlling authorities are faced with the unparalleled problem of ensuring the safety, efficacy and, at the same time, fast availability of the drugs.

3. A queueing network model for the drug development process

3.1. *Some modelling issues*

At a first glance, different DDPs may seem to have unique features. The peculiarities may arise from a variety of reasons ranging from the difference in the organizational structures, the nature of the targeted disease, to the difference in the management philosophy. On the other hand, there are reasons that account for the similarity in DDPs, the most important being the tight government (FDA) control over the pharmaceutical industry and its activities—from drug manufacturing to drug development. These stringent requirements make the DDP uniform across different pharmaceutical companies.

Basa and Allen (1994) bring out the important issues in the project management of the DDP and various important decisions at different stages and summarize the effects of different organizational structures on the DDP. But this work makes no attempt to identify the high leverage points in the DDP which would be critical input to planning and managing the DDP. Our approach here is to focus on the similarities in different DDPs, to create a process model that is a fairly general representation and to identify the leverage points, ignoring the superfluous minutiae (typically associated with an idiosyncratic DDP), which may otherwise hinder a systematic study for these DDPs under a unified framework. Once such a general model is analysed over a wide range of input and system parameters to identify the high leverage points, it can be fine-tuned to suit a particular pharmaceutical company.

A model that faithfully captures the various important facets of the DDPs should include:

(1) multiple, concurrent projects,
(2) contention for resources,
(3) variability of project initiations and task execution times,
(4) reworks and feedbacks, and
(5) termination of projects before completion.

The resources in the model are the teams in the organization which are dedicated to a single stage of the DDP, for example, compound synthesis, pharmacology etc. The choice of the customer would reflect the chosen level of granularity in the model. There could be two options: we could either study each and every compound on which any study is initiated and follow its path in the network as a tagged job or we could just represent the whole project as symbolic of all the compounds that are evaluated as part of the project. In the first case, the customer comes out of the network if a compound is rejected at any of the stages, but the project still goes on. A project gets terminated when all the compounds that enter the network as part of the same project are rejected. In the second case, the customer comes out of the network if a particular DDP is terminated at any of the stages.

To arrive at a good decision, it would be relevant to discuss the merits and drawbacks of both the approaches. In the first case the network population downstream would be orders of magnitudes different from the population upstream as, for a particular targeted drug, typically 5000–10 000 new compounds enter the DDP contending for evaluation as the active ingredient in the targeted drug. But in the later downstream stages only one of these 5000–10 000 compounds emerges as the final candidate. Also one has to keep track of all the compounds that enter the DDP as part of a particular project to evaluate whether or not the project is to make an exit from the network. Further, it would be very cumbersome to represent cases where a particular project is terminated on the basis of reasons other than simple compound evaluations, for example, some strategic decisions. In case that happens, one has to remove all the compounds from all the various stages of the network which are part of the same project. In the second case, the amount of book-keeping involved is much less compared with the first case but there is a risk of losing some important insights. To keep matters simple and tractable, we chose to adopt the second strategy, for example, to represent the customers as the development projects themselves rather than the compounds.

In the DDO under study, at any given time, many new drug development projects are in progress at different phases. This causes contention for engineering/human

resources and results in delays at various points. We should like to model the resulting congestion. Often, different phases of the same development project could be contending for a given resource.

There are various decisions that are taken at different stages of the drug development process that affect the flow of activities. (Basa and Allen 1994) provide a set of important decisions that are taken during the drug development. Here we consider those decisions which could possibly affect the flow of activities in the process:

- toxic findings may lead to synthesis of improved analogues,
- change pre-clinical testing owing to discovery of possible toxicity or carcinogenicity in humans,
- add or change toxicology and pharmacology or other pre-clinical tests at FDA request,
- discontinue owing to toxicity, carcinogenicity, etc in non-human models,
- Go/No-Go after Phase I, II, III,
- discontinue for a therapeutic indication during the clinical trials,
- change clinical testing plans owing to lack of efficacy,
- change clinical testing plans owing to toxicity or side effects,
- alter clinical testing plans at FDA request,
- alter clinical testing plans because of scarcity of subjects,
- add more non-human model studies owing to perceived lack of adequate pre-clinical profiling.

Figure 2 represents a typical DDP with the various decisions involved at different stages.

This is a set of decisions that are relevant to pre-clinical and clinical testing. There are other strategic/managerial decisions that might decide the Go/No-Go for a drug development project. All these decisions are important and a good model should capture them. Re-entrant lines are one such model.

3.2. *Re-entrant lines*

Re-entrant lines (Kumar 1993) are appropriate for modelling queueing systems with distinct multiple job visits to service centres. In a re-entrant line, the parts visit the same server several times at different stages of processing, before exiting the system, thus making the flow *re-entrant*. A re-entrant line can be described as follows. There is set of *service centres* $\{1, 2, \ldots, m\}$. Service centre $i \in \{1, 2, \ldots, m\}$ has n_i logical or physical buffers, $b_{i1}, b_{i2}, \ldots, b_{in_i}$. For $j \in \{1, 2, \ldots, n_i\}$, the buffer b_{ij} contains parts visiting service centre i for the jth stage of processing. A part visits these buffers in a given sequence and any service centre is typically visited several times in the route of a part.

Figure 3 shows a typical re-entrant line with three service centres and 11 buffers. Parts enter the system at buffer b_{11} and visit the centres according to a deterministic route as shown. Finished parts emerge from centre 3 after undergoing processing following a wait in b_{33}. Note that every part in this example line visits centre 1 three times, centre 2 five times, and centre 3 three times.

In the re-entrant line shown in figure 3, the route of a job is deterministic. On the other hand, we can have re-entrant lines with probabilistic or Markovian routing

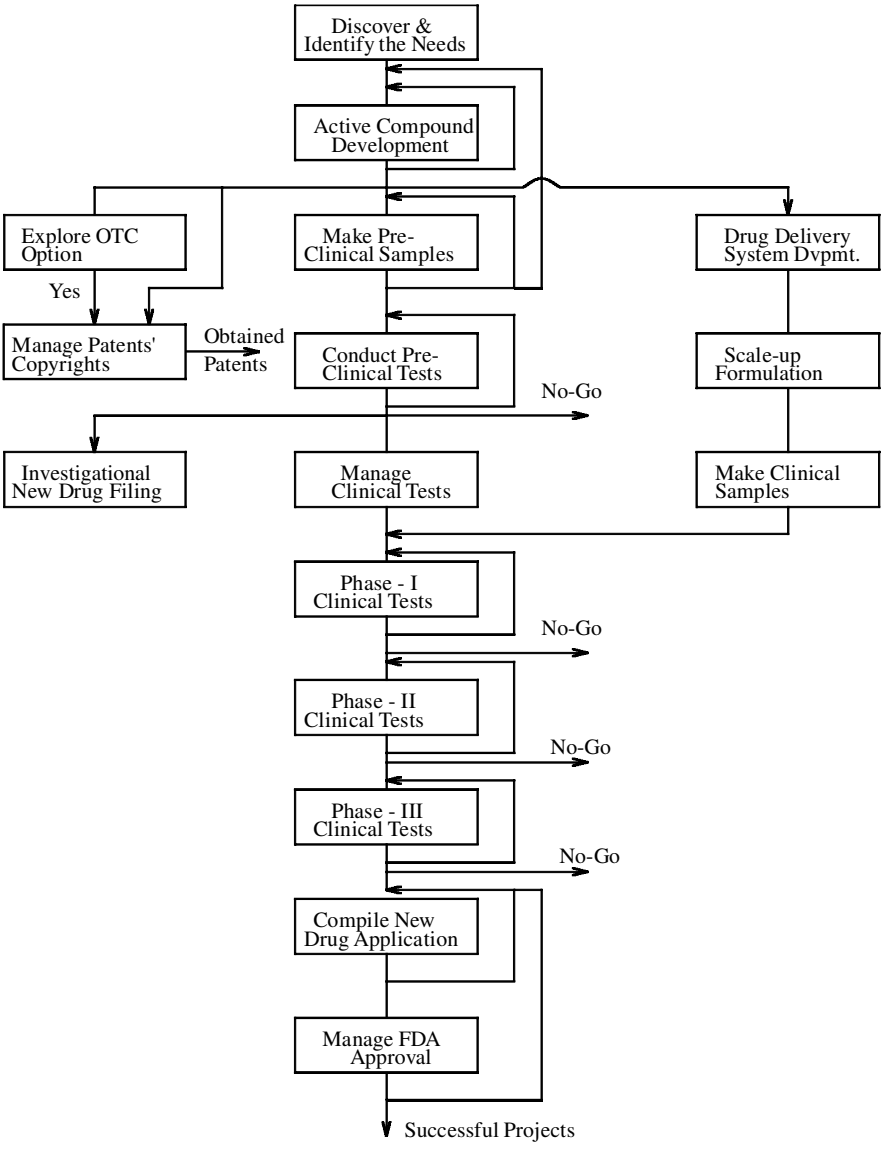

Figure 2. Project flow and decisions in a DDP.

(Narahari and Khan 1997), where we specify for each pair of buffers, say Buffer x and Buffer y, the probability $P(x, y)$ which gives the probability that a job goes to Buffer y next, after finishing its stay and service in Buffer x. The model that we develop for a multi-project DDO will be of this type.

There are two important decisions that have significant effects on the performance of a re-entrant line. These are: *input release policies*, that specify when to release fresh jobs into the system; and *scheduling policies*, that specify which job to process next when a server becomes available.

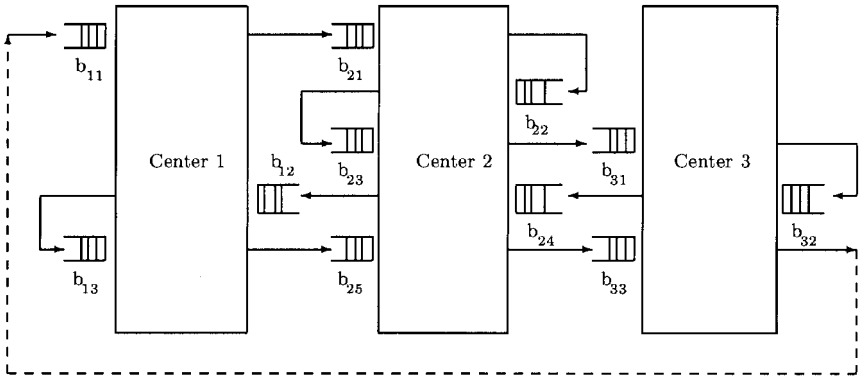

Figure 3. A re-entrant line with three stations and 11 buffers.

3.2.1. *Scheduling in re-entrant lines*

The scheduling problem in a re-entrant line becomes interesting because several parts at different stages of processing may be in contention with one another for service at the same service centre. Several researchers have studied the issue of scheduling in re-entrant lines (Kumar 1993). Distributed scheduling policies based on buffer priorities and due dates have been formulated and investigated by Kumar (1993) and Lu *et al.* (1994). Kumar (1993) has investigated, among others, the following fixed buffer priority policies: first buffer first serve (FBFS) and last buffer first serve (LBFS). In a buffer priority policy, the buffers are assigned different static priorities. After finishing the service on a part, the service centre will pick up a part from the buffer having the highest priority (if one is available, of course). For example, in the case of LBFS, we order the n_i buffers of processing centre i as $b_{in_i}, b_{i,(n_i-1)}, \ldots, b_{i2}, b_{i1}$ in decreasing order of priority. The next part selected for processing is the one that has finished most of its processing and, hence, the one with the least amount of processing remaining. Thus we may say that each processing centre *myopically* tries to clear parts from the system as fast as possible. Other popular policies are due-date based policies such as earliest due date (EDD) first and least slack (LS) first.

Fluctuation smoothing policies (Lu *et al.* 1994) are a special class of least slack scheduling policies (Kumar 1993). In the least slack policies, for every job π that enters the network, there is an associated real number $\beta(\pi)$. Also to each buffer b_{ij}, $\forall i = 1, \ldots, m; j = 1, \ldots, n_i$ there is associated a real number γ_{ij}, which is usually an estimate of the mean time a job in buffer b_{ij} will spend in the network before leaving the network. If a job is located in buffer b_{ij}, the slack $s(\pi)$, is defined by,

$$s(\pi) := \beta(\pi) - \gamma_{ij}.$$

A least slack scheduling policy gives highest priority to the job π for which the slack is minimum. Whenever the server is to choose the next part after a service completion, it selects a part with the least slack. Now a particular choice of $\beta(\pi)$ and γ_{ij} will give the particular least slack policy a unique capability. See Lu *et al.* (1994) for a good overview of fluctuation smoothing policies. We shall look at three such fluctuation smoothing policies.

Reducing the variance of lateness. Suppose each job π (project in our case) has an associated due date $d(\pi)$ (delivery time promised for the project). If we choose $\beta(\pi) = d(\pi)$, the resulting scheduling policy is found to reduce the variance of lateness of jobs and is called the fluctuation smoothing policy for variance of lateness (FSVL).

Reducing the variance of cycle-time. If we make $\beta(\pi)$ the time at which the resulting scheduling policy is found to reduce the variance of cycle time of jobs, is called the fluctuation smoothing policy for variance of cycle time (FSVCT).

Reducing the mean cycle-time. Suppose π is the nth job entering the network and λ is the average arrival rate of jobs into the network. If we choose $\beta(\pi) = n/\lambda$,the resulting scheduling policy is found to reduce the mean cycle time of jobs and is called the fluctuation smoothing policy for mean cycle time (FSMCT).

3.3. *Re-entrant line model of a multi-project drug development organization*

Figure 4 depicts a re-entrant line model of a typical multi-project DDO. Each node here represents an aggregated, parallel set of activities and the single server in each node is a functional or cross-functional team carrying out this set of activities. Table 1 describes the function of these seven nodes. It is possible that a given human resource is involved in two or more of the parallel activities corresponding to the given node. The server is thus a conglomerate of all these human resources and the service time corresponds to the most time-consuming activity among the parallel activities. This is the reason why the model is coarse and aggregates much detail. However, it is possible to parameterize such a model and infer very useful issues and insights by experimenting with the model.

At each of the nodes, there are multiple buffers that contain queued up jobs. Various new drug development projects that are in progress in different stages

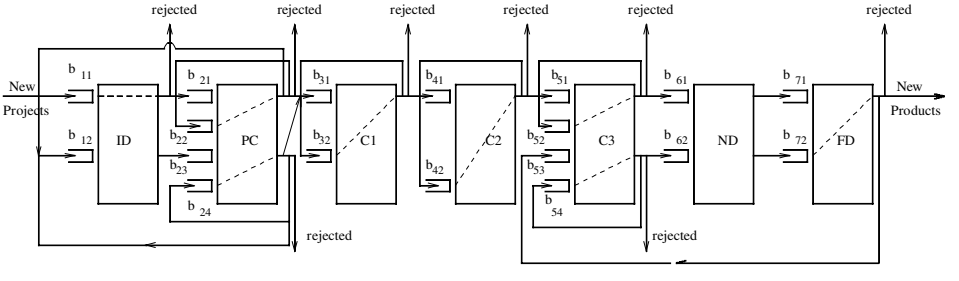

Figure 4. A re-entrant line model of a DDO.

Description		Buffers
ID	Initial discovery	b_{11}, b_{12}
PC	Pre-clinical trials	b_{21}, b_{22}, b_{23}, b_{24}
C1	First phase of clinical trials	b_{31}, b_{32}
C2	Second phase of clinical trials	b_{41}, b_{42}
C3	Third phase of clinical trials	b_{51}, b_{52}, b_{53}, b_{54}
ND	Filing for New Drug Application	b_{61}, b_{62}
FD	Review by Food and Drugs Authority	b_{71}, b_{72}

Table 1. Description of the nodes of the re-entrant line model.

represent the customers or jobs in the network. The buffers at a node contain jobs visiting the node for the first time, for the second time, etc. Note that new projects first enter the buffer b_{11} at the ID node and successfully completed projects (successfully developed new drugs) leave the network from node FD. Each project undergoes a sequence of activities in the manner shown in the network. A project obviously can visit a node several times because of reworks.

For our experimentation, we assume that the inter-arrival time between successive new project initiations and also the service times at various nodes are distributed according to a probability distribution whose mean and variance are known. These values can be obtained using measured data from a DDO. Company specific data is, however, scarcely available in the public domain, many recent web sites provide empirically observed values for these durations (Web 5–8).

The re-entrant line model contains seven stations (nodes) and 18 buffers. The multiple buffers at a given station contain projects that are revisiting that station for rework from different stages. For example, Station 5 (C3) has four buffers, b_{51}, b_{52}, b_{53}, b_{54}. Buffer b_{51} contains projects that are visiting this node for the first time. Buffer b_{52} contains projects that are visiting this station for rework after having undergone an operation or rework at this stage and before going to the ND node. The jobs in buffer b_{53} correspond to those projects that are visiting this station for rework after having undergone an operation or rework at the FD node. Likewise, we can describe the jobs in the other buffers also. The processing time distributions could be different for customers in different buffers. A routing probability has a source buffer and a destination buffer and describes the (Markovian) probability that a project, after completing service at the source buffer, next queues up for service at the destination buffer.

The re-entrant line model is thus described by the following parameters: number of nodes, number of buffers at each node, mean and variance of inter-arrival time between successive new project initiations, mean and variance of the service time distribution at each buffer, routing probabilities, and the scheduling policy to be followed at each node. Sophisticated scheduling policies can be defined to select which buffer and job to process next at a given station. The policies include: buffer priority policies, such as FBFS or LBFS (Kumar 1993), due-date based policies (Kumar 1993), and fluctuation smoothing policies (Lu *et al.* 1994).

Based on the data available from many web sites (Web 1–9) and published data, we have chosen the mean service times for the seven servers in our model as shown in table 2. The units are in months. Also, based on the above data, we have chosen the routing probabilities as in table 3.

We can now set up a base model for our experimentation. Assume that new drug development projects corresponding to the most promising drug formulations are initiated on an average once every year. If we assume that the arrivals of new development projects are a Poisson process, its rate therefore would be 1 per year.

Station	ID	PC	C1	C2	C3	ND	FD
Mean service time on first visit (months)	36	42	12	24	36	19.2	9.5

Table 2. Mean service time for the queueing network.

Source buffer	Destination buffer	Routing probability
ID – b_{11}	PC – b_{21}	0.5985
ID – b_{12}	PC – b_{23}	0.9025
PC – b_{21}	ID – b_{12}	0.2
PC – b_{21}	PC – b_{22}	0.099
PC – b_{21}	PC – b_{31}	0.00665
PC – b_{22}	ID – b_{12}	0.1
PC – b_{22}	PC – b_{22}	0.099
PC – b_{22}	C1 – b_{31}	0.76
PC – b_{23}	ID – b_{12}	0.02
PC – b_{23}	PC – b_{24}	0.0475
PC – b_{23}	C1 – b_{31}	0.893
PC – b_{24}	ID – b_{12}	0.01
PC – b_{24}	ID – b_{24}	0.019
PC – b_{24}	C1 – b_{31}	0.893
C1 – b_{31}	C1 – b_{32}	0.2375
C1 – b_{31}	C2 – b_{41}	0.665
C1 – b_{32}	C1 – b_{32}	0.099
C1 – b_{32}	C2 – b_{41}	0.8075
C2 – b_{41}	C2 – b_{42}	0.38
C2 – b_{41}	C3 – b_{51}	0.4465
C2 – b_{42}	C2 – b_{42}	0.1425
C2 – b_{42}	C3 – b_{51}	0.76
C3 – b_{51}	C3 – b_{52}	0.19
C3 – b_{51}	ND – b_{61}	0.75
C3 – b_{52}	C3 – b_{52}	0.0665
C3 – b_{52}	ND – b_{61}	0.9
C3 – b_{53}	C3 – b_{54}	0.0475
C3 – b_{53}	ND – b_{54}	0.95
C3 – b_{54}	C3 – b_{54}	0.19
C3 – b_{54}	ND – b_{62}	0.5
ND – b_{61}	FD – b_{71}	1
ND – b_{62}	FD – b_{72}	1
FD – b_{71}	C3 – b_{53}	0.38
FD – b_{72}	C3 – b_{53}	0.19

Table 3. Routing probabilities for the re-entrant line model (all other probabilities are zero).

Also, assume that the service times at the stations are all mutually independent exponential random variables. Let the mean processing time of jobs visiting a node for the first time be given as in table 2 and assume that the mean processing time of jobs revisiting the node is half of that for the first visit. Furthermore, assume that the routing probabilities are as in table 3 and non-pre-emptive FCFS (First Come First Served) scheduling policy is followed at each node for selecting the next job to process. Under these assumptions, it is found that the DDLT has a mean value of 13.68 years and a standard deviation of 3.638 years.

Validation of the above model is very important in view of the large number of assumptions in model formulation and in view of the fact that only empirical data are available. The estimated mean DDLT of 13.68 years agrees with the range of 12 to 14 years which is well known for drug development processes the world over. A certain amount of approximation and in accuracy is inevitable with any model of this type.

4. Compression of Drug Development Lead Time

4.1. *Lead time reduction through effective scheduling*

We now demonstrate that better lead time performance can be achieved by intelligently scheduling internal work in the product development network. We look at mean lead time and variance of lead time. We consider five different scheduling strategies: FCFS, FBFS, LBFS, FSMCT, and FSVCT.

Table 4 shows the results for a typical scenario. We have assumed Poisson arrivals and exponential processing times. Results are provided for three cases: inter-arrival time = 12 months (low IAT); inter-arrival time = 11 months (medium IAT); and inter-arrival time = 10 months (high IAT). The processing times of a job on its first visit to a station are assumed to be as in table 2. On subsequent visits to a station,the processing time means are assumed to be half of their original values. The results in table 4 are obtained by a detailed simulation under each policy, where each simulation is run to complete about 10 000 projects and the performance measures are computed after deleting an appropriate number of initial transients. A confidence level of 0.95 is considered for these results.

Note that the FSMCT and LBFS policies are attractive for minimizing DDLT while the FSVCT policy, as expected, minimizes the variance of DDLT. Applying fluctuation smoothing to improve lead time performance is a very attractive technique for lead time compression. This is because we do not need to add any additional capacity to the system resources, or incur any overheads such as rejecting some projects. We choose the way in which to prioritize work corresponding to internal flows or internal processes only. In the model, at any given station, we distinguish between work by considering the history of sojourn in the network and the due dates that different jobs are carrying and schedule the jobs depending on what is required to be minimized. In simple terms, we rush jobs that are likely to bring down the mean lead time and slow down the progress of jobs that have a much greater impact on the mean lead time. In the context of a DDO, the scheduling strategies have implications on how the internal subprocesses of various current projects are scheduled. The attractive aspect of employing these policies arises by virtue of not having to commit any additional resources or personnel for various

		Low IAT 12 months	Medium IAT 11 months	High IAT 10 months
FCFS policy	Mean DDLT (years)	13.68	13.81	14.04
	Variance of DDLT	175.1	178.02	195.02
FBFS Policy	Mean DDLT (years)	13.78	14.566	14.95
	Variance of DDLT	228.93	257.55	289.98
LBFS Policy	Mean DDLT (years)	12.96	13.39	13.76
	Variance of DDLT	134.76	147.75	152.93
FSMCT Policy	Mean DDLT (years)	11.63	11.93	12.15
	Variance of DDLT	161.91	184.49	191.92
FSVCT Policy	Mean DDLT (years)	13.68	13.84	13.99
	Variance of DDLT	90.44	100.12	103.42

Table 4. Performance of a DDO under fluctuation smoothing policies (IAT, inter-arrival time).

project tasks. We are only prioritizing individual work elements in an appropriate way. A drug development manager can implement this effectively.

4.2. *Lead time reduction through critical mass-based project management*

Critical mass is that property of a research team provided by the presence of several interactive and mutually catalytic scientists who function collectively to produce a dynamic research intellect (Basa and Allen 1994). The importance of critical mass for achievement of excellence in pharmaceutical product development is well recognized. Clinical drug development via a critical mass research team is capable of speeding up the DDP.

The effect of critical mass-based project teams is similar to that of cross-functional teams advocated in the concurrent engineering literature (Gatenby *et al.* 1994). It is quite a standard argument in the concurrent engineering literature that, by making product development work more cross-functional, the rework loops are reduced. At the same time, each individual activity will need more resources, more discussion, more interaction, and consequently more time. In a DDP, such cross-functional teams can potentially be employed in clinical trials stages C1, C2, and C3 (see figure 4). To capture the effect of these teams, we can decrease the feedback probabilities concerning these stages and simultaneously increase the processing times at these stages. In our experiment, we have increased the processing times of stages C1, C2, and C3 by 20% to account for the additional time entailed by increased cross-functional work and assumed that the associated feedback probabilities are reduced by 10%. If we do this and evaluate the lead time performance, we get results as shown in table 5. The gains in lead time performance are quite clear, if we compare the results in tables 4 and 5.

4.3. *Further opportunities for lead time compression*

There are numerous other opportunities for lead time reduction that can be investigated. many such experiments, in the context of factory floors and product design, are reported in Hopp *et al.* (1990), Harrison and Lock (1995), Adler *et al.* (1995), Hopp and Spearman (1996), Suri (1996), Narahari *et al.* (1999). Well known techniques include: reducing the variability of processing times through strict process

		Low IAT 12 months	Medium IAT 11 months	High IAT 10 months
FCFS Policy	Mean DDLT (years)	13.13	13.411	13.763
	Variance of DDLT	160.4	169.92	178.44
FBFS Policy	Mean DDLT (years)	13.52	13.88	14.445
	Variance of DDLT	209.45	223.37	255.43
LBFS Policy	Mean DDLT (years)	12.78	13.02	13.34
	Variance of DDLT	120.33	138.23	145.87
FSMCT Policy	Mean DDLT (years)	11.40	11.65	11.98
	Variance of DDLT	144.43	167.49	181.95
FSVCT Policy	Mean DDLT (years)	13.20	13.34	13.70
	Variance of DDLT	80.22	94.18	101.44

Table 5. Performance of a DDO with critical mass-based project teams (IAT, inter-arrival time).

control, reducing the variability of inter-arrival times through input control and load balancing among different stages of processing. Such experiments can be easily carried out on the re-entrant line under study.

5. Conclusions

In this paper, we have exploited the discrete event dynamical nature of a product development process and modelled it in the same manner as a production process using queueing networks. The project dynamics in a drug development organization is slow compared with that of a production processor a factory floor process. However, we have demonstrated the validity of queueing network models in capturing the slow dynamics adequately enough to reveal insights into lead time reduction.

Needless to say, the model presented is still not representative of all the details and distinctive aspects of a DDO. Only a comprehensive simulation model provides a partial answer to the problem of creating a faithful replica of a given DDO. What we have attempted and succeeded in doing here is to come up with a good analytical model that captures certain important performance determinants of a DDO and used the model towards a deeper understanding of project management issues. The model captures the effect of various lead time reduction strategies at the level of abstraction of a product development manager. Such models can be used by managers in aggregate project planning and project management. In order to use such models in the detailed planning of projects in a multi-project DDO, one has to enrich the models and also the analysis techniques. In this sense, the paper certainly throws open several interesting issues for further investigation.

The models described can become the foundation of a software tool that can be used by managers in multi-project DDOs.

Acknowledgements

This research was supported in part by the Office of Naval Research and the Department of Science and Technology grant N00014-93-1017. It was also supported by a Research Grant from the Ministry of Human Resources Development, New Delhi, on Competitive Manufacturing Systems. We should like to thank Rahul Rathore for his active participation in the initial parts of this research during 1996. He has contributed to parts of § 2 and § 3.1 and also prepared the figures. The simulation results here were obtained using a software package developed by Palla Nagaraju. The simulations were carried out by Debjit Roy.

References

ADLER, P. S., MANDELBAUM, A., NGUYEN, V. and SCHWERER, E., 1994, From project to process management: Strategies for improving development cycle time. Technical report, Sloan School of Management, Massachusetts Institute of Technology.

ADLER, P. S., MANDELBAUM, A., NGUYEN, V. and SCHWERER, E., 1995, From project to process management: An empirically-based framework for analyzing product development time. *Management Science*, **41**, 458–484.

ADLER, P. S., MANDELBAUM, A., NGUYEN, V. and SCHWERER, E., 1996, Getting the most out of your product development process. *Harvard Business Review*, **22**, 134–152.

BASA, F. and ALLEN, T. J., 1994, Managing the drug development process: A preliminary report on project management in the pharmaceutical industry. Technical Report 9-94, Sloan School of Management, Massachusetts Institute of Technology.

BOURLAND, K., 1994, Cellular production at spartan. Technical report, Teaching Case, Tuck School of Business, Dartmouth College, Hanover, New Hampshire.

BOURLAND, K. and SURI, R., 1992, Spartan industries. Technical report, Case Study,Tuck School of Business, Dartmouth College, Hanover, New Hampshire.

COHEN, M. A., ELIASHBERG, J. and HO, T. H., 1997, New product development: The performance and time-to-market tradeoff. *Management Science*, **42**, 173–186.

GATENBY, D. A., LEE, P. M., HOWARD, R. E., HUSHYAR, K., LAYENDECKER, R. and WESNER, J., 1994, Concurrent engineering: an enabler for fast, high quality product realization. *AT & T Technical Journal*, pp. 34–47.

HARRISON, J. M. and LOCH, C. H., 1995, Operations management and reengineering. Technical report, The Graduate School of Business, Stanford University.

HAUPTMAN, O. and HIRJI, K. K., 1996, The influence of process concurrency on project outcomes in product development: An empirical study of cross-functional teams. *IEEE Transactions on Engineering Management*, **43**, 153–164.

HOPP, W. J. and SPEARMAN, M. L., 1996, *Factory Physics: Foundations of Manufacturing Management* (Irwin: McGraw-Hill).

HOPP, W. J., SPEARMAN. M. L. and WOODRUFF. D. L., 1990. Practical strategies for lead time reduction. *Manufacturing Review*, **3**, 78–84.

KNOOP, S. J. and WORDEN, D. E., 1988. The pharmaceutical drug development process: An overview. *Drug Information Journal*, **22**, 259–268.

KUMAR, P. R., 1993, Re-Entrant lines. *Queueing Systems: Theory and Applications*, **13**, 87–110.

LU, S. H., RAMASWAMY, D. and KUMAR, P. R., 1994, Efficient scheduling policies to reduce mean and variance of cycle-time in semiconductor manufacturing plants. *IEEE Transactions on Semiconductor Manufacturing*, **7**, 374–388.

NARAHARI, Y. and KHAN, L. M., 1997, Modeling re-entrant manufacturing systems with inspections. *Journal of Manufacturing Systems*, **16**, 367–378.

NARAHARI, Y., VISWANADHAM., N. and KIRAN KUMAR, V., 1999, Lead time modeling and acceleration of new product development. *IEEE Transactions on Robotics and Automation*, **15**, 882–896.

SARANTAPOULOS, P. D., ALTIOK, T. and ELSAYED, E. A., 1995, Manufacturing in the pharmaceutical industry. *Journal of Manufacturing Systems*, **14**, 452–467.

SMITH, P. and REINERTSEN, D., 1991, *Developing Products in Half the Time* (New York: Van Nostrand Reinhold).

SURI, R., 1996, Using queueing networks to support quick response manufacturing. *Proceedings of the IIE Conference of the Institute of Industrial Engineers*, USA, pp. 603–609.

VISWANADHAM, N. and NARAHARI, Y., 1992. *Performance Modeling of Automated Manufacturing Systems* (Englewood Cliffs, NJ: Prentice Hall).

WEB 1 World Wide Web, http://www.pfizer.com/pfizerinc/about/pubs/journal/opportunity

WEB 2 World Wide Web, http://www.hcnor.com/rx-help/approval.html

WEB 3 World Wide Web, http://www.pharma.org/charts/approval.html

WEB 4 World Wide Web, http://www.wabio.com/dis/drug-approv.html

WEB 5 World Wide Web, http://www.mhsource.com

WEB 6 World Wide Web, http://www.allp.com/drug-dev.htm

WEB 7 World Wide Web, http://www.critpath.org/research/process.htm

WEB 8 World Wide Web, http://www.fda.gov/fdac/special/newdrug/speeding.html

WEB 9 World Wide Web, http://www.fda.gov/cder/rdml

WHEELWRIGHT, S. C. and CLARK, K. B., 1992, *Revolutionizing Product Development: Quantum Leaps in Speed, Efficiency and Quality* (New York: The Free Press).

YEVICH, J. P., 1991, Drug development: From discovery to marketing. In *A Text-book of Drug Design and Development* (Reading, MA: Harwood), pp. 606–630.

ZIRGER, B. J. and HARTLEY, J. L., 1996, The effect of acceleration techniques on product development time. *IEEE Transactions on Engineering Managements*, **43**, 143–150.