Forecasting Resource Requirements for Drug Development Long Range Planning

by

Angela Thedinga

B.S., Chemical Engineering
University of Wisconsin – Madison, 2004

Submitted to the MIT Sloan School of Management and the
Department of Chemical Engineering
in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration
and
Master of Science in Chemical Engineering

In conjunction with the Leaders for Global Operations Program at the
Massachusetts Institute of Technology
June 2010

ARCHIVES

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 0 8 2010
LIBRARIES

© Massachusetts Institute of Technology
All rights reserved

Signature of Author_____
May 7th, 2010
MIT Sloan School of Management, Department of Chemical Engineering

Certified by_____
Ernst Berndt, Thesis Supervisor
Louis E. Seley Professor of Applied Economics

Certified by_____
Charles Cooney, Thesis Supervisor
Robert T. Haslam Professor of Chemical Engineering

Accepted by_____
Debbie Berechman
Executive Director of MBA Program, MIT Sloan School of Management

Accepted by_____
William Deen, Professor of Chemical Engineering
Chairman, Committee for Graduate Students Chemical Engineering

This page has been intentionally left blank.

Forecasting Resource Requirements for Drug Development Long Range Planning

by

Angela Thedinga

Submitted to the MIT Sloan School of Management and the Department of Chemical Engineering on May 7th, 2010 in Partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Chemical Engineering

ABSTRACT

This thesis investigates the use of a task-based Monte Carlo simulation model to forecast headcount and manufacturing capacity requirements for a drug development organization. A pharmaceutical drug development group is responsible for designing the manufacturing process for new potential drug products, testing the product quality, and supplying product for clinical trials. The drug development process is complex and uncertain. The speed to market is critical to a company's success. Therefore, it is important to have an adequate number of employees and available manufacturing capacity to support timely and efficient drug development. The employees and manufacturing capacity can either be supplied internally or externally, through contract manufacturing organizations.

This thesis formulates and empirically evaluates a simulation model designed using the Novartis Biologics drug development process and is adaptable to other pharmaceutical organization. The model demonstrates 7% accuracy when compared with historical data, and estimates within 13% of the currently accepted manufacturing capacity forecasting tool. Additionally, three case studies are included to demonstrate how the model can be used to evaluate strategic decisions. The case studies include: a drug development process improvement evaluation, an outsourcing evaluation, and an "at risk" development evaluation.

Thesis Supervisor: Ernst R. Berndt
Title: Louis E. Seley Professor of Applied Economics

Thesis Supervisor: Charles L. Cooney
Title: Robert T. Haslam Professor of Chemical Engineering

ACKNOWLEDGEMENTS

I have been fortunate to work with many excellent people. I would like to express my appreciation to the following individuals:

Don Rosenfield and Leaders for Global Operations (LGO) Staff – Thank you for your dedication to developing future operations leaders. Your commitment to continuous improvement shows, as you have crafted a phenomenal educational experience. Thank you!

Philippe Marchal, Corinna Sonderegger, Glen Swistara and Yuan Xu at Novartis – A sincere thank you for your support and guidance through the research internship. The experience at Novartis Biologics in Basel, Switzerland has been amazing and I have learned so much through your support.

Professors Ernst Berndt and Charles Cooney – Thank you for your mentorship and guidance from research approach to career advice. I appreciate your time and support.

Doug Fearing, Nichole Rothaupt, and Margo de Naray – Thank you for editing.

My LGO classmates – Thank you for sharing your insights and experiences. I cannot count how many times I have leaned on you for your expertise and/or advice, and I look forward to keeping in contact in the future.

My parents, Don and Teresa Cleaver – Thank you both for your support. Dad, thank you for helping me with the paper route and teaching me the value of hard work. Mom, thanks for teaching me how to think creatively.

My neighbors, Lisa Coughlin and Scott Apher – Thank you for all of the love and care you gave to Lola during these past two years. Paul and I will never forget your help.

My husband, Paul Thedinga – Thank you for your support throughout this two year journey. You have graciously made adjustments and sacrifices to support my education and our future. I am so grateful, and am excited to start our next journey together.

Angela Thedinga
Massachusetts Institute of Technology
Cambridge, Massachusetts
May 7th, 2010

Table of Contents

FIGURES AND TABLES

CHAPTER 1 – Introduction and Overview

1.1 Problem Statement

The drug development process is complex and with many uncertainties that complicate forecasting headcount and plant capacity requirements. The ability to forecast resources is important because speed to market is critical and delays are costly. Business decisions involving plant expansions and headcount expansions require foresight because the actual decision making occurs years before benefits are fully realized. New employees can take up to a year to recruit, hire and fully train, and new manufacturing capacity can take three to five years to build and validate. With the slow moving, complex, and uncertain drug development process, it is very challenging to estimate resource needs into the future.

1.2 Hypothesis

Employee and plant capacity requirements for the complex and variable drug development process can be generalized, modeled, and presented in ways useful for long term strategic decision making.

1.3 Thesis Goals

The goal of this thesis is three fold:

- Simulate the drug development process in a flexible and scalable model
- Evaluate the reliability and accuracy of the model
- Demonstrate how the model can be used for strategic decision making

In order to achieve these goals, data from Novartis Biologics, which focuses on large molecule drug development, will be used to design and test the model.

1.4 Results

Through this thesis research at Novartis Biologics in Basel, Switzerland, a task-based Monte Carlo simulation model was created to forecast drug development resource requirements. The model forecasts the number of required headcount, measured in full time equivalents (FTEs[1]), and the manufacturing capacity, measured in the number of campaigns or manufacturing weeks. The model was designed to be flexible and scalable using Microsoft Excel, coupled with Oracle Crystal Ball software. The model is flexible in that tasks and new projects are easily assigned to new functional groups or to new sites. The model is scalable in that new product types can be added by copying the existing templates and modifying the tasks to represent that new product type. The model structure described in this thesis could be used for any pharmaceutical development process by changing the input variables and assumptions as appropriate.

When comparing the model projections against actual Novartis data, the model forecasts within 7% of the actual FTEs hired and estimates manufacturing capacity requirements within +/- three manufacturing campaigns. The model forecasts the manufacturing utilization within 13% of the currently accepted forecasting method. When compared with the overall industry data for large molecule development, the model overestimates the actual number of launched products by up to 90%. This percentage suggests that even though the model closely estimates the Novartis pipeline, the overall industry trends differ by company.

This thesis also demonstrates that the model is not only useful for forecasting future resources, but also for evaluating different scenarios to make strategic decisions regarding resource or capacity planning. The three case studies presented in this thesis are: a drug development process improvement evaluation, an outsourcing evaluation, and an "at risk" development evaluation.

This thesis meets the goals stated above and concludes with a number of recommendations. First, a successful model depends on reliable data input. To ensure that the data are as accurate

[1] Many companies measure human resources in terms of full time equivalents (FTEs). Typically, one FTE represents the work performed by one full time equivalent employee working for 40 hours per week. This unit is used to measure resources required for tasks, or asses the amount of resources available. If there is a task that requires 20 hours per week every week, then that task requires 0.5 FTE. If a part-time employee works 30 hours per week, then they are considered 0.75 FTE.

as possible, it is recommended to create a structured process that gathers model input on a periodic basis and delivers forecast results to the appropriate stakeholders in a consistent and timely fashion. Second, the model currently considers a standard set of assumptions for all projects. However there is also more specific resource requirement information for the projects that are currently in progress. By incorporating the specific information about current projects and keeping the general information for future projects, the short term forecasting accuracy of the model can be improved. Lastly, the model could be expanded for use in other areas of Novartis or it could be used in more detail or granularity within Novartis Biologics. It will be important to choose a specific application for the model and simplify the assumptions and structure to best serve its overall objective.

1.5 Thesis Overview

This thesis is organized into six chapters. The first chapter (this chapter) provides an introduction and an overview to the thesis and its contents. The second chapter summarizes the drug development process, reasons for forecasting, and forecasting techniques. Chapter three details the model framework, assumptions, and calculations. The fourth chapter evaluates this model using three perspectives: (1) comparing model projections against historical project pipeline data from the pharmaceutical industry, (2) comparing model projections against historical Novartis headcount and manufacturing data, and (3) comparing model projections against current manufacturing capacity planning methods. Chapter five demonstrates how this model can be used to provide information for strategic decision making. The specific evaluations are: a drug development process improvement evaluation, an outsourcing evaluation, and an "at risk" development evaluation. Lastly, chapter six summarizes the key findings and conclusions.

CHAPTER 2 – Background

This chapter outlines the drug development process, and provides reasons and common practices for forecasting drug development resources.

2.1 Drug Development Process in the BioPharmaceutical Industry

Today there are two main types of pharmaceutical products: "small molecule" and "large molecule" products. Small molecule products can be chemically synthesized and large molecule products are grown in living cells that are genetically altered to produce the desired product (Schuler, 2002). The large molecule products are also called "biologics." Since the manufacturing equipment required for these two methods varies greatly, pharmaceutical manufacturing capacity is divided to produce either one of these two products.

This thesis is supported by data collected at Novartis Biologics. The parent company, Novartis AG, develops both new molecular entities (NME), which are new drugs under patent protections, and generics, which are drugs off patent protection. Traditionally, generics have been small molecule products; however, some of the first large molecule products have reached patent expiration. This has launched the emergence of "biosimilars," which are the generic versions of large molecule drug products. Novartis AG has launched the first two FDA and/or EU approved biosimilars, Omnitrope® and Binacrit®/Hexel®, under the brand name "Sandoz" (Novartis AG, 2010).

Novartis Biologics supports both NME and generic products. New molecular entity and generic products are manufactured using one of two manufacturing methods, cell culture or microbial fermentation. The four product and manufacturing combinations are:

- Cell culture – New molecular entities (NMEs) that are produced using mammalian cell culture are classified as "cell culture" products. This is typically, but not limited to, monoclonal antibodies.
- Microbial – NMEs that are produced using a microbial expression system such as *E. coli* or yeast are classified as "microbial" products. This manufacturing method requires different manufacturing steps and different equipment than cell culture products.

- Biosimilar cell culture – Generic version of a drug that is nearing patent expiration and is produced using mammalian cell culture are classified as "biosimilar cell culture" products.
- Biosimilar microbial – Generic version of a drug that is nearing patent expiration and is produced using microbial expression systems are classified as "biosimilar microbial" products.

Drug development covers starts when a new potential drug is discovered and continues until the commercial launch of that product. Drug development organizations do not produce product for sale, but rather product for testing to collect information regarding:

- Drug manufacturing requirements, such as Chemistry, Manufacturing and Controls (CMC) information required by regulatory agencies like the FDA and EMEA
- Drug safety and efficacy data, supported through animal studies and clinical trials
- Drug stability data

In order to obtain this information, development groups must perform numerous small scale manufacturing runs, develop analytical test methods, and execute those tests many times. Since an organization produces information, the "demand" is the number of projects requiring development and testing. In this way, it is the project that moves from group to group, while the headcount resources and laboratory space are analogous to stationary manufacturing equipment on an assembly line (Repenning, 2000). In this case, the "product" is either the information generated by the functional groups or the clinical material that will eventually produce information through clinical trial results. The overall development process for each product is illustrated in Figures 1 and 2. The differences between NME and biosimilar development paths are due to the different regulatory requirements.

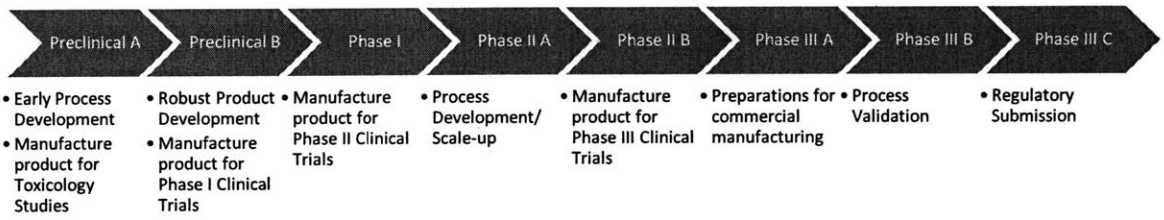

Figure 1 - New Molecular Entity (cell culture and microbial) Development Path

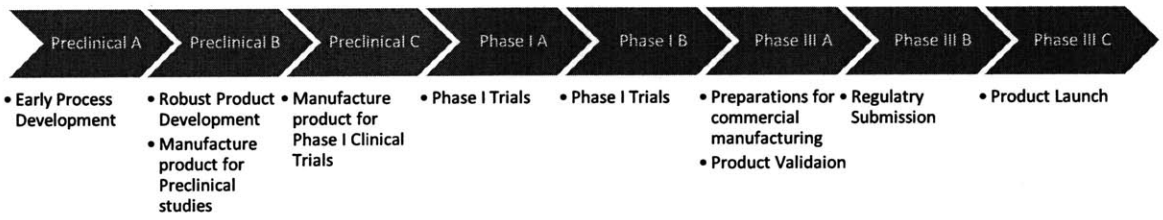

Figure 2 - Biosimilar (biosimilar cell culture and biosimilar microbial) Development Path

Each drug typically takes seven to nine years to complete development (Parexel, 2009). Many drug projects are terminated for a variety of reasons, which range from toxicological concerns to poor clinical results, or even a saturated market with too much competition for it to launch. The probability that a project will proceed in the drug development process is measured as the "probability of success."

Different functional groups are needed during the various development stages. For instance, formulation resources are required in the very beginning of product development, as well as the very end when the product is launched and new dosage levels are requested. In contrast, the drug product analytics are frequently tested and process analytical resources are required in all eight steps.

In addition to needing specific resource types, different drug projects will require varying levels of those resources. For instance, an "easy" drug candidate that can be manufactured using well established manufacturing methods and evaluated with known test methods will require fewer

resources than a drug candidate with unusual manufacturing characteristics or undeveloped test methods. Other factors that influence resource requirements include drug delivery methods, number of formulations, and manufacturing yields.

Typically, large pharmaceutical companies like Novartis will have a number of projects in the "pipeline," meaning that they are in various stages of development. Novartis Biologics is experiencing significant growth as large molecule products are representing an increasing portion of the overall Novartis pipeline (Novartis Annual Report, 2009). With the growing pipeline, Novartis Biologics is evaluating headcount resource and manufacturing capacity requirements to ensure quick and efficient drug development.

2.2 Forecasting Practices

Companies forecast because some decisions must be made years before the impact of that decision is known. If a company chooses to expand manufacturing capacity it can take three to five years to purchase land, design and construct facilities, and gain manufacturing approval from regulatory agencies. Drug development requires employees with specific talent sets. The recruiting, hiring, and training of these employees can take up to a year for some positions. Additionally, office space and laboratory space can take two years to build and commission. For these reasons, pharmaceutical companies commonly create strategic plans that outline expected manufacturing, headcount and office space requirements for the upcoming five years. The purpose of forecasting is to identify where resource or capacity gaps exist, and decide how to address them.

There are a number of ways to forecast. Many times, management will request functional groups to provide estimated resource requirements for a single predicted pipeline scenario, specifying the number of projects expected in each development stage for upcoming years. Then, all resource decisions are made based on that scenario. However, it is nearly impossible to predict the exact scenario that will occur. Additionally, it can be very time consuming for management to collect and assemble the data from many sources; and once the information is gathered, this information is only relevant for a single scenario. If the main scenario or business strategy

changes, this manual forecasting process must be repeated. This process of manually collecting the information unnecessarily wastes time.

Alternatively, modeling and simulation can also be used for forecasting. In modeling, one uses a set of assumptions to calculate an expected outcome. Simulation involves varying those assumptions in a way that represents the many possible scenarios that could occur. These scenarios are tracked to understand the range of possible outcomes. Using modeling and simulation for strategic planning began in the early 1970's when computers became a helpful tool for rapid computing (Krueger, 1972). Healthcare companies have now started using computer modeling for drug development and strategic capacity planning (Hynes, 2007 and Heinzle, 2006). According to health care modeling expert James Stahl (Stahl, 2008), models and simulations are used for one or more of the following reasons:

- To test something that is impossible to test through direct experimentation
- To better understand or predict the outcome of a complex system
- To aid in decision making

In this thesis, modeling is used for the second and third reasons. Drug development is very complex with long timelines. Resource decisions must be made today in order to satisfy requirements that will occur in one to three years.

To adequately staff development projects, pharmaceutical management teams can either internally increase resources (headcount) or capacity (laboratory and office space, and manufacturing), or outsource to a contract research organization (CRO) or contract manufacturing organization (CMO). While outsourcing may be more expensive on a per man-hour basis, it provides flexibility that capacity and headcount expansion cannot. Furthermore, many industries face a long capacity expansion lead time. Therefore, capacity must be built while products are in development, at a time when the success of each specific product is uncertain. This could be a large reason for the steady increase of CRO and CMO involvement in drug development and clinical manufacturing (Parexel, 2008). Alternatively, CMOs offer efficiency in building capacity according to the overall industry needs. Rather than individually building an overabundance of capacity, pharmaceutical companies can "share" this capacity risk through CMO contracts. In a way, the pharmaceutical companies can use this as a hedging

strategy (Beckman, 2008). When deciding which activities to outsource, companies will typically hire contract organizations to perform data-intensive projects and the company will perform knowledge-intensive projects internally (Azoulay, 2004).

In order to make strategic decisions using the best information available, decision makers like to know the forecasted headcount resources and manufacturing capacity requirements as well as the variability of those requirements. This thesis develops a modeling tool to supply a better understanding of the complex drug development system in order to assist in strategic decision-making regarding future resources and capacity.

CHAPTER 3 – Methods and Tools

This chapter describes the model that forecasts resource requirements for drug development. The model foundation was initially designed by Tamara Conant and is outlined in her thesis, *Modeling Variability for Biologics Strategic Planning* (Conant, 2009). Conant created a model to forecast FTE resources for Novartis Biologics. As part of this thesis research, the model was modified by the current author to also forecast manufacturing capacity and allocate FTE resource and manufacturing capacity requirements per site.

3.1 Problem Description

This model forecasts FTE and manufacturing capacity requirements for Novartis Biologics in future years by answering the following two questions:

- Question A: How many projects are anticipated per development stage in future years?
- Question B: What level of FTE resources, manufacturing capacity, and financial resources are required per site per year?

Novartis develops multiple product types, and each product has sequential development stages (described in Chapter 2.1). Every development stage requires an average number of workdays from different functional groups. Some functional groups are only located at one site, whereas other functional groups are available at multiple sites and might specialize on certain product types. Activities that are always performed at the same site are "fixed activities," while activities that can be performed at multiple sites are "flexible activities." Some development stages require manufacturing campaigns. The manufacturing sites available for these campaigns will depend upon the product type and development stage. Each development stage takes an average amount of time to complete; however, this number can vary depending on the product complexity.

3.2 Model Assumptions and Input Data

The model is based on fixed and flexible assumptions. All user adjustable parameters are flexible assumptions and can be changed according the user's information. Flexible assumptions

are listed below. The variable names are given for reference, and will be further defined later in this section.

- Number of current projects ($c_{p,d}$)
- Number of future projects ($f_{y,p}$)
- Probability of success between development stages ($S_{p,d}$)
- Percent of projects performed at risk ($R_{p,d}$)
- Resources required per development stage and per functional group ($\mathbf{FTE_{p,s}}$, $\mathbf{MC_{p,s}}$, $\mathbf{MS_{p,s}}$, for both fixed and flexible activities). See Appendix D for more information on how this information is obtained.
- Site allocation of flexible activities that can be performed at multiple sites ($x_{p,c,s}$)
- Time required for each development stage (T_d)

Fixed assumptions are inherent to the model design and cannot be changed without changing the model structure. Fixed assumptions are listed here:

- All projects are represented by a "standard" project, requiring an average number of resources (FTEs, manufacturing capacity, and financial resources).
- The resources required for a "standard" project this year are assumed to be the same number of resources required for that type of project for the next ten years. No organizational efficiency, productivity improvement, or technological advances are factored into the model.
- All employees within the same functional group at any site are equally capable of working on any type of project, and each employee can work on an unlimited fraction of projects. Each FTE is only limited by the number of work days in the year.
- Manufacturing sites are interchangeable and campaigns are assigned per the site allocation ($x_{p,c,s}$), which is defined by the user.
- Activities in the next stage can only be started when all of the tasks in a previous stage are completed.
- The progression of projects through the pipeline is fractional, not discrete. In other words, if a project has a 50% chance of progressing from stage 4 to stage 5 ($S_{p,4} = 0.50$), and for one project in stage 4, the model will consider 0.50 projects progressing to stage 5.

- The model assumes risk neutrality and uses expected values. Fractional projects are added. For example, four projects that all have a 50% chance of progressing are considered two full projects.
- The organization is split into a finite number of groups and will remain that way for the period of the forecast (model does not predict organizational structure changes).
- The model assumes that in each simulation run, a development stage takes the same amount of time for all new molecular entity or biosimilar projects. For instance, if the simulation generates a 1.5 year duration for development stage 4, then all projects will take 1.5 years to complete stage 4 for all ten years of the forecast.
- Regardless of the phase duration time, the work required for that development stage is spread evenly over the years required to complete the work.
- The model assumes that all future projects will start in development stage 1. This assumption is not true for many in-licensed projects. Projects that are currently in-licensed are included in the model by assigning the appropriate development stage to the current project parameter, but all future projects are assumed to enter the pipeline at development stage 1.

The assumptions listed above are important to understand when using the model for decision making purposes. In the future, if any assumptions become inappropriate the model should be adjusted according to the change required. Within Novartis, there are two individuals who serve as the NME and biosimilar resource and capacity planning heads. These two individuals are the model owners and are responsible for maintaining the model and communicating assumptions and results to Novartis decision makers.

The data-oriented notation used in the model description is listed below. This information is supplied by the model user. Individual variables are in italics, and matrices are in bold.

$c_{p,d}$ = number of current (ongoing) projects of product type, p, in development stage, d.

C = p x d matrix describing the current number of projects for each product type in each development stage.

$f_{y,p}$ = number of future projects expected in year, y, for product type, p.

F = y x p matrix describing the number of projects expected for year, y. All new projects that start each year are assumed to start at development stage 1.

$S_{p,d}$ = probability of success for product type, p, at development stage, d.

$R_{p,d}$ = fraction of projects performed at risk for product type, p, at development stage, d.

T_d = time required to complete development stage, d.

$\mathbf{FTE}_{p,s}$ = (g x d) matrix describing the number of FTEs required from each functional group, g, to complete each development stage, d. There are multiple matrices, one for each product type, p, requiring resources at site, s.

$\mathbf{MC}_{p,s}$ = (g x d) matrix describing the number of manufacturing campaigns required from each functional group, g, to complete each development stage, d. There are multiple matrices, one for each product type, p, requiring resources at site s.

$\mathbf{MS}_{p,s}$ = (g x d) matrix describing the materials and services costs required from each functional group, g, to complete each development stage, d. There are multiple matrices, one for each product type, p, requiring resources at site, s.

$\mathbf{FTE\_flex}_{p,c}$ = (g x d) matrix describing the number of FTEs required from each functional group, g, to complete each development stage, d. This work is flexible and can be performed at a number of the sites. There are multiple matrices, one for each product type, p, requiring resources at center, c.

$\mathbf{MC\_flex}_{p,c}$ = (g x d) matrix describing the number of manufacturing campaigns required from each functional group, g, to complete each development stage, d. These campaigns can be performed at a number of the sites. There are multiple matrices, one for each product type, p, requiring resources at center, c.

$\mathbf{MS\_flex}_{p,s}$ = (g x d) matrix describing the materials and services costs required from each functional group, g, to complete each development stage, d. These costs are flexible and can be charged to a number of the sites. There are multiple matrices, one for each product type, p, requiring resources at center, c.

$x_{p,c,s}$ = fraction of flexible resources for product type, p, belonging to center, c, that are performed at site, s.

3.3 Model Calculations

As mentioned in Section 3.1, the model addresses two main questions.

Question A: How many projects are anticipated per development stage in future years?

First, a forecast vector is calculated for each current project development stage $c_{p,d}$, determining how many will proceed to the next development stage. A project will "graduate" to the next development stage if it has a successful clinical trial outcome, or if it is developed at risk[2]. The project forecast vectors are calculated as:

$$<F_{cp,d}> = <c_{p,d},\ c_{p,d}[S_{p,1}+R_{p,1}(1-S_{p,1})],\ c_{p,d}S_{p,1}\ [S_{p,2}+R_{p,2}(1-S_{p,2})],\ c_{p,d}S_{p,1}S_{p,2}[S_{p,3}+R_{p,3}(1-S_{p,3})],\ \ldots>$$

(Equation 1)

The matrix contains as many terms as the number of development stages (i.e. eight terms for eight development stages). A separate forecast vector is calculated for each $c_{p,d}$. The term will be 0 for all stages already completed and the probability of success $S_{p,d} = 1$ for all development stages that have already passed (i.e. $<F_{cp,5}>= <0, 0, 0, 0, c_{p,5}, c_{p,5}[S_{p,5}+R_{p,5}(1-S_{p,5})]$, $c_{p,5}S_{p,5}[S_{p,6}+R_{p,6}(1-S_{p,6})]$, $c_{p,5}S_{p,5}S_{p,6}[S_{p,7}+R_{p,7}(1-S_{p,7})]>$)

A forecast vector is also calculated for each of the future projects, $f_{y,p.}$ Similarly, this vector is calculated as:

$$<F_{fy,p}> = <f_{y,p},\ f_{y,p}[S_{p,1}+R_{p,1}(1-S_{p,1})],\ f_{y,p}\ S_{p,1}[S_{p,2}+R_{p,2}(1-S_{p,2})],\ f_{y,p}S_{p,1}S_{p,2}[S_{p,3}+R_{p,3}(1-S_{p,3})],\ \ldots>$$

(Equation 2)

Again, the vector contains as many terms as the number of development stages, and estimates the number of projects that start in year, y, that will proceed into the next development stage.

Time matrices (**T**) allocate the work over future years. Multiplying the forecast vector by the time matrix will allocate the resources into the appropriate year, y. Time matrix calculation is

[2] Projects developed at risk will proceed to the next development step before the clinical trial outcome is known.

described in detail in Appendix C. In general, the time matrix for current projects will be the identity matrix if each development stage lasts exactly one year and there are no time delays. The time matrix for future projects will be zeros until the year the future projects start; after that year, the time matrix will be the identity matrix.

The total number of projects in each development stage, d, over the next y years is calculated by:

$$\mathbf{N_p} = \langle F_c \rangle * \mathbf{T} + \langle F_f \rangle * \mathbf{T} \qquad \text{(Equation 3)}$$

$\mathbf{N_p}$ is a (d x y) matrix where d is development stage and y is the year. If each development stage takes one year, and there are no time delays, the formulas to calculate each element of N_p are shown below, where each row represents the number of projects in that stage of development and each column represents the number of projects in that year:

$$\boldsymbol{N_p}$$

$$= \begin{array}{cccccc} & y_0 & y_1 & y_2 & .. & y_8 \\ d_1 & c_{p,1} & f_{1,p} & f_{2,p} & .. & f_{8,p} \\ d_2 & c_{p,2} & c_{p,1}(S_{p,1} + (1 - S_{p,1})R_{p,1}) & f_{1,p}(S_1 + (1 - S_{p,1})R_{p,1}) & .. & f_{7,p}(S_1 + (1 - S_{p,1})R_{p,1}) \\ d_3 & c_{p,3} & c_{p,2}(S_{p,2} + (1 - S_{p,2})R_{p,2}) & c_{p,1}S_1(S_2 + (1 - S_{p,2})R_{p,2}) & .. & f_{6,p}S_1(S_2 + (1 - S_{p,2})R_{p,2}) \\ .. & .. & .. & .. & .. & ... \\ d_8 & c_{p,8} & c_{p,2}(S_{p,8} + (1 - S_{p,8})R_{p,8}) & c_{p,7}S_7(S_8 + (1 - S_{p,8})R_{p,8}) & .. & f_{1,p}S_1S_2S_3S_4S_5S_6S_7(S_8 + (1 - S_{p,8})R_{p,8}) \end{array}$$

(Equation 4)

Example

To illustrate the model equations above, consider the following simplified case with two products ($p = \{p_1,p_2\}$), forecasting this year and the next two years ($y = \{y_0,y_1,y_2\}$), and given that each product has two development stages ($d = \{d_1,d_2\}$). Assume that no steps are performed at risk ($R_{p,d}=0$), and the probability of success between development stage 1 and development stage 2 is 50% ($S_{1,1} = S_{2,1} = 0.5$). Also assume that each development step takes one year, and there is no time delay ($T_d = 1$ for both development stages). The current projects expected for this year ($y = y_0$) and the future projects ($y = y_1,y_2$) are given below:

$$\boldsymbol{C} = \begin{array}{c|cc} & d_1 & d_2 \\ p_1 & 6 & 5 \\ p_2 & 3 & 2 \end{array} \quad \text{and} \quad \boldsymbol{F} = \begin{array}{c|cc} & y_1 & y_2 \\ p_1 & 7 & 7 \\ p_2 & 4 & 4 \end{array}$$

In other words, there are six type 1 products in development stage one and five in development stage two. Seven new projects per year are expected for type 1 products. The number of projects forecasted per year is calculated using Equation 4 for each product type:

$$\boldsymbol{N_1} = \begin{array}{c|ccc} & y_0 & y_1 & y_2 \\ d_1 & 6 & 7 & 7 \\ d_2 & 5 & 3 & 3.5 \end{array} \quad \text{and} \quad \boldsymbol{N_2} = \begin{array}{c|ccc} & y_0 & y_1 & y_2 \\ d_1 & 3 & 4 & 4 \\ d_2 & 2 & 1.5 & 2 \end{array}$$

The $\mathbf{N_p}$ matrix answers Question A, giving the number of projects expected per development stage per year. This information is used to address Question B.

Question B: What level of FTE resources, manufacturing capacity, and financial resources are required per site per year?

FTE Calculations

The model uses the $\mathbf{N_p}$ matrix and the FTE resource assumptions to calculate the full time equivalents (FTEs) required per site. The FTEs required at site, s, are:

$$\mathbf{FTE_s} = \sum_{p=1}^{P} \boldsymbol{N_p} \times \boldsymbol{FTE_{p,s}} + \sum_{p=1}^{P} \sum_{c=1}^{C} \boldsymbol{x_{p,c,s}} \left(\boldsymbol{N_p} \times \boldsymbol{FTE\_flex_{p,c}}\right) \qquad \text{(Equation 5)}$$

Equation 5 calculates a matrix (g x y) describing the number of FTEs required in each of the g functional groups for each of the y years. The first term considers the fixed site activities always performed at site, s. The project matrices $\mathbf{N_p}$ are multiplied by the respective fixed site activity matrix, $\mathbf{FTE_{p,s}}$, detailing the FTE support for project type, p, at site, s. Since there are multiple product types, the matrices for all products (p=1 to p=P) are added together. Similarly, the flexible site activity matrices are multiplied by the $\mathbf{N_p}$ matrix for each product, and the site allocation factor $x_{p,c,s}$ is applied. Again, all of the flexible site activities for all product types are added together for site, s. Adding the FTE requirements for fixed site activities and flexible site activities gives the grand total of FTEs required at site, s, for each of the functional groups, g, in upcoming years, y.

Manufacturing Campaign Calculations

Calculated in a similar fashion, the forecasted manufacturing campaigns are:

$$\mathbf{MC_s} = \sum_{p=1}^{P} \boldsymbol{N_p} \times \boldsymbol{MC_{p,s}} + \sum_{p=1}^{P} \sum_{c=1}^{C} \boldsymbol{x_{p,c,s}} \, (\boldsymbol{N_p} \times \boldsymbol{MC\_flex_{p,c}}) \qquad \text{(Equation 6)}$$

This is a matrix (g x y) describing the number of manufacturing campaigns required in each of the g functional groups for each of the y years. The project matrices $\mathbf{N_p}$ are multiplied by the respective fixed site manufacturing campaign matrix, $\mathbf{MC_{p,s}}$, specifying the manufacturing campaigns required for project type, p, at site, s. Then the matrices for all products (p=1 to p=P) are added together. Similarly, the flexible site manufacturing campaign matrices are multiplied by the $\mathbf{N_p}$ matrix for each product, and the site allocation factor $x_{p,c,s}$ is applied. Again, all of the flexible site manufacturing campaigns for all product types are added together for site, s. Adding the fixed site manufacturing campaigns and flexible site manufacturing campaigns gives the manufacturing campaign grand total required at site, s, for each functional group, g, in upcoming years, y.

Financial Calculations

The materials and services required are:

$$\mathbf{MS_s} = \sum_{p=1}^{P} \boldsymbol{N_p} \times \boldsymbol{MS_{p,s}} + \sum_{p=1}^{P} \sum_{c=1}^{C} \boldsymbol{x_{p,c,s}} \, (\boldsymbol{N_p} \times \boldsymbol{MS\_flex_{p,c}}) \qquad \text{(Equation 7)}$$

Equation 7 calculates a matrix (g x y) describing the total materials and services costs required in each functional group, g, for each year, y. In the same way as the FTE and manufacturing campaign calculations, the project matrices $\mathbf{N_p}$ are multiplied by the respective fixed site materials and services matrices, $\mathbf{MS_{p,s}}$, specifying the materials and service costs expected for project type, p, at site, s. The matrices for all products (p=1 to p=P) are added together. Again, the flexible site materials and services matrices are multiplied by the $\mathbf{N_p}$ matrix for each product, and the site allocation factor $x_{p,c,s}$ is applied. Again, all of the flexible site materials and services for all product types are added together for site, s. Adding the fixed site and flexible site materials and services gives the grand total materials and services anticipated for site, s, for each functional group, g, in upcoming years, y.

Additionally, one can calculate the total costs by considering FTEs, manufacturing campaigns, and materials and services. To do this, some additional information is needed: $\langle FTE\_COST_s \rangle$ = a vector (with g terms) listing the average cost per FTE in each of the g functional groups at site s, MC_WKS = average weeks per manufacturing campaign, and MC_COST_s = cost per week of manufacturing time at site s. The overall site cost, $\langle OSC_s \rangle$, at site s is calculated using:

$$\langle OSC_s \rangle = \mathbf{FTE_s} * \langle FTE\_COST_s \rangle + \mathbf{MC_s} * MC\_WKS * MC\_COST + \mathbf{MS_s}$$

This $\langle OSC_s \rangle$ vector (with y terms) describes the overall site costs for the next y years.

Example (continued)

Continuing the previous example, assume that there are two sites ($s = \{s_1, s_2\}$) and two functional groups per site ($g = \{g_1, g_2\}$). The site specific work is defined in $\boldsymbol{FTE_{p,s}}$ and the work that could be performed at multiple sites is defined by $\boldsymbol{FTE\_flex_{p,c}}$. For both products, site 1 does 10% of the center 1 work and site 2 does 90% of center 1 work ($x_{p,1,1} = 0.10$ and $x_{p,1,2} = 0.90$). The following FTE information is also given.

Site 1 resources for p_1 and p_2: $FTE_{1,1} = \begin{array}{c|cc} & d_1 & d_2 \\ g_1 & 10 & 4 \\ g_2 & 9 & 8 \end{array}$ and $FTE_{2,1} = \begin{array}{c|cc} & d_1 & d_2 \\ g_1 & 15 & 8 \\ g_2 & 8 & 7 \end{array}$

Site 2 resources for p_1 and p_2: $FTE_{1,2} = \begin{array}{c|cc} & d_1 & d_2 \\ g_1 & 0 & 2 \\ g_2 & 2 & 15 \end{array}$ and $FTE_{2,2} = \begin{array}{c|cc} & d_1 & d_2 \\ g_1 & 0 & 5 \\ g_2 & 5 & 20 \end{array}$

Center 1 resources for p_1 and p_2:

$FTE\_flex_{1,1} = \begin{array}{c|cc} & d_1 & d_2 \\ g_1 & 0 & 1 \\ g_2 & 2 & 0 \end{array}$ and $FTE\_flex_{2,1} = \begin{array}{c|cc} & d_1 & d_2 \\ g_1 & 1 & 3 \\ g_2 & 4 & 2 \end{array}$

Using the N_1 and N_2 calculated in the previous example:

$N_1 = \begin{array}{c|ccc} & y_0 & y_1 & y_2 \\ d_1 & 6 & 7 & 7 \\ d_2 & 5 & 3 & 3.5 \end{array}$ and $N_2 = \begin{array}{c|ccc} & y_0 & y_1 & y_2 \\ d_1 & 3 & 4 & 4 \\ d_2 & 2 & 1.5 & 2 \end{array}$

We can calculate the total FTEs required at sites 1 and 2 in the next two years. The first terms are shown below to illustrate the calculation method. The first terms of FTE_1 (number of g_1 and g_2 FTEs required in this year, y_0) are shown below:

$(FTE_1)_{g_1,y_0} = (6*10+5*4)+(3*15+2*8)+0.1*(6*0+5*1)+0.1*(3*1+2*3) = 142$

$(FTE_1)_{g_2,y_0} = (6*9+5*8)+(3*8+2*7)+0.1*(6*2+5*0)+0.1*(3*4+2*2) = 135$

The FTE forecast results for sites 1 and 2 for this example are:

$FTE_1 = \begin{array}{c|ccc} & y_0 & y_1 & y_2 \\ g_1 & 142 & 155 & 161 \\ g_2 & 135 & 133 & 140 \end{array}$ and $FTE_2 = \begin{array}{c|ccc} & y_0 & y_1 & y_2 \\ g_1 & 33 & 24 & 29 \\ g_2 & 167 & 139 & 157 \end{array}$

The number of clinical campaigns as well as the materials and services costs can be calculated in a similar fashion.

The model-oriented notation is summarized below. This information is calculated by the model. Individual variables are in *italics*, vectors are contained in <brackets>, and matrices are in **bold**.

$<F_{cp,d}>$ = Forecast vector summarizing the number of $c_{p,d}$ projects that will graduate into the remaining development stages, d. The vector contains d terms, one for each development stage.

$<F_{fy,p}>$ = Forecast vector summarizing the number of $f_{y,p}$ projects that will graduate into the remaining development stages, d. The vector contains d terms, one for each development stage.

T = Time matrix for each current project, $c_{p,d}$, and future project, $f_{y,p}$, that allocates the respective vector into the appropriate year. If there are no time delays, and each development year takes one year, then **T** is an identity matrix. If there are time delays and $T_d \neq 1$, then **T** is calculated according to Appendix C.

$\mathbf{N_p}$ = Number of projects expected per year, y, in development stage, d. There is one (d x y) matrix for each product type, p.

$\mathbf{FTE_s}$ = Total FTEs forecasted for each functional group, g, needed at site, s, for each year, y. There is one (g x y) matrix for each product type, p.

$\mathbf{MC_s}$ = Total manufacturing campaigns forecasted for each functional group, g, needed at site, s, for each year, y. There is one (g x y) matrix for each product type, p.

$\mathbf{MS_s}$ = Total materials and services forecasted for each functional group, g, needed at site, s, for each year, y. There is one (g x y) matrix for each product type, p.

$<OSC_s>$ = Overall site costs for site, s, considering FTE, manufacturing campaigns, and materials and service costs. This is a vector consisting of y terms, forecasting the cost per year for the upcoming y years.

3.4 Novartis Model

The model designed for Novartis has seven product types (p = {cell culture 2-step, cell culture 3-step, microbial, biosimilar cell culture, biosimilar microbial, vaccines, external}) and five functional groups ($g = \{g_A, g_B, g_C, g_D, g_E\}$). The model is usually used to project the next seven years ($y = \{y_0, y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$). Many of the model input variables listed in Section 3.2 carry a degree of uncertainty. In order to model this uncertainty, Crystal Ball Monte Carlo simulation software is used to simulate the variability and test the sensitivity of the resource forecasts. Crystal Ball, an Oracle software program, simulates many scenarios by populating cells with randomly generated numbers according to probability distributions defined by the user. The software runs many trials and records the outputs. This allows the user to observe the variability of the model projections. See references for more information on Monte Carlo simulation and the Crystal Ball software (Mun, 2006 and Charnes, 2007).

First, the user must define probability distributions for the future projects ($f_{y,p}$), development stage duration (T_d) and probabilities of success ($S_{p,d}$). For instance, the user can specify that the time to complete development stage 1 has a BetaPERT probability distribution with a maximum of 1.5 years and a minimum of 1.0 years. The user can then run the model for a set number of simulations (1000 trials were used for this model) and determine the resource variation. Figure 3 shows the results of a simulation run involving 1000 trials and graphing the "high" and "low" estimates. In this case the 10% percentile and 90% percentile outcomes are graphed; however, that can be adjusted by the user.

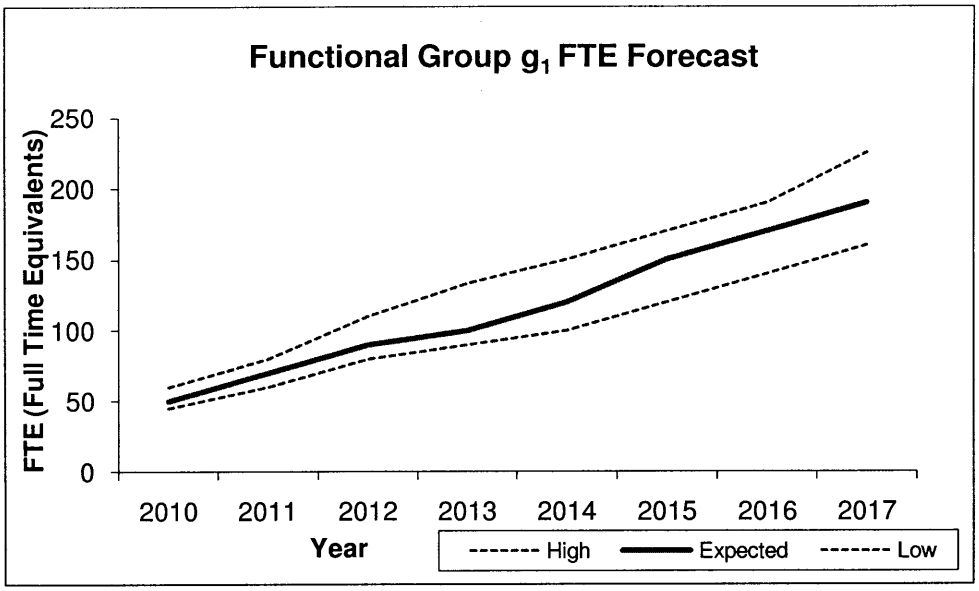

Figure 3 - Example of Model Output, FTE Forecast for functional group, g_1.

Similar graphs are made for the manufacturing campaign and materials and services forecasts. In summary, this model calculates the number of projects expected over the next seven years and the resources required to support future projects. Through simulation, the model also projects the potential variability of those resource needs. This information is useful for making long term resource and capacity decisions.

CHAPTER 4 – Model Empirical Evaluation

This chapter uses three evaluations to demonstrate that the drug development process can be described by a flexible and scalable model, and that the model produces reliable output. The first part (Section 4.1) empirically evaluates the model's ability to forecast the correct number of projects ($\mathbf{N_p}$). The second part (Section 4.2) empirically assesses the model's ability to forecast resources required ($\mathbf{FTE_s}$, $\mathbf{MC_s}$). The third part (Section 4.3) empirically examines how similar the model predictions are with a currently accepted Novartis model. The purpose of these empirical evaluations is to compare the model to actual data and currently accepted practices.

4.1 Historical Pipeline Comparison

If the model is used with today's information, one would need to wait up to eight years to know if the model is accurate. Alternatively, one can take data from the past, and run the model at a certain point in history. The model projection can be compared against what actually happened to assess the model accuracy.

This empirical assessment uses historical data, which are available for the biologics industry as a whole to demonstrate how the model can be used to predict the number of projects each year ($\mathbf{N_p}$), and how assumptions can be adjusted in order to best model the actual data. The following graph summarizes a study performed by Pharmaprojects (Parexel, 2009).

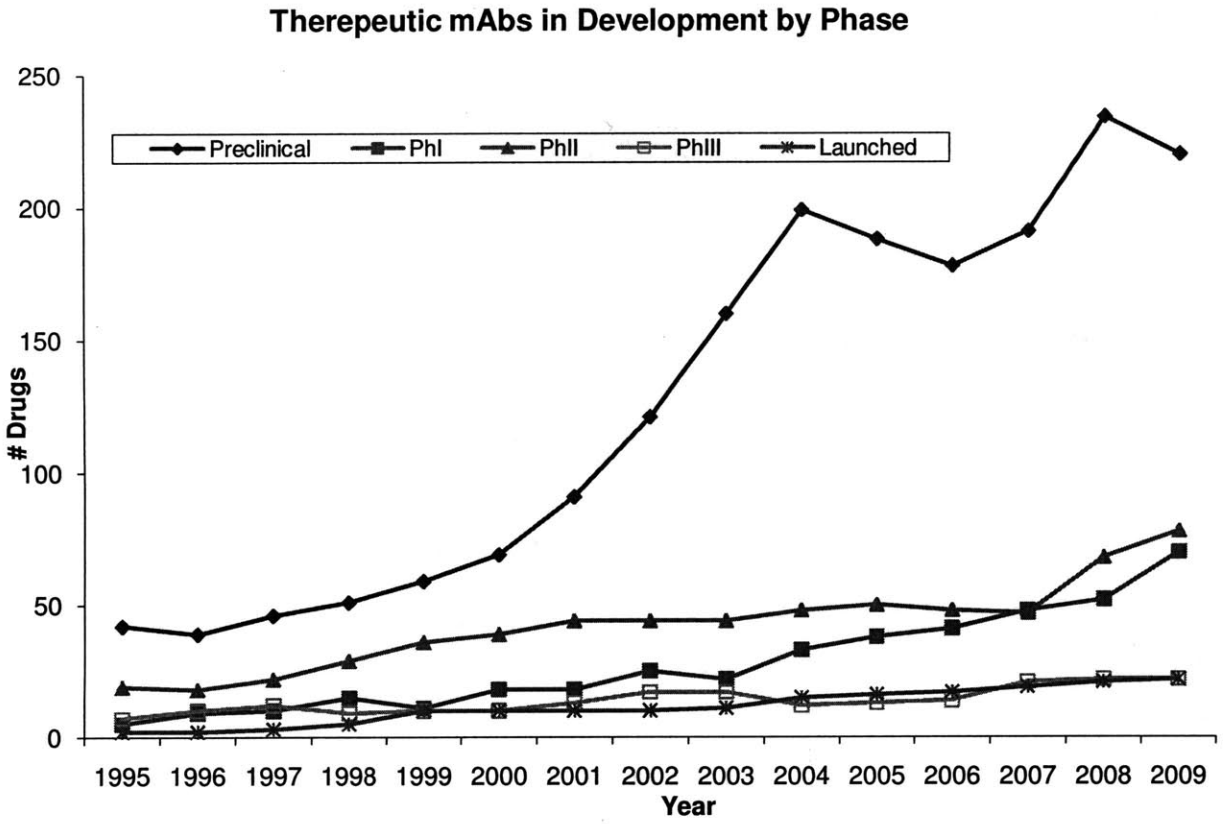

Figure 4 - Therapeutic mAbs in Development by Phase, 1995-2009

These data show the actual number of monoclonal antibodies in each development phase from 1995 to 2009. The square markers show the number of preclinical projects per year, the circles show the number of Phase I projects, and so forth according the legend. In this empirical evaluation, the 1995 data and the growth of the preclinical phase represent incoming projects. The model is used to forecast the number of projects expected for the following seven years; and this forecast is then compared to the number of projects that actually occurred during those years.

1995 Industry Comparison

Figure 5 summarizes the probabilities of success and phase duration assumptions. A 42% probability of success between Phases I and II means that at the end of Phase I, 42% of the projects will proceed to Phase II and 58% will be terminated.

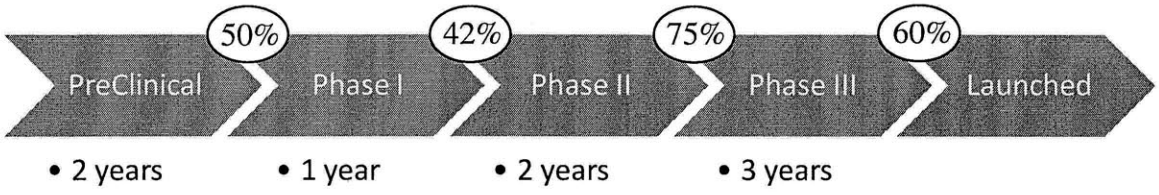

Figure 5 - Probability of Success and Phase Duration Assumptions for 1995 Industry Comparison

The incoming projects are given by the number of preclinical projects specified by the Parexel data. Running the model with the assumptions above and comparing that information with the actual data yields the results shown in Figures 6-9.

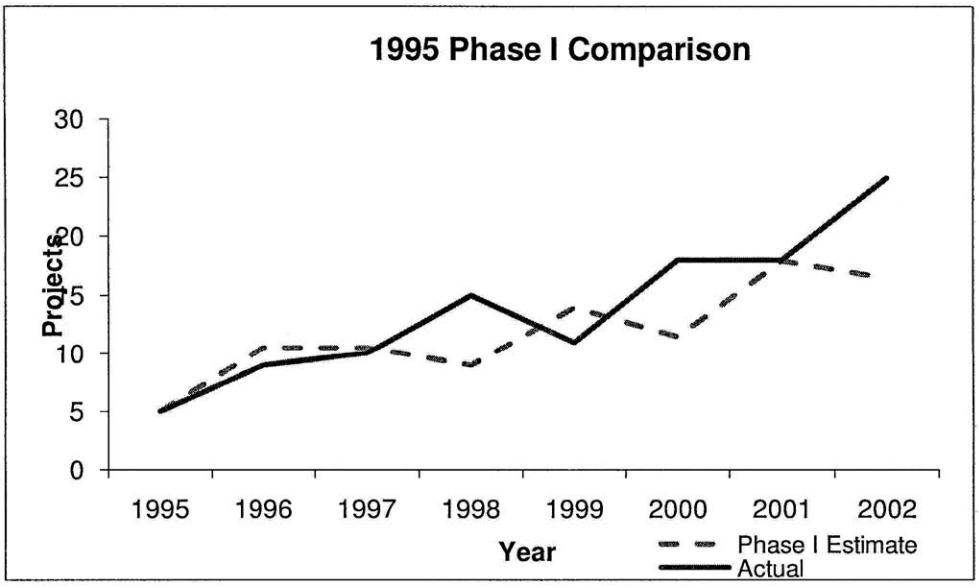

Figure 6 - Phase I Model Forecast and Actual Number of Projects

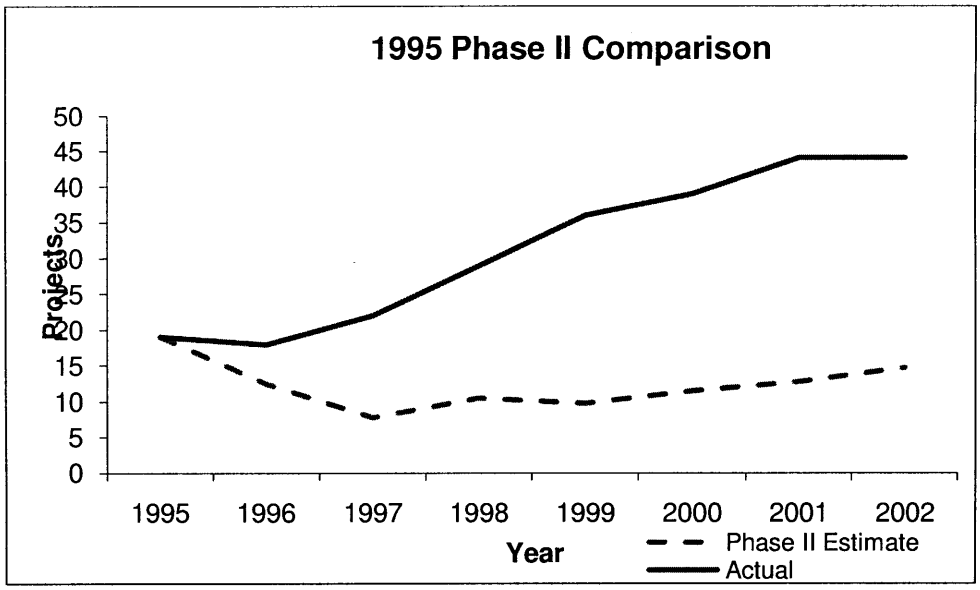

Figure 7 - Phase II Model Forecast and Actual Number of Projects

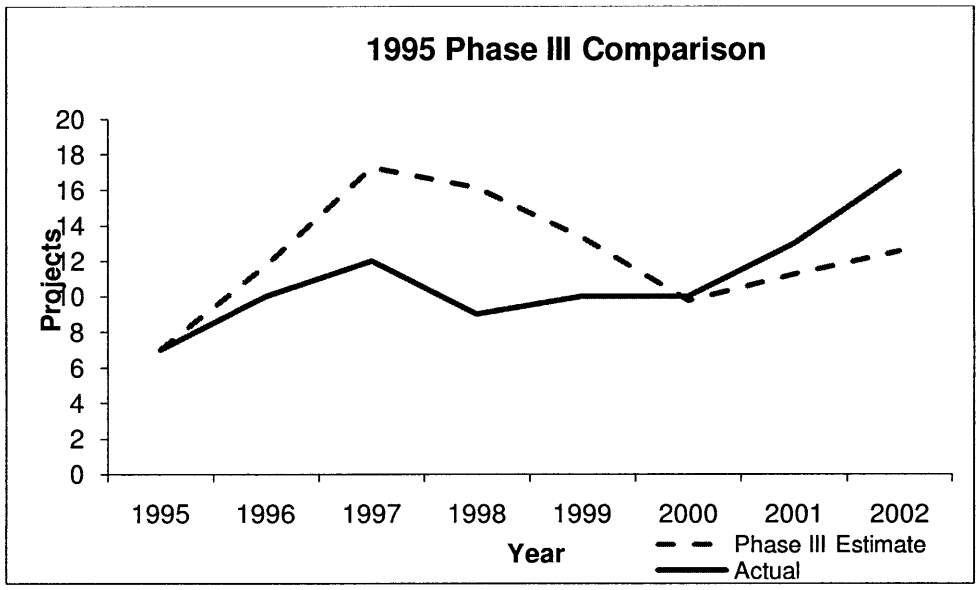

Figure 8 - Phase III Model Forecast and Actual Number of Projects

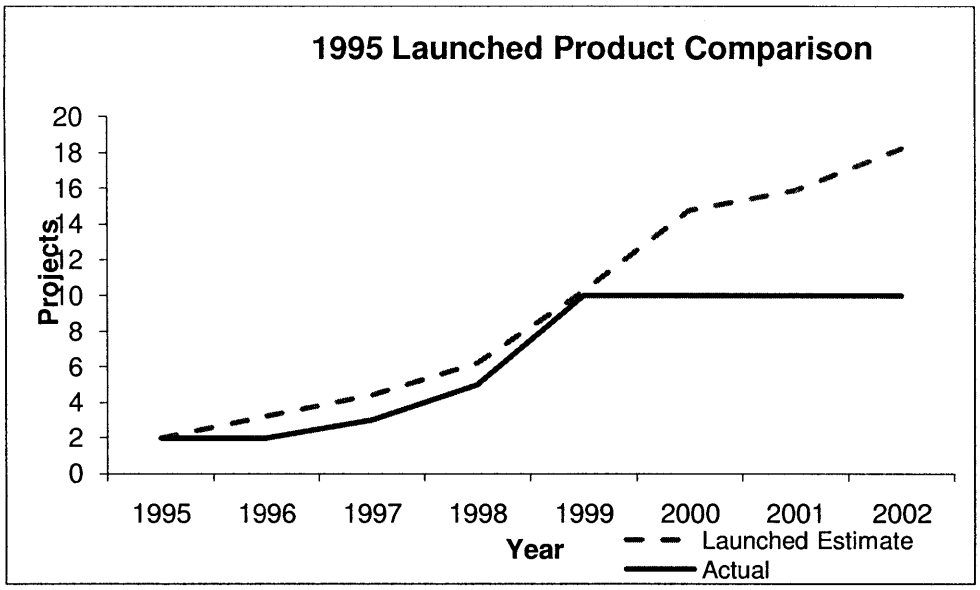

Figure 9 - Launched Product Forecast and Actual Number of Projects

The model predicted the expected number of projects in Phase I within +/- 7 projects (or +/- 35%) for the 1995 forecast. This prediction is reasonably accurate for the purposes of long range planning. However, the Phase II forecast was consistently lower than the actual number of projects in Phase II. This trend is explored in detail below. The Phase III comparison estimates the number of projects within +/- 7 projects (estimating high by up to 75% in 1998). Lastly, the Launched Product comparison matches very nicely until after year 2000, where the actual data flat-lines. There may have been an external factor involved during this period, as it appears to be uncharacteristic considering the pipeline and past number of projects.

1995, 2000, and 2005 Industry Comparison

To test if the model consistently over or underestimates projects, and to see if these model assumptions hold true over time, 2000 and 2005 were also evaluated in the same way as the 1995 case above. The percent difference was calculated between the model estimates and the actual values, where percent difference = (model value - actual value)/actual value. So if the percent difference is negative, the model underestimates the actual number of projects and if the percent difference is positive then the model overestimates the actual number of projects. Figures 10-13 below show the evaluation results.

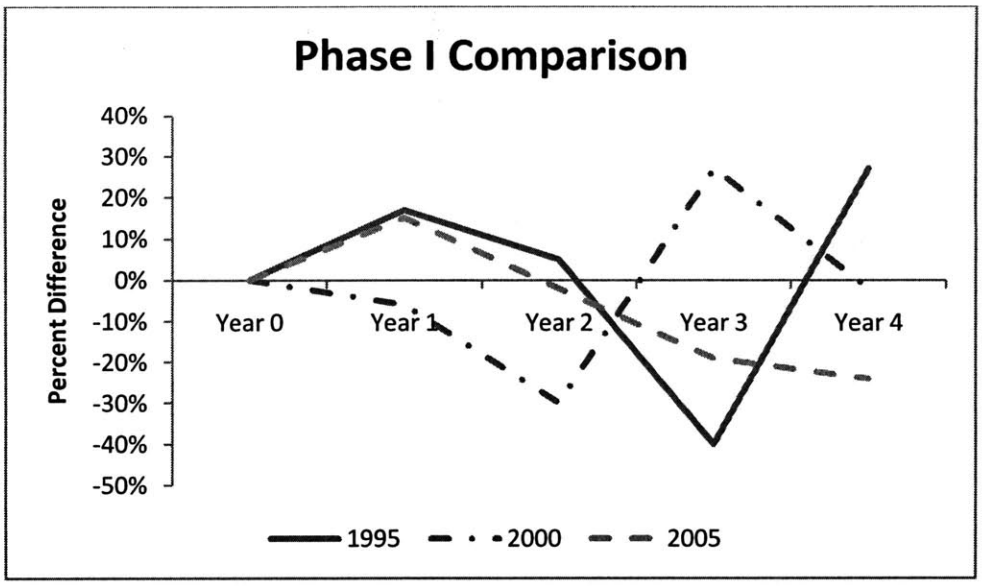

Figure 10 - Phase I Comparison between Forecast and Actual (in Percent Difference) for 1995, 2000, and 2005.

The model estimated the Phase I projects reasonably well, within +/- 40%. There is no overestimating or underestimating trend.

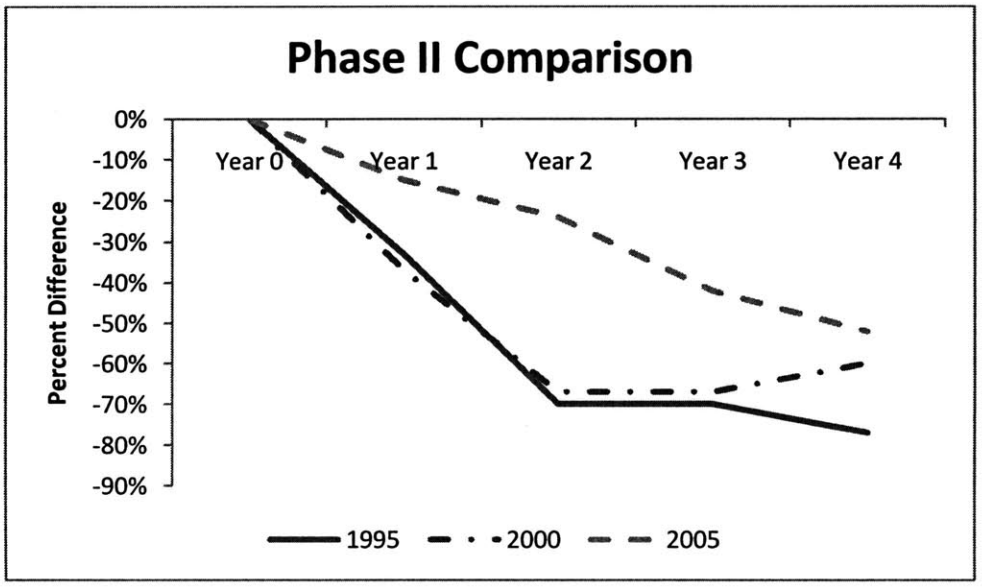

Figure 11 - Phase II Comparison between Forecast and Actual (in Percent Difference) for 1995, 2000, and 2005.

The model consistently underestimates the number of projects in Phase II. Possible reasons for this trend:

- The time duration for Phase II could be too low, and projects could actually be spending more than two years in this phase.
- The probability of success between Phase I and Phase II could be lower than the model assumes, and less than 42% of the Phase I projects could be moving into Phase II.

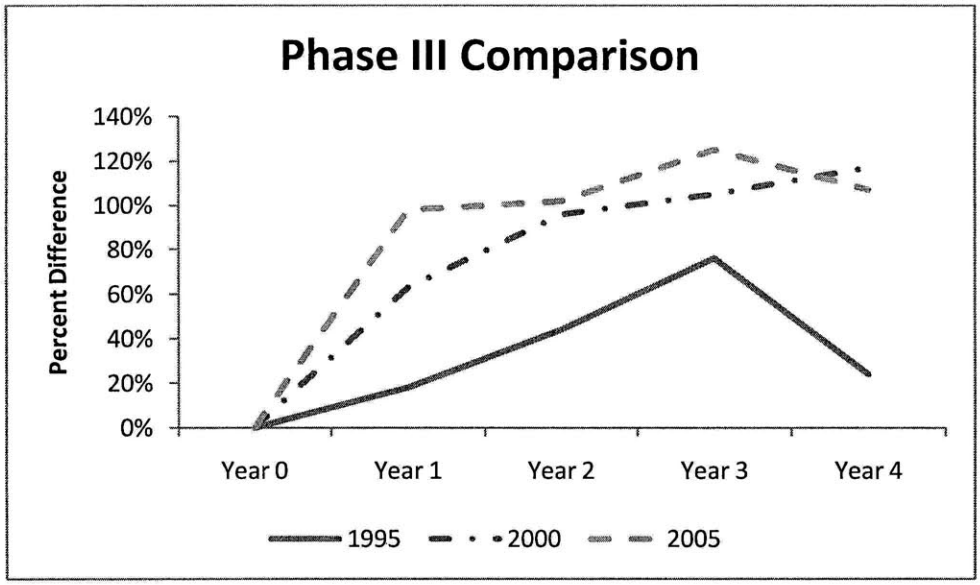

Figure 12 - Phase III Comparison between Forecast and Actual (in Percent Difference) for 1995, 2000, and 2005.

In all cases, the model overestimates the number of projects in Phase III, sometimes by as much as 130%. Possible reasons for this trend:

- The Phase II time duration assumption could be too short, and the model could be moving projects into Phase III prematurely.
- The probability of success from Phase II to Phase III used in the model could be too high, or less than 75% of Phase II projects move into Phase III.
- The model may be overestimating the Phase III duration time.

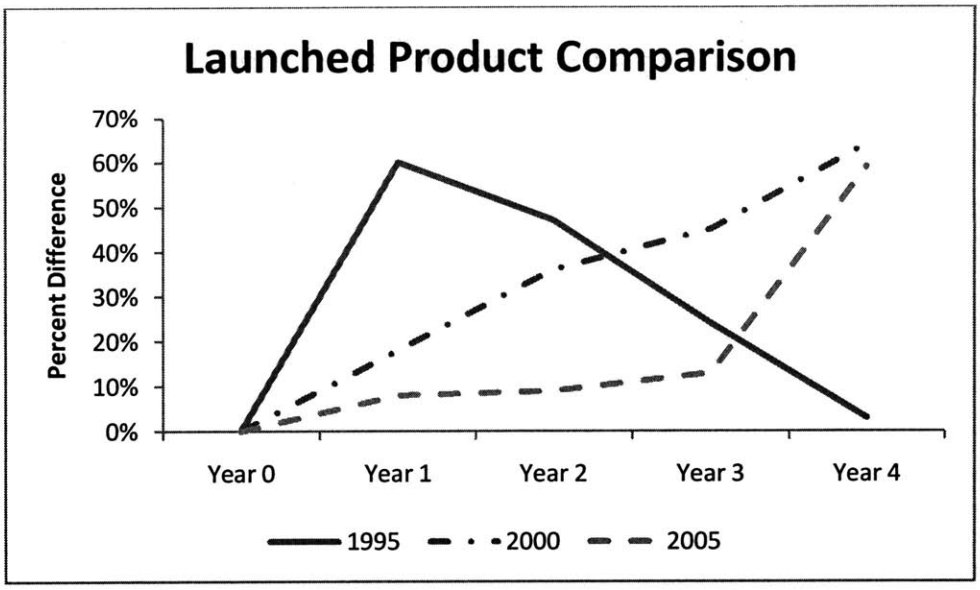

Figure 13 - Launched Product Comparison between Forecast and Actual (in Percent Difference) for 1995, 2000, and 2005.

The model also tends to overestimate the number of launched products, in some cases by as much as 60%. It is possible that the probability of success is actually too high in this case or that projects are actually spending more time in previous phases and the model is running at an accelerated pace.

To empirically assess the accuracy of the probability of success assumptions, the overall number of projects predicted by the model is compared with the actual number of projects. Results are shown in Figure 14.

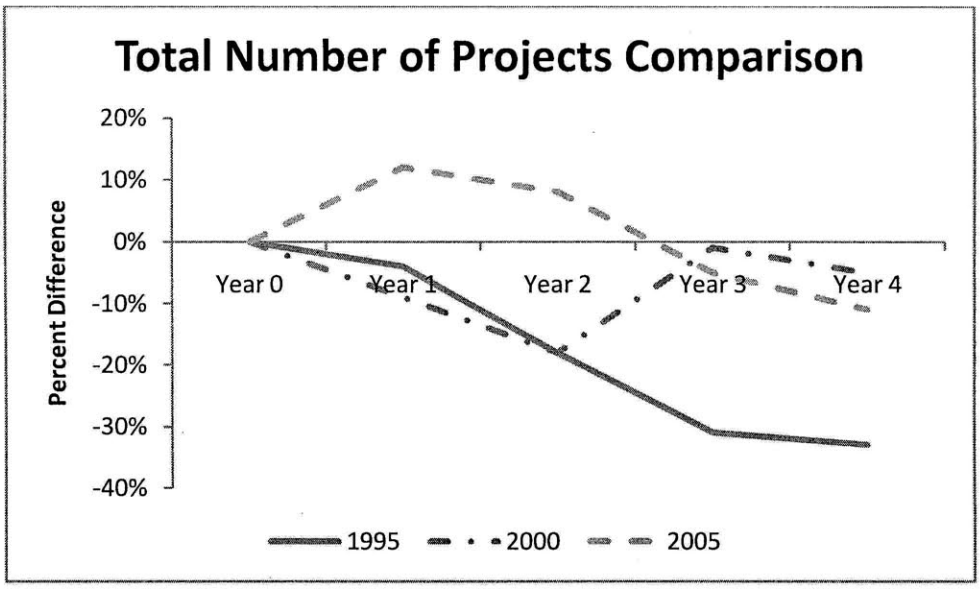

Figure 14 - Total Project Number Comparison between Forecast and Actual (in Percent Difference) for 1995, 2000, and 2005.

It is interesting to note that even though the model tends to overestimate the number of projects in Phases II, III and Launched products, the total number of projects is reasonably estimated by the model. This suggests that the model is retaining a good number of projects, suggesting that the phase durations are in need of adjustment, not the probabilities of success.

1995, 2000, and 2005 Industry Comparison with Adjusted Assumptions

According to an interview with a clinical trial coordinator, Phase II and Phase III duration can be most variable. Phase III is typically a continuation of Phase II and the transition point between the two phases can be somewhat arbitrary. Understanding this, the time for Phase II was adjusted to four years and Phase III was reduced to two years. Figure 15 illustrates the new model assumptions.

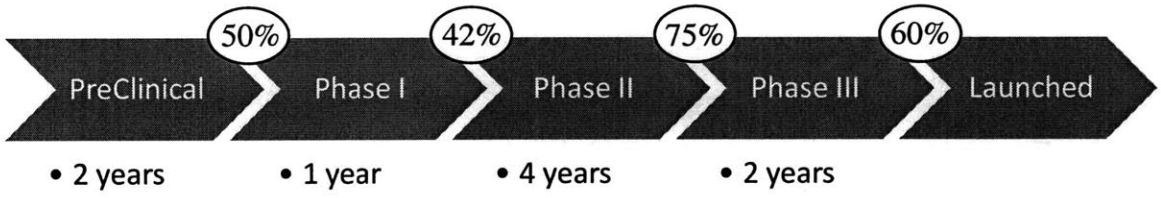

Figure 15 - Adjusted Assumptions for Industry Comparison

Again, the percent difference between the model and actual data can be compared for the 1995, 2000, and 2005 forecasts. The results are shown in Figures 16-19. Since neither the probability of success nor the phase duration of Phase I changed, the Phase I model forecasts did not change. They are still represented in Figure 10.

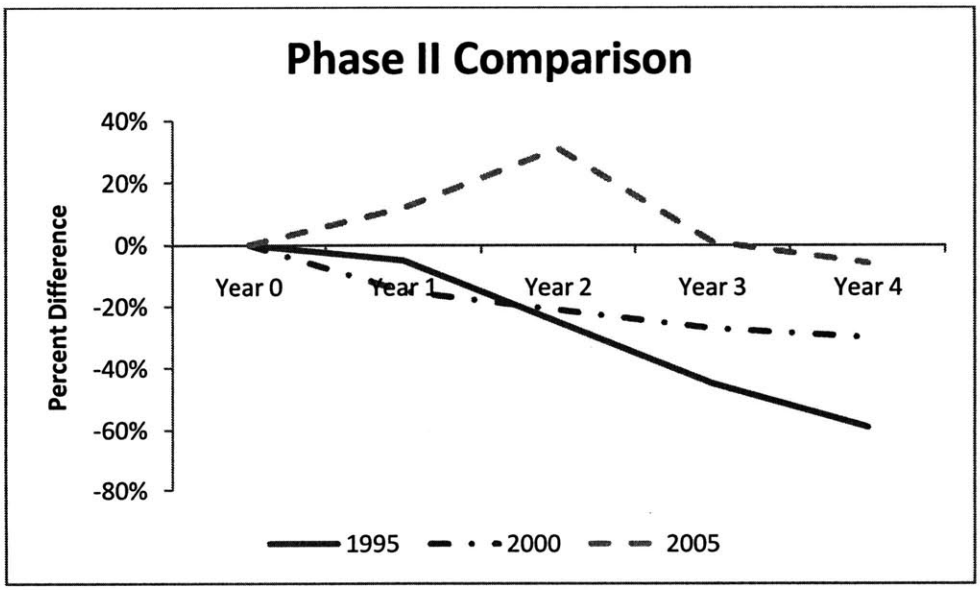

Figure 16 - Phase II Comparison between Forecast with Adjusted Assumptions and Actual (in Percent Difference) for 1995, 2000, and 2005.

The Phase II forecasts were much more accurate using the new assumptions; it appears that four years in Phase II may better represent the actual case. The 1995 forecast was estimating about 50% low for four years out into the future, but that is still significantly better than in the initial case. Additionally, the model predicts 2005 quite well, overestimating by 30% at the most.

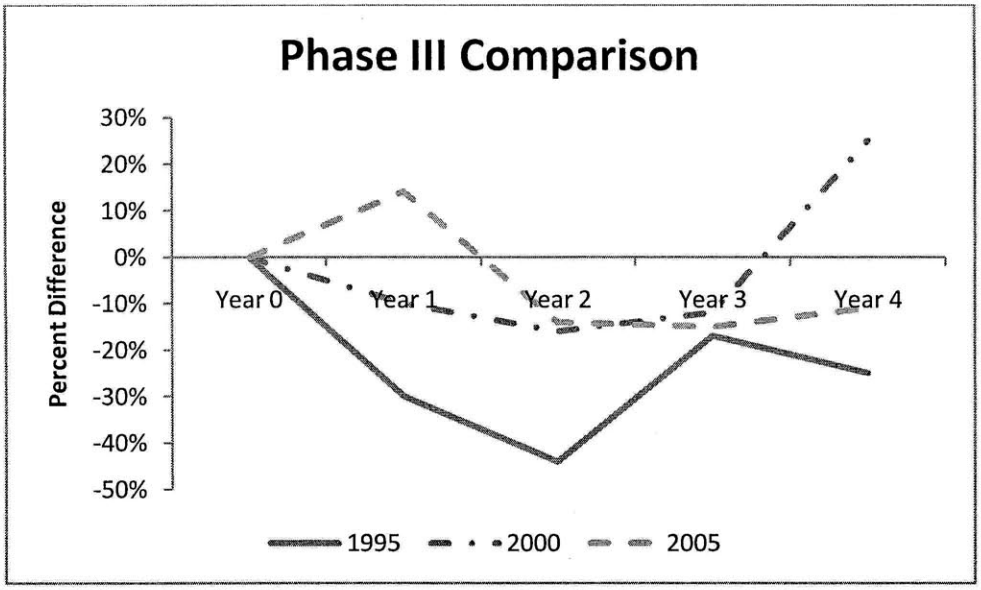

Figure 17 - Phase III Comparison between Forecast with Adjusted Assumptions and Actual (in Percent Difference) for 1995, 2000, and 2005.

The Phase III comparison in this case is also more accurate than the first evaluation. Using the original assumptions, the model was overestimating by as much as 130%; the model now estimates within +/- 40%.

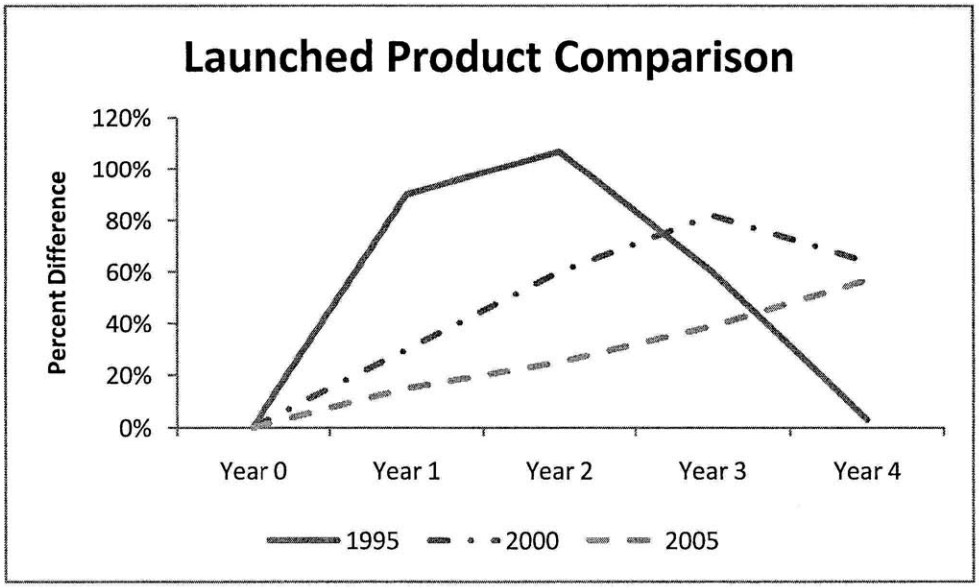

Figure 18 - Launched Product Comparison between Forecast with Adjusted Assumptions and Actual (in Percent Difference) for 1995, 2000, and 2005.

The model continues to overestimate the numbers of launched products. It is possible that the probability of success is too high at this step.

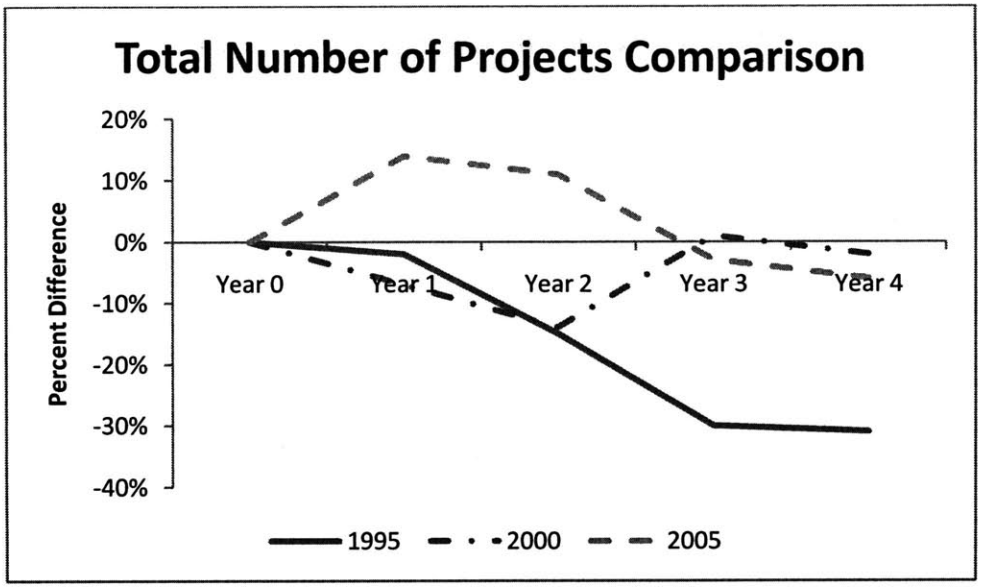

Figure 19 - Total Project Comparison between Forecast with Adjusted Assumptions and Actual (in Percent Difference) for 1995, 2000, and 2005.

Not surprisingly, the total number of products did not change significantly, only increasing slightly due to the retention of projects in Phase II for an additional two years. The percent difference seems reasonable, especially for the 2000 and 2005 forecasts.

Conclusions

As demonstrated by these empirical evaluations, the drug development process can be forecasted using a flexible model. This chapter demonstrates that a model can be used to study historical data, and assumptions can be adapted such that the model represents these data. After all adjustments, this model predicts the number of cell culture projects in the pharmaceutical industry within +/- 40%. This error is rather large for a number of reasons.

First, errors may reflect the fact that assumptions are based on data from many companies for many different medical indications. Probabilities of success can depend on companies, and phase duration can depend on medical indication. Aggressive companies with an appetite for risk may push more products through clinical trials, while a conservative company that is risk adverse or capital constrained may choose a limited number of products to move to clinical trials.

It is possible that the industry is better at antibody design today than it was in 1995, so probabilities of success may improve over time. Phase duration is dependent on drug indication. A cancer drug may have an accelerated Phase I trial to treat terminal cancer patients quickly, while, a hypertension drug may go through extended Phase I, Phase II and Phase III trials. Therefore, there are many reasons as to why it is difficult to assign phase durations and probabilities of success to represent a wide range of companies and drug indications.

Nevertheless, for the purpose of long range planning, general assumptions must be used to describe a diverse pipeline of projects. It is impossible to know today the specific drug products that will be successful, and the indications for which they will be used. However, if the project pipeline is large enough, average values for probabilities of success and phase duration can be helpful in determining resource needs. It is important that a company that uses a model like the one described in this thesis also examines the phase duration and probabilities of success using its own internal data as well, adjusting assumptions to best estimate the future pipeline as demonstrated in this section.

4.2 Historical Resource Comparison

The purpose of this empirical exercise is to demonstrate that the model can provide reliable FTE and manufacturing forecasts. FTE and manufacturing capacity information is held confidential by pharmaceutical manufacturers. In this case, confidential Novartis data was used for this evaluation, and the results are shown in terms of percent difference per year, nominal difference, or percent capacity utilized, and does not reveal confidential information.

2007 FTE Comparison

Novartis has presentations from 2007, outlining the "current projects" at that point in time. Using that information, the model can generate an FTE forecast beginning in 2007. The model output can then be compared against the actual number of FTEs hired in 2007, 2008, and 2009. The percent difference is calculated as percent difference = (model value - actual value)/actual value. Figure 20 summarizes the results of the comparison, and shows that the model estimated the actual number of total FTEs required within 7% for each year (2007, 2008, and 2009).

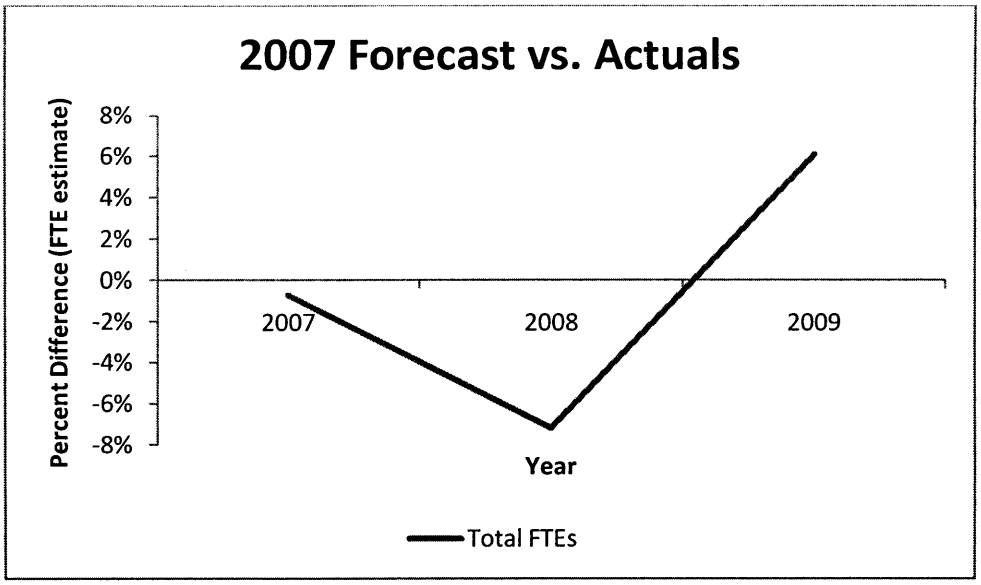

Figure 20 - 2007 Model Forecast comparison actual FTEs for 2007, 2008, and 2009, measured in percent difference.

The model underestimated 2007 FTEs by 1%, underestimated the 2008 FTEs by 7%, and overestimated the 2009 FTEs by 7%. These estimates are within acceptable error for long range planning purposes, where +/- 10% can be considered good. If the model projects that 107 FTEs are required and the actual number hired is 100 FTEs, then this margin of error is very reasonable for long range planning purposes where an approximate estimate is more important than precision.

Since the model also breaks down the resource requirements per functional group, model forecast per group can be compared with the actual number of FTEs hired per group over a three year time span. The results are shown in Figure 21.

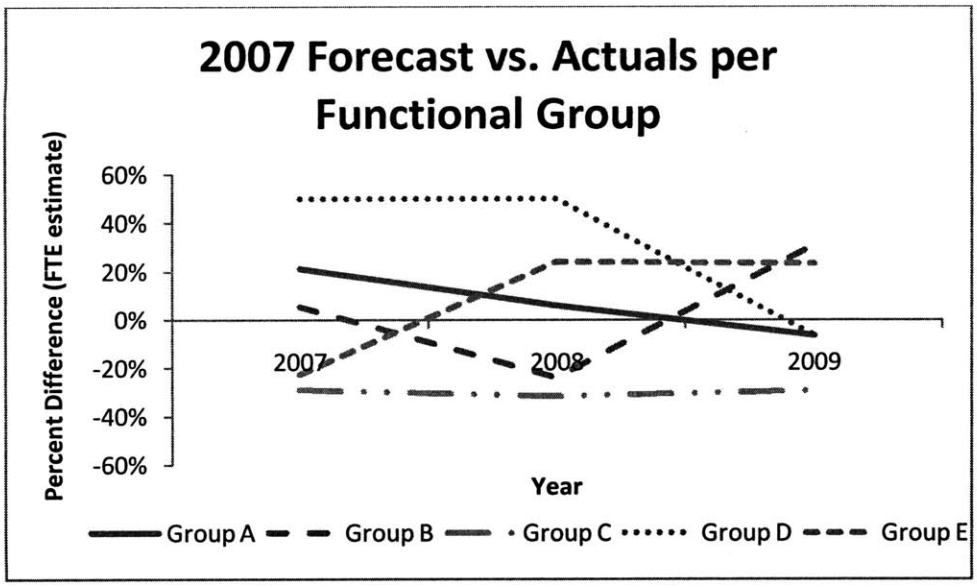

Figure 21 - 2007 Model Forecast comparison against actual FTEs per functional group for 2007, 2008, and 2009, measured in percent difference.

The model does not forecast the functional group FTEs as well as it forecasts the total FTEs. Group D was overestimated by as much as 50% in 2007 and 2008. However, since this is the smallest group in the department, this error was dwarfed by the overall total FTE approximation. Additionally, the model consistently underestimates the Group C FTE requirements. This encourages Novartis Biologics to re-evaluate the FTEs per project assumptions for Group C; some tasks may be missing or have low work requirement estimates. The model approximates within 20% of the actual required FTEs for Groups A, B and E for 2007, 2008, and 2009.

This exercise exposes a number of additional challenges regarding model forecasting. In the period of these three years, the Novartis Biologics development organization has had a number of reorganizations. The model is made to forecast the headcount needed for the current organizational structure. However, when using the model for 2007, it was realized that some of the functional groups did not exist in the same way that they do today. For the purpose of this exercise, FTEs were classified into the appropriate bucket as the organization is structured today. Even so, it is recognized that the organization will continue to change. It is not possible to predict exactly how the organization will change. So it is most useful to model the current organizational structure, understanding that the model must be updated as the organization changes.

This exercise also showed that future advancements that could change resource and capacity utilization are not considered in the model. In 2007, there was only one type of product development method, mAb 3-step. In recent years, another development method was implemented, mAb 2-step. This will be discussed in further detail in Chapter 5.1. If this model were used in 2007, it would have been difficult to predict that there would be over five types of product development platforms in 2009. This is another fact that must be realized in using the model: it only considers development platforms that are currently in use. If new types of manufacturing methods or new types of products are discovered in the future, it is difficult to predict these advancements using today's information.

Despite these shortcomings and "unknowns," resource forecasting remains a valuable exercise. Even though the organizational structure or development platforms may change, decisions to expand capacity, hire FTEs, or contract work must be considered and evaluated, using the best information available.

2006 Manufacturing Campaign Comparison

A similar comparison is performed for the manufacturing capacity. Using Novartis 2006 presentations, the model is run as if the year were 2006. The numbers of manufacturing campaigns per site are forecasted and compared to the actual number of cell culture manufacturing campaigns. Novartis has two sites manufacturing Phase I/II clinical trial material, and two sites manufacturing Phase III clinical trial material. The nominal difference is calculated as (model value – actual value). The results are shown in Figures 22 and 23. A positive number means that the model overestimated the number of campaigns by this amount, and a negative number means that the model underestimated the number of campaigns by this amount.

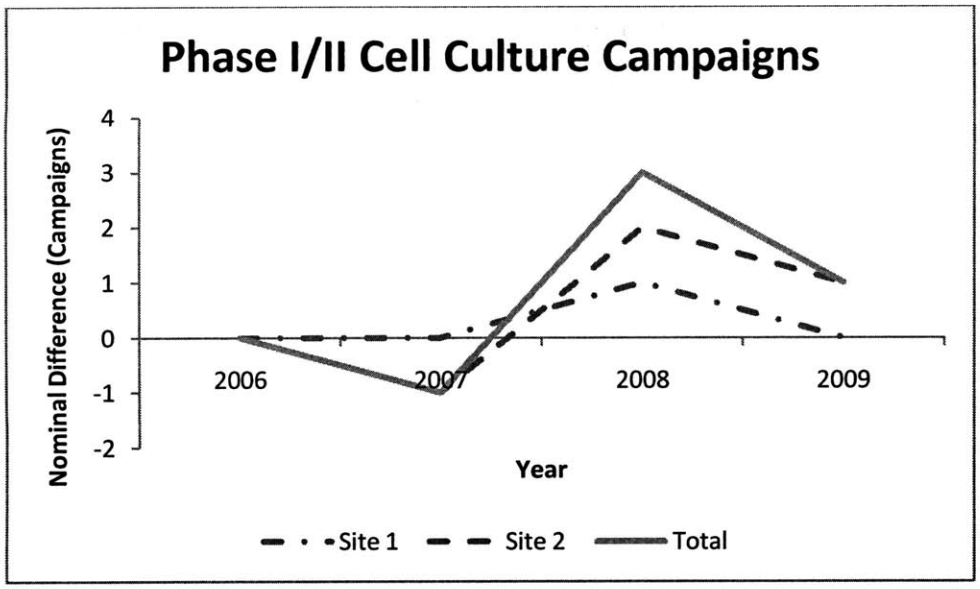

Figure 22 - Difference between model forecast and actual number of Phase I/II Cell Culture Campaigns

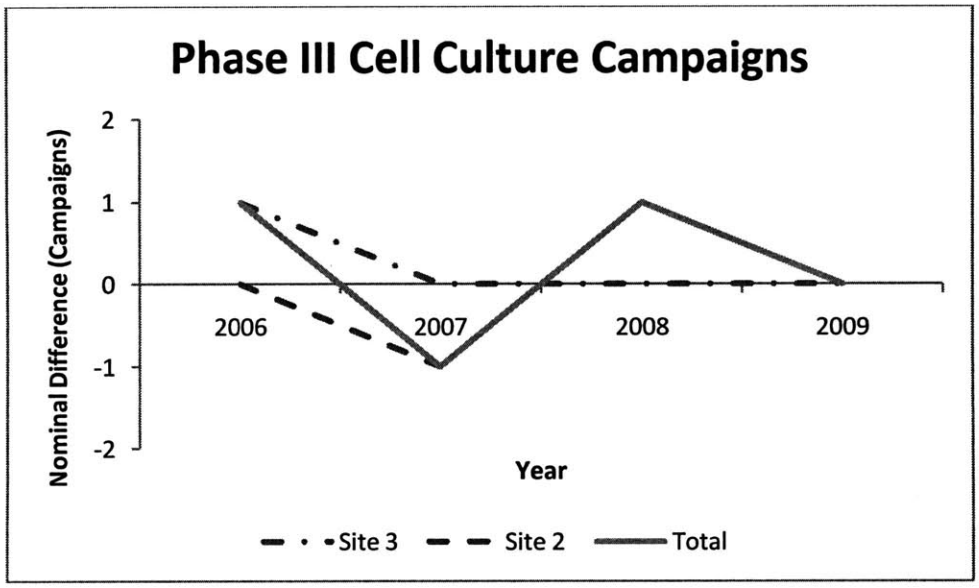

Figure 23 - Difference between model forecast and actual number of Phase III Cell Culture Campaigns

As expected, the 2006 campaign predictions are within one campaign of the actual number of campaigns. However, the model overestimates the Phase I/II cell culture campaigns by up to three additional campaigns in 2008. Additionally, Site 1 Phase I/II campaigns seem to be consistently overestimated by at least one campaign per year. For the purpose of long range

planning, +/- three manufacturing campaigns is acceptable per Novartis management. However, due to the Phase I/II trends, it would be worth investigating reasons for this overestimation. The campaigns specified in the work packages may be either too frequent or in the incorrect development stages. Since the Phase III estimates are close to the actual numbers of campaigns performed, high probabilities of success in early stages are most likely not the reason for manufacturing campaign overestimation.

Conclusion

When using historical Novartis data, the model forecasts FTE resource and manufacturing capacity requirements reasonably. Any company using a model such as this one should perform a similar assessment using actual internal data. As Novartis gains more experience in biopharmaceutical development, this evaluation should be repeated to expose assumptions that may need adjustment.

4.3 Current Capacity Forecasting Comparison

In 2007, the commercial manufacturing organization developed an independent modeling tool to estimate manufacturing capacity requirements for commercial projects. In 2009, clinical trial manufacturing data were added to the existing model. The commercial project model is on a Microsoft Access database and extracts information from Microsoft Project files. If the campaign dates for a project are known, they are entered into the Microsoft Project file; if the campaign dates are unknown, the model uses a standard timeline for each project. Both the "existing model" used by the commercial organization and the "new model" described in this thesis use the same probabilities of success. Both models were used to forecast manufacturing requirements for 2010-2018, only considering projects that start in 2010. The results are compared in terms of percent capacity utilization and are shown in Figure 24.

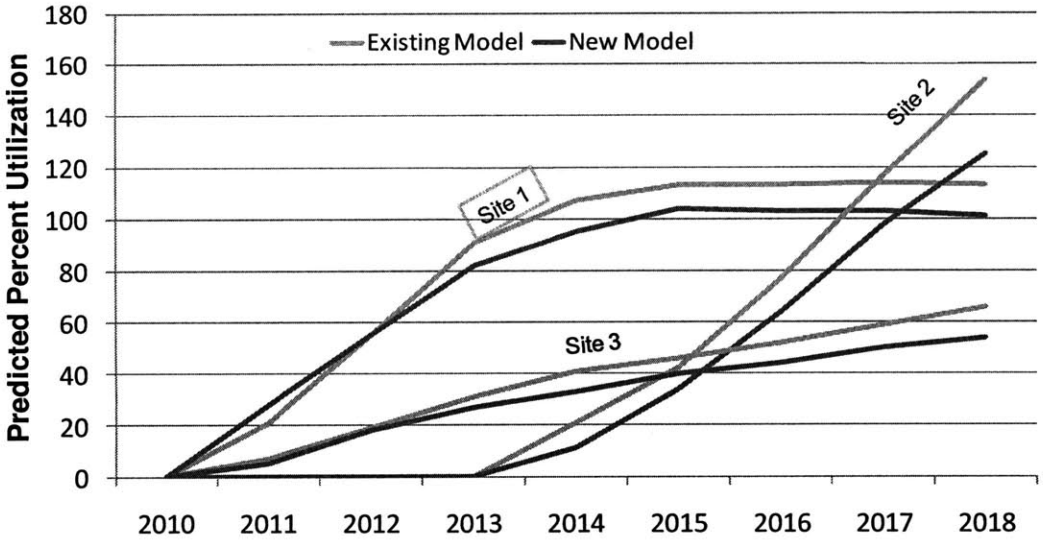

Figure 24 - Comparison of Existing Model with New Model for Manufacturing Capacity Forecasting

Both models use the same manufacturing timeline, so a close comparison is expected. The average difference between the new model and the existing model is around 13%, with the existing model usually estimating higher than the new model. This variation could be due to the "time delay" factor that is used in the new model, and not used in the existing model. Fewer manufacturing campaigns will be required as projects are pushed into the future by the delay. As delays compound, the difference between these two models will increase over time. One can see that the models are much closer in 2011 than 2018. Including delays are more realistic, and are an acceptable difference according to the Novartis model owners.

Conclusion

This empirical evaluation is an enlightening way to examine how the model compares to currently accepted practices. Figure 25 compares the two models in the evaluation. The advantage of the new model is that it is faster to run, and includes tools to estimate development headcount and financial requirements. The existing model is designed for commercial manufacturing planning, and focuses on the commercial aspects of planning.

	Existing Commercial Model	New Development Model
Simulation Capability?	No	Yes
Can incorporate specific project timelines?	Yes	No
Time to run a forecast?	45 minutes	3 minutes
Forecasting capabilities?	Commercial Manufacturing Development Manufacturing	Development Manufacturing Development FTEs
Forecasting timeline?	8 years	8 years

Figure 25 – Commercial and Development Model Qualitative Comparison

Regardless of the modeling application, it is very useful to compare the new model with the currently accepted best practice. This allows a better understanding of the new model's capability and accuracy. The exercise also exposes opportunities to improve the new model if it is lacking important capabilities.

CHAPTER 5 – Case Studies on Strategic Decision Evaluations

This chapter illustrates how a long term resource model is used to evaluate strategic decisions. The impact on resources is evaluated for three potential evaluations: a drug development process improvement evaluation, an outsourcing evaluation, and an "at risk" development evaluation.

5.1 Development Process Improvement Evaluation

The resource planning model is used to evaluate the impact of development process improvements, when tasks are either added or removed, and estimate the effect on the number of resources required. In this section, an actual example is presented and other possible uses for the model are discussed.

Historically, Novartis Biologics has developed cell culture products using a three-step approach:

1. Platform Process – Develop cell line, master cell bank, and general process. Product from this step is used for Preclinical studies and Phase I clinical trials.
2. Advanced Process – The general process is reviewed and revised to improve product titer and yields. Product from this step is used for Phase II clinical trials.
3. Final Process – Lastly, the advanced process is adjusted and finalized for large scale commercial manufacturing. Product from this step is used for Phase III clinical trials.

Every time the manufacturing process changes, there is potential that the product has also changed due to different manufacturing conditions. Clinical trials begin using the platform process product. After each process change, "comparability studies" must be performed to show that the product is comparable to the product used in the ongoing clinical trials.

Ideally, there would be minimal process changes to reduce the number of comparability studies and develop drugs more efficiently. Novartis Biologics has gained experience in drug development, improving its knowledge base and ability to efficiently develop manufacturing processes. Additionally, the process development group has been able to produce more efficient cell lines, allowing for higher titers. With these advancements, Novartis Biologics is now able to reduce cell culture development to a two-step process:

1. Platform/Advanced Process - Develop cell line, master cell bank, and manufacturing process for development use. Product from this process is used for Preclinical studies, and Phase I and Phase II clinical trials.
2. Final Process – Manufacturing process is finalized for large scale commercial manufacturing. Product from this process is used for Phase III clinical trials.

This process improvement offers many resource savings by reducing the number of FTEs and manufacturing campaigns required. The model developed through this thesis offers a method to evaluate the resource savings.

As mentioned in Chapter 3.4, the model considers seven different product types for Novartis. Two of those product types are cell culture two-step (p = cc2) and cell culture three-step (p = cc3). As explained here, different resources are required for these two product types ($\mathbf{FTE_{cc2,s}} \neq \mathbf{FTE_{cc3,s}}$) Figure 26 compares the two-step and three-step processes using the model development framework.

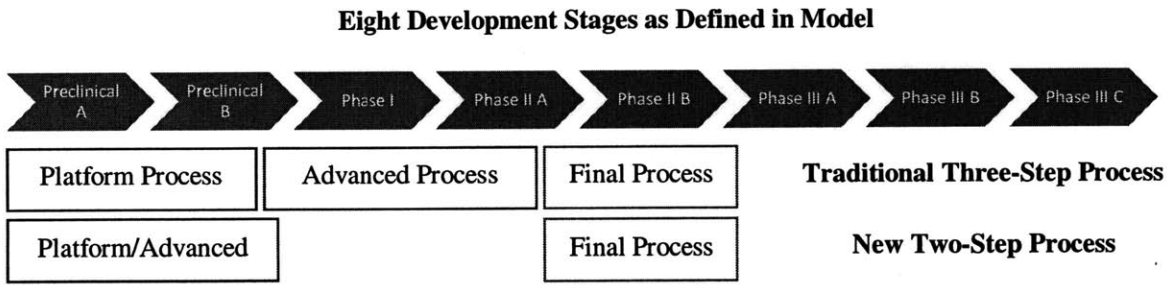

Figure 26 - Three-step and Two-step development processes over the eight development stages.

The current 2010 project plan includes several newer projects that will be developed using the two-step development approach. However, certain projects that started with three-step development and are further along in the development stage must continue with the three-step development strategy. Additionally, some projects are particularly challenging, or exhibit low titers, and will require an additional development stage. For these reasons, the cell culture projects in 2010 are a blend of two-step and three-step development projects. Using the model, the resources required to execute the current cell culture projects were projected and compared to the number of FTEs that would be required to run all cell culture projects with the two-step

method, or all cell culture projects with the three-step method. The results of this comparison are shown in Figure 27. The actual FTEs have been removed, but percent differences are shown in the dialog boxes.

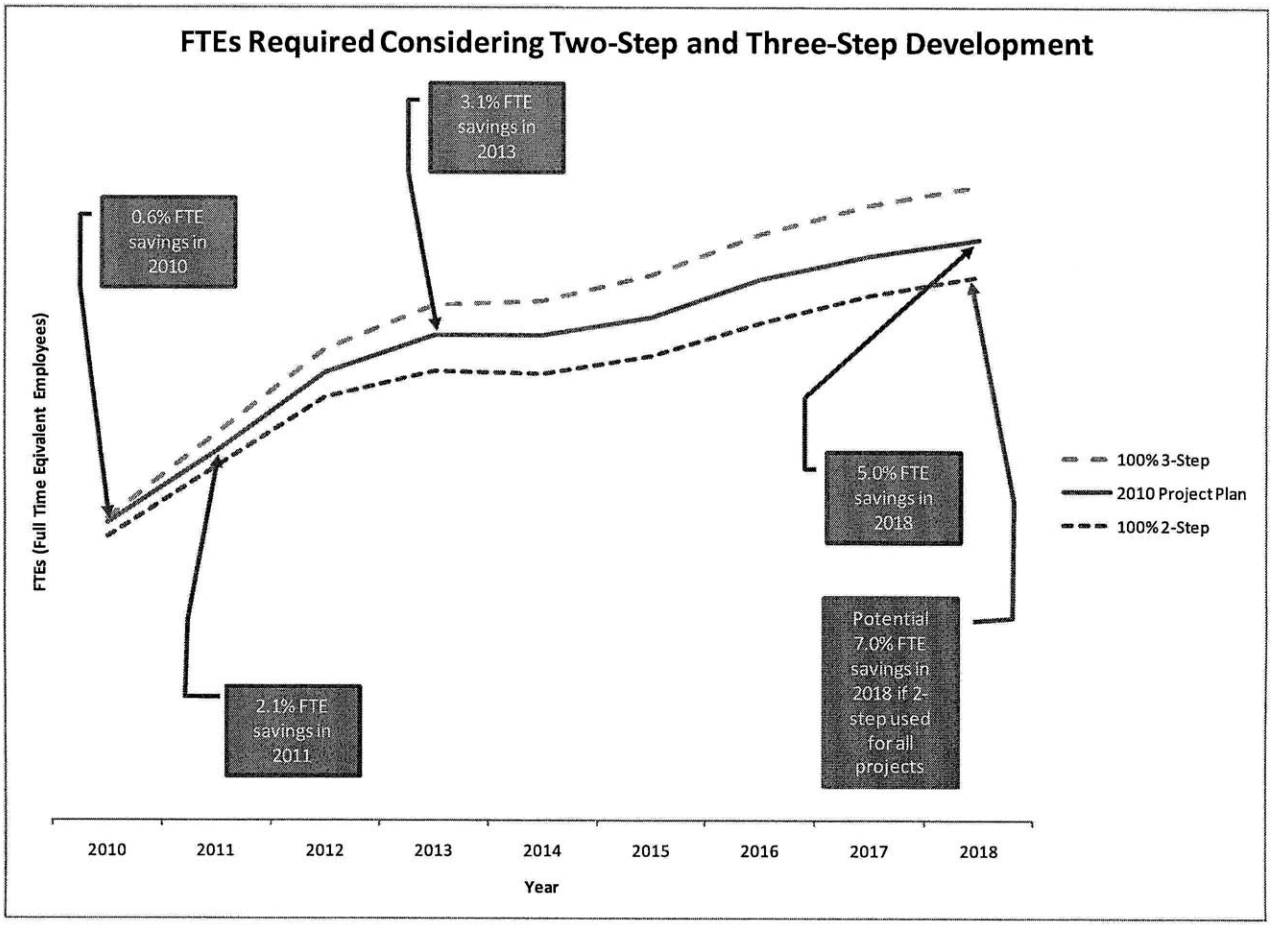

Figure 27 - Comparing FTEs Required for Two-Step and Three-Step Process Development

By using the two-step development for some of its products, Novartis Biologics is saving 0.6% in FTE resources for 2010, showing that this is a more efficient development approach. In 2011, two-step development enables 2.1% in FTE savings. Continuing to use the two-step development as planned would offer a 5.0% FTE savings in 2018. If all cell culture products could be developed using the 2-step method, there is potential for 7.0% FTE savings in 2018. Since Novartis Biologics is expanding quickly, these FTE savings through more efficient drug development practices are critical to success.

Conclusion

This evaluation demonstrates how potential business process improvements could be evaluated using the resource model. Quality by Design (QbD) is another initiative within Novartis, as well as within many other pharmaceutical companies. QbD recommends moving many quality studies (requiring resources) earlier in process development. Since detailed quality testing earlier means testing many projects that will eventually be terminated, QbD could potentially require a greater number of resources. The impact on resources could be evaluated in a way similar to the three-step vs. two-step development resource requirements evaluation.

5.2 Outsourcing Evaluation

Since Novartis Biologics has been growing quickly, it has not had the internal resources or capacity available to complete all of the development tasks necessary. Novartis Biologics has initiated outsourcing to prevent slowing down projects. In the past year, management has developed an outsourcing strategy, and built close relationships with its primary outsourcing partner. Novartis Biologics has found that supporting an outsourced project still requires some level of internal resources from every functional group. This model is used to determine the impact that outsourcing has on internal FTEs and manufacturing capacity requirements.

In the following example, the current outsourcing strategy is compared to three other strategies. Figures 28 and 29 show the differences in FTEs and manufacturing campaigns using the different outsourcing strategies. The "Current Strategy" represents the resource and capacity forecasts if Novartis continued to use the current outsourcing levels. Option 1 represents outsourcing one more project per year, Option 2 represents outsourcing two more projects per year, and Option 3 represents outsourcing three more projects per year. In order to run this analysis, all assumptions are fixed except for the classification of incoming projects. One of the product types listed in Chapter 3.4 is "External" ($p = ex$); this means that the development work is outsourced and performed externally. In this case, the resources required for external (or outsourced) products at each site are $FTE_{ex,s}$ and $MC_{ex,s}$, for FTEs and manufacturing campaigns, respectively.

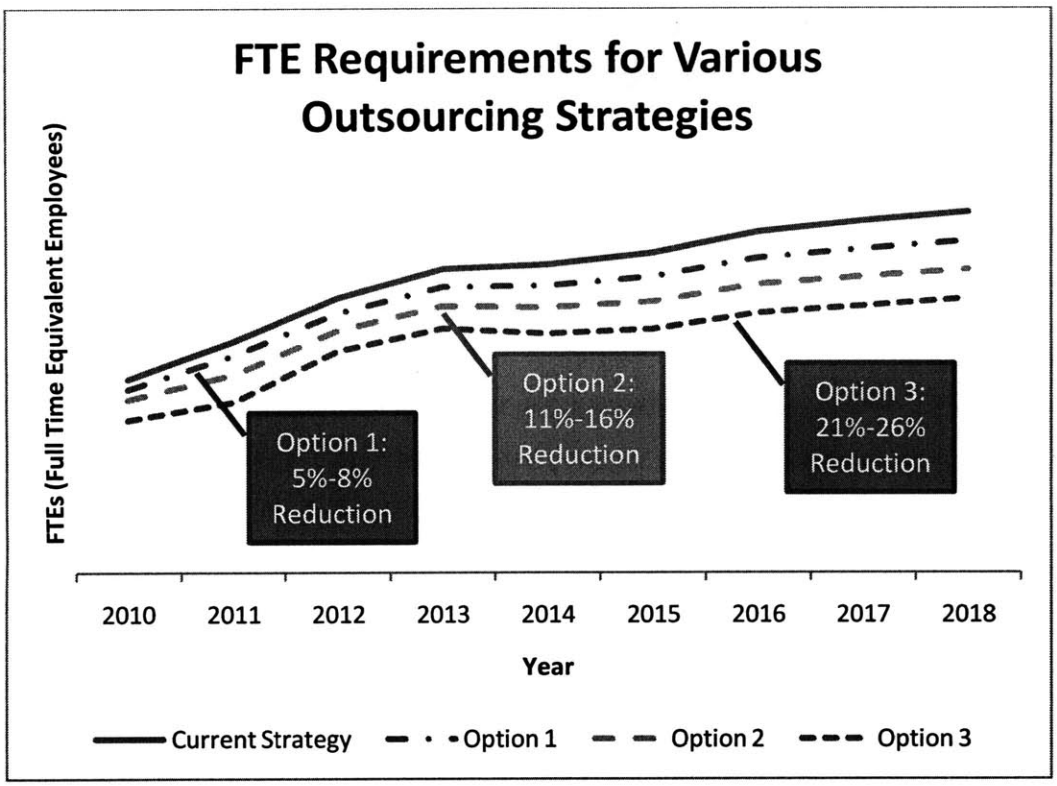

Figure 28 - FTE Requirements for Various Outsourcing Strategies

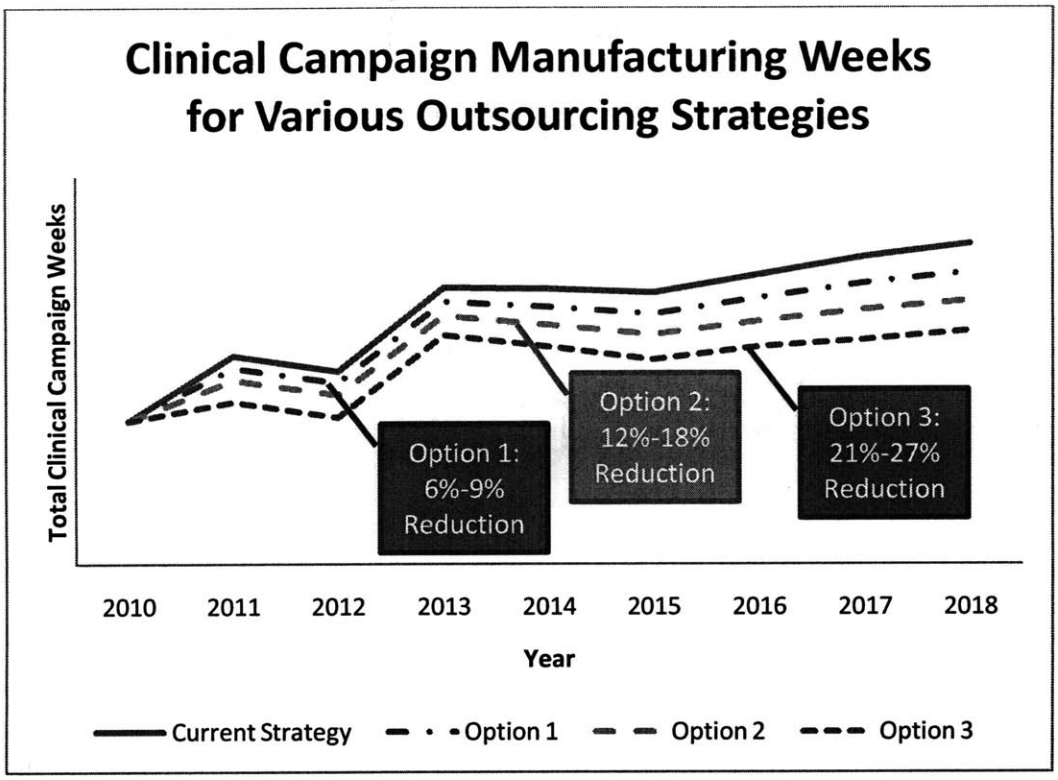

Figure 29 - Clinical Campaign Manufacturing weeks for Various Outsourcing Strategies

In the different strategies, more projects were outsourced, and the number of "in-house" projects was reduced according to the increase in outsourcing. Since the outsourcing strategy can only be applied to projects going forward (i.e., projects currently started in-house cannot be outsourced), there is less of an impact today and more of an impact in the future.

Conclusion

The information generated from this model is used to estimate the financial savings of reduced internal resources vs. the additional cost of outsourcing. For instance, Option 1 reduces FTE requirements by 5%-8% and manufacturing requirements by 6%-9%. If the cost savings of that resource or capacity reduction is greater than the cost of outsourcing, this may be a good case to support increasing the outsourcing strategy. Also, the forecast evaluation allows management to examine manufacturing capacity requirements. When expanding, Novartis has the choice between building more manufacturing capacity or outsourcing that capacity. This model offers data to aid in that decision.

5.3 At Risk Development Evaluation

Many pharmaceutical companies accelerate development activities to ensure there are no clinical trial delays. For select products, management may choose to perform some activities before clinical outcomes that normally trigger those activities are complete. These activities are considered at risk, because development is proceeding at the risk of a negative clinical trial outcome. If the product fails, these "at risk" activities are a waste. If the product succeeds, then it can be advantageous to be ready with the next development step, which could help some products reach the market more rapidly, thereby increasing the opportunity for market share gain.

This case study demonstrates how the model can be used to examine the impact of at risk development on the number of resources required. As defined in Chapter 3.2, the fraction of projects performed at risk is represented by $R_{p,d}$. By changing $R_{p,d}$ and holding all other assumptions constant, resources required for various at risk strategies can be compared. The four "at risk" strategies are shown in Figure 30. For instance, an aggressive at risk strategy is to continue development of 80% of the projects waiting for milestone 1 results. This represents that 80% of projects waiting for milestone 1 results are continuing with further development "at risk."

	% At Risk Assumptions for Different Strategies			
Milestone	No at risk	Modest	Aggressive	All at risk
1	0%	50%	80%	100%
2	0%	50%	80%	100%
3	0%	25%	20%	100%
4	0%	10%	10%	100%

Figure 30 - Assumptions for different "at risk" development strategies

Figure 31 shows the FTE requirements for the different development strategies defined in Figure 30. As the total number of projects increases in the future, the additional resources needed to satisfy projects at risk increase as well. In 2010, performing all projects "at risk" increases the FTE requirement by 10% over the "no at risk" strategy. In 2018, the resources necessary for an "all at risk" strategy is 25% more than those needed for a "no at risk" strategy.

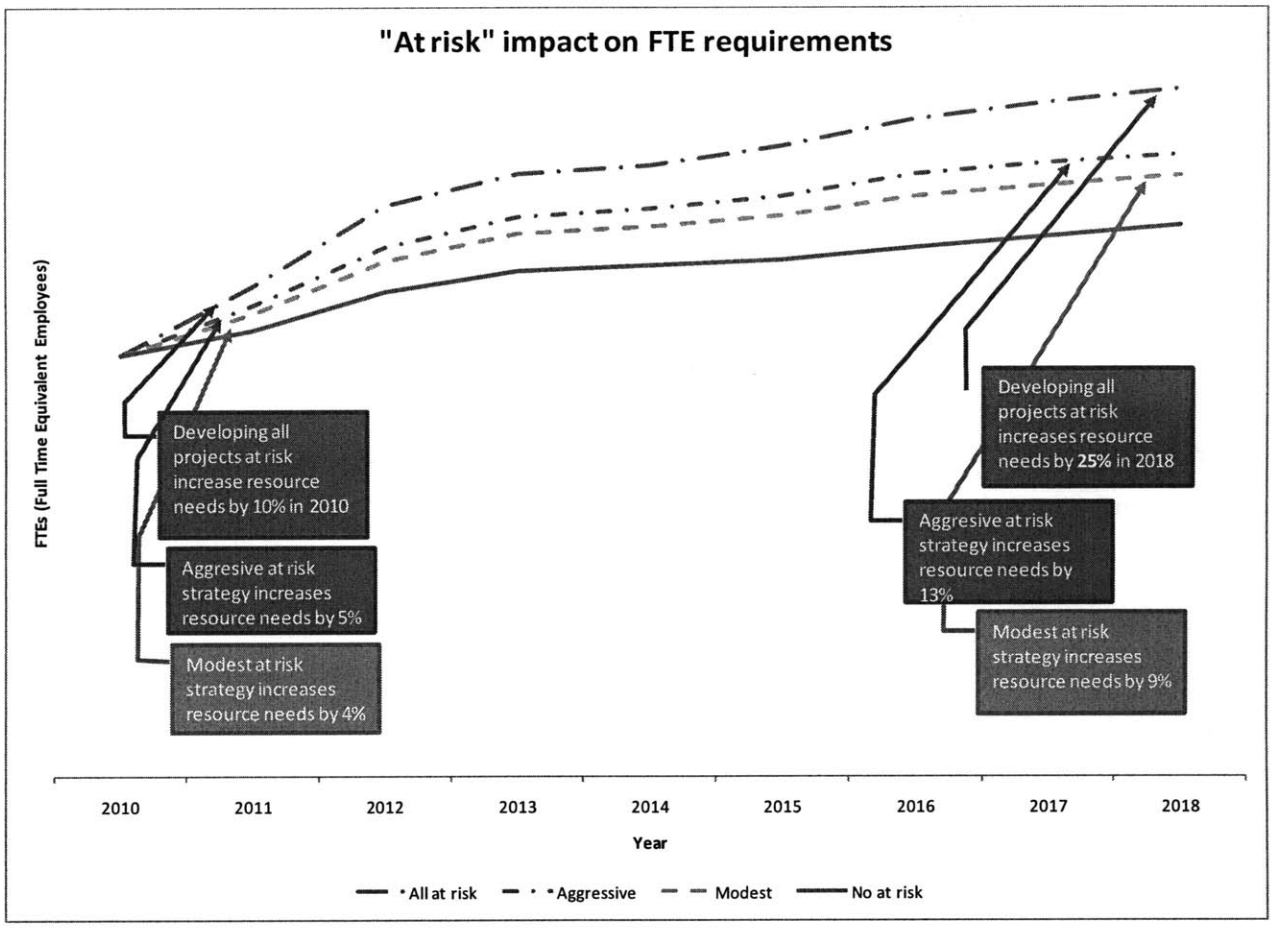

Figure 31 - FTE Projections Considering Various At Risk Strategies

In practice, at risk decisions are made by analyzing each project individually. The cost of proceeding at risk is compared to the potential advantages of the product reaching the market at a faster speed. However, for long term forecasting, one cannot know which projects will be successful and what markets the successful projects will serve. Therefore, it is impossible to make individual "at risk" decisions today for projects well into the future. The purpose of this model is to decouple the individual projects and make a general assumption for the future (considering that the specific projects which will remain in the pipeline are uncertain). By making a general assumption, one can understand how at risk projects might affect the number of resources required.

Conclusion

This final case shows how adjustable assumptions are modified to assess the assumption sensitivity. A similar evaluation could be performed to analyze the impact of the probabilities of success assumptions.

The cases outlined in this chapter are only three examples of how this model can be used to aid in strategic decision making. The impact of process improvements, outsourcing decisions, and business risks can quickly be evaluated. If the model is maintained and updated according to business practices, it can be used to evaluate future strategic decisions impacting FTEs and/or manufacturing capacity.

CHAPTER 6 – Observations and Conclusions

This chapter summarizes the key observations and recommendations resulting from this thesis.

6.1 Key Observations

The key observations are divided into three sections. First the model evaluation results and accuracy are presented. Second, many of the improvements and advantages that the model has shown to date are discussed. Third, some of the potential challenges that may arise using the model in the future are highlighted.

Model Evaluation Results

The model is able to forecast the number of future projects in clinical trials with 40% accuracy when using industry-wide data. The forecast of the total number of launched projects was only accurate within 90%. However, when considering Novartis Biologics internal data, the model was able to predict the overall FTE requirements within 7% accuracy and FTEs by functional group within 50% accuracy (this result is not materially significant because that functional group is very small). The manufacturing campaigns per site are estimated within three campaigns. When comparing currently accepted manufacturing projections for 2010 and beyond, the model estimated within 13%. According to Novartis capacity planners, this degree of uncertainty is acceptable for long range planning purposes and is an improvement over the previous forecasting methods.

Model Advantages

This model delivered additional improvements to the forecasting and strategic planning processes at Novartis. First, the model offers a platform to share common assumptions. Before this model was developed and utilized, each functional group forecasted its own resource requirements independently and project assumptions were not aligned. The model use now helps to align assumptions across functional groups. Additionally, users can now perform sensitivity analyses according to how assumptions may change and evaluate their impact without needing to contact each functional group individually. Secondly, the model created a dialog among many of the functional groups at multiple sites. Due to this model, some of these groups engaged in

conversations about workflow and tasks required for each development stage. This creates a better understanding of the work required by each group, which can reduce rework and prevent tasks from "slipping through the cracks." Thirdly, the model improves communication among the upstream and downstream development groups within Novartis. There is a gradual transition between development and commercial manufacturing, as many commercial sites will produce the last clinical trial material in preparation for large scale manufacturing. With this model, the commercial manufacturing groups also gain better insight into the number of incoming projects. Lastly, the model helps Novartis Biologics plan capacity expansions strategically, deciding when to outsource and when to expand their capacity. This provides a solid foundation to evaluate and justify these decisions to upper level management within Novartis.

Potential Challenges

Although this model presents advantages, there are also many remaining limitations and challenges to overcome. First, the model is very complex and can only be used by a trained set of users (designated as the capacity planning heads). Because of this, each of the functional groups must communicate more closely with these heads to make sure that all tasks and requirements are sufficiently incorporated into the model. Then, as process changes occur, the functional groups need to update the strategic heads rather than just updating their own assumptions. The model needs accurate data to produce reliable results; this data may be difficult to collect and update on a regular basis.

Secondly, due to the model's complexity, it is very easy to overlook assumptions that could make a significant difference in the outcome. Currently, the capacity planning heads are very aware of these assumptions, but if more people use the model, there is a danger in making inappropriate assumptions. Additionally, the model is equipped with the capability of using Crystal Ball software. However, this software is expensive (approximately $1,800 per user) and requires a learning curve, which means that not everyone will have the access or the ability to use that software.

Thirdly, the Novartis Biologics is undergoing many organizational changes. In some cases, it could be either easy or difficult to accurately reflect those changes in the model. Lastly, capacity expansions can be a highly political topic in a corporate environment. Even though the model

offers a legitimate basis to make expansion decisions, its complexity could lend itself to the presumption that someone who understands the details has the control, and someone who is less familiar with the details loses control. For this reason, a model must be presented transparently, exposing all assumptions, for people to understand and believe in its results. The successful implementation of this model will depend not only on data reliability, but also on transparency and general acceptance.

6.2 Recommendations

For a model to be useful, it must contain accurate data, produce reliable results, and be accepted by its users. A well defined process to ensure that data are collected from the correct sources and updated periodically is important for obtaining accurate data. This process could involve an annual data review. Following the review, the two model owners could generate a periodic report to distribute to the appropriate stakeholders. The report should present the data that is most helpful to key decision makers because understanding the concerns of these individuals will give insight to more improvements. Close dialog among the model owners and upper management/key decision makers may help gain support for and increase the usefulness of the model.

Since it is impossible to predict the exact pipeline composition in the future, the model only considers "average" resource requirements. In other words, each project is assumed to require the same number of resources as other projects of the same product type. This is important for future projects, where the resource requirements are not known. Interestingly, Novartis Biologics uses another resource management program for short term resource management. Technical project leaders input resource requirements for currently ongoing projects. If the model can easily use information downloaded from the short term planning source, then this could significantly help in estimating short term resources.

The model is able to run many scenarios using the Crystal Ball software, but the scenarios are based on expected value calculations that reflect risk neutrality. Suppose that two projects have a 50% probability of success and each project that succeeds requires 100 FTEs for development support. According to the current design, the model will always expect an average of one project

to pass and one project to fail; this setup will always expect a requirement of 100 FTEs. If the model were designed with discrete event simulation and a project was either "on" or "off", then the output would show that 25% of the time no projects pass and 0 FTEs are required, 50% of the time one project passes and 100 FTEs are required, and 25% of the time both projects pass and 200 FTEs are required. This gives greater insight to the possible variability of the outcomes, and would be a better use of the Monte Carlo simulation capability. Even though the model gives an accurate estimate of the expected value, the simulation does not represent the possible outcome variations, providing another opportunity for future model improvement.

Finally, the model's complexity may become a hindrance to its expansion into other applications. For instance, the model is currently being considered for expansion in two directions. It could be expanded further for Novartis Biologics to include more detailed financial assumptions. Most often, when making strategic decisions, the financial results play a significant role in that decision, which means that this would be a very useful addition for Novartis Biologics. On the other hand, Novartis as a whole is considering how to best use common manufacturing capacity across divisions. A model such as this one could be an excellent tool for that application. However, to be successful, the model would need to be less granular and specific and at the same time, cover more product variety. It is difficult to create a "one model fits all" solution that adequately balances the complexity that makes it accurate with the simplicity that makes it broadly applicable. This could be an upcoming challenge for this model. The recommendation is to choose a specific purpose and application for the model and craft the model such that it satisfies that purpose.

6.3 Conclusions

This thesis demonstrates the following three points:

- The drug development process can be described by a flexible and scalable model.
- Reliable assumptions enable the model to provide reliable resource forecasts.
- The model can be used for strategic decisions such as process improvements, outsourcing, and decisions to develop products at risk.

The first two points are supported by empirical evaluations of the model, described in Chapter 4. All results were satisfactory to the model users and support the theory that the drug development process can be modeled and accurately forecasted. The third point is supported by the case studies in Chapter 5. The three cases examined are a very limited set of the possibilities for which this model could be used for strategic decision making. Simulation modeling is a useful resource planning tool, especially in areas of high uncertainty and variability.

BIBLIOGRAPHY

Azoulay, Pierre. 2004. Capturing Knowledge within and across Firm Boundaries: Evidence from Clinical Development. *The American Economic Review*, 94(5), 1591-1612.

Beckman, Sara L., and Don B. Rosenfield. 2008. *Operations Strategy: Competing in the 21st Century*. New York: McGraw-Hill/Irwin.

Charnes, John. 2007. *Financial Modeling with Crystal Ball and Excel*. New Jersey: John Wiley & Sons, Inc.

Conant, Tamara. 2009. Modeling Variability for Biologics Strategic Planning. MS Thesis, Massachusetts Institute of Technology.

Heinzle, Elmar, Arno P. Biwer, and Charles L. Cooney. 2006. *Development of Sustainable Bioprocesses: Modeling and Assessment*. West Sussex, England: John Wiley & Sons, Ltd.

Hynes, Martin D. III., Robert D. Swarts, Nichole R. Egli, and Jeanette M. Colonna. 2007. Development, Implementation, and Validation of a Capacity Management System. *Drug Information Journal* 41: 213-220.

Krueger, Donald A., and John M. Kohlmeier. 1972. Financial Modeling and 'What If' Budgeting. *Management Accounting 53*, 25.

Mun, Johnathan. 2006. *Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization Techniques.* New Jersey: John Wiley & Sons, Inc.

Novartis AG. 2010. More about Biosimilars. Accessed on 24 April 2010 via http://www.sandoz.com/site/en/product_range/more_about_biosimilars/glossary.shtml.

Novartis Annual Report. 2009. Basel, Switzerland: Novartis AG.

PAREXEL'S Bio/Pharmaceutical R&D Statistical Sourcebook. 2009. Boston, MA: PAREXEL International Corp.

Repenning, Nelson P.. 2000. A dynamic model of resource allocation in multi-project research and development systems. *System Dynamics Review 16*, no. 3, (October 1): 173.

Schuler, Michael L., and Fikret Kargi. 2002. *Bioprocess Engineering: Basic Concepts.* Upper Saddle River, NJ: Prentice Hall PTR.

Stahl, J.. 2008. Modelling Methods for Pharmacoeconomics and Health Technology Assessment: An Overview and Guide. *PharmacoEconomics 26*, no. 2, (February 1): 131-148.

APPENDIX A: Glossary

CMC (Chemistry, Manufacturing, and Controls): Information a pharmaceutical company must submit to the FDA to certify that the chemistry, manufacturing and controls are in order to produce product of a consistent quality.

CMO (Contract Manufacturing Organization): Companies that perform drug development or manufacturing on contract for a pharmaceutical company.

FDA (United States Food and Drug Administration): Agency that reviews and approves all food and drug products that are sold in the United States.

FTE (Full Time Equivalent): This is the amount of work equal to one full time employee. For example, two half time employees are considered one full time equivalent.

mAb (Monoclonal Antibody): Large molecule drug product that can be designed to interact with or mimic the immune system. This is a biologic compound that is typically produced in mammalian cell culture.

NME (New Molecular Entity): A compound or molecule that is patent protected.

APPENDIX B: Model Data Collection

Interviews were conducted with 33 Novartis employees from many functions including: process and analytical development, project leadership, finance, management, capacity and operational planning, portfolio management, manufacturing planners, and plant managers. Of the 33 people interviewed, two individuals are the model owners.

Model development meetings were held with the two model owners once per month for six months. The following bullets outline the timeline of the model design process:

- Month One: Understanding current state and project scope
  The first meeting was used to discuss Tamara's original model and understand what the stakeholders liked the best about it, and where the model needed improvement. Many ideas were posed and prioritized on a list.
- Month Two: Present "mock-up" models to gain user feedback
  At the second meeting, several "mock-ups," meaning examples of how the model could look, were discussed. This approach afforded flexibility, since the mock-ups could be modified quickly while talking with the key users.
- Month Three: Model test run
  The first version of the new model was distributed a few days before the third meeting, and the users were able to try using the model, and give further feedback to improve usability. This approach was very helpful in identifying bugs or improper assumptions within the model. By adjusting the inputs and reviewing the sensitivity of the outputs, areas for additional data collection could be identified. After the third meeting, all of the agreed upon changes were implemented into the final model, and the focus shifted onto reliable data collection.
- Months Four and Five: Data collection
  The fourth and fifth meetings focused entirely on collecting the correct data, updating assumptions and identifying bugs and corrections within the model.
- Month Six: Model Evaluation
  Preliminary results from the empirical evaluations (Chapter 4) were reviewed and discussed.

APPENDIX C: Time Factor Matrix Calculation Detail

The standard logic for each cell (except the "Launched" row):

- There are two situations where project work is NOT performed (time factor = 0)
    - All work for previous stages have not been completed

      Excel code: IF SUM(years for all previous stages)>year, THEN time factor = 0
    - All work for that stage was completed in previous years

      Excel code: IF SUM(years for all previous stages + this stage)<year, THEN time factor = 0
- There are two situations where project work IS performed (0 < time factor < 1)
    - Stage is completed in that year

      Excel code: IF both cases above are false

      ANDIF SUM(years for all previous stages + this stage)>year, THEN time factor = (1-all fractions of stage completed in previous years)
    - Stage is worked on but not completed

      Excel code: IF all cases above are false, THEN time factor = {(year-years or all previous stages-fractions of stage performed in previous years*years for this stage)/(years for this stage)}

The launched row will equal 1 if all of the development stages 1-8 are complete; otherwise, it will equal 0.

Table C1 shows the matrix when the time in each phase is longer or shorter than expected, and the logic programmed into a few cells are shown to demonstrate the logic.

Table C1 – Illustration of Time Factor Calculation

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
4														
5	**Stage 1 Project Time factor**	Dev. Stage	Yrs per Stage	**2010** 1	**2011** 2	**2012** 3	**2013** 4	**2014** 5	**2015** 6	**2016** 7	**2017** 8	**2018** 9	**2019** 10	**2020** 11
6	CSP	1	2	0.5	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	Preclinical	2	2	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	PhI/PoC/PhIIa	3	1	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
9	PhIIb	4	1	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
10		5	1	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
11	PhIII	6	0.33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
12		7	0.33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
13	Submission	8	0.33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
14	Launched			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0

For instance, cell **F7** contains the following logic:

=IF(SUM(C6:C6)>F5;0;IF(SUM(C6:C7)<=E5;0;IF(SUM(C6:C7)<=F5;1-SUM(D7:E7);(F5-C6-SUM(D7:E7)*C7)/C7)))

To break it down as described before:

=IF(SUM(**C6:C6**)>=**F5**;0;	FALSE, stage is not waiting for previous stages
IF(SUM(**C6:C7**)<=**E5**;0;	FALSE, stage is not already complete
IF(SUM(**C6:C7**)<=**F5**;1-SUM(**D7:E7**);	FALSE, stage is not completed within the year
(**F5-C6**-SUM(**D7:E7**)*C7)/C7)))	(3-2-0)/2 = **0.5 = 2012 Preclinical time factor**

Similarly, cell I8 contains the following logic:

=IF(SUM(C6:C7)>I5;0;IF(SUM(C6:C8)<=H5;0;IF(SUM(C6:C8)<=I5;1-SUM(D8:H8);(I5-SUM(C6:C7)-SUM(D8:H8)*C8)/C8)))

Again, described in detail:

=IF(SUM(C6:C7)>=I5;0;	FALSE, stage is not waiting for previous stages
IF(SUM(C6:C8)<=H5;0;	TRUE, **0 = 2016 PhI/PoC/PhII time factor**
IF(SUM(C6:C8)<=I5;1-SUM(D8:H8);	
(I5-SUM(C6:C7)-SUM(D8:H8)*C8)/C8)))	

This time factor matrix is slightly different for projects that start in the future versus projects that have already started. For projects that start in the future, the time factors for each year prior to the start will be zero. For projects that have already started, the current stages do not depend on the length of the previous stages that have already been completed (all completed stages will have a required time of zero).

APPENDIX D: Work Package Calculation Detail

The resources required per product per development stage are determined through a compilation of data from many sources. Each functional group provides a list of tasks that they perform, and a specified number of workdays for each task for each development stage. All of the tasks are compiled to create a work package for each drug product. These work packages define all of the development tasks required for the eight stages of development, along with the number of FTEs, manufacturing campaigns and materials and services costs required. Table D1 shows an example of the work package input and some of the resulting FTE, MC (manufacturing campaign) and MS (materials and services) matrices.

Table D1 - Work package task input section (top), fixed site activities matrices for Site 1 (middle), and flexible site activity matrices for Launch Site (bottom)

Work Package Example for Microbial (m)

			FTEs								Manufacturing Campaigns								Materials and Service Costs (kUSD)							
Activity	Function	Site	S1	S2	S3	S4	S5	S6	S7	S8	S1	S2	S3	S4	S5	S6	S7	S8	S1	S2	S3	S4	S5	S6	S7	S8
Task 1	Group A	Site 1	15	25	20	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100	50	0	0	0	0	0
Task 2	Group B	Site 1	0	0	15	15	20	4	2	3	0	0	0	0	0	0	0	0	0	0	25	50	25	0	0	0
Task 3	Group A	Site 2	10	10	10	10	10	10	10	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Task 4	Group C	CC	5	15	10	5	0	0	0	0	0	1	1	1	0	0	0	0	0	100	100	0	0	0	0	0
Task 5	Group B	DS	0	0	5	15	10	5	0	0	0	0	0	1	1	1	0	0	0	0	0	200	150	0	0	0
Task 6	Group C	LS	0	0	0	0	5	10	15	30	0	0	0	0	0	1	2	3	0	0	0	0	0	300	200	100

Fixed Site Activities - Site 1

	$FTE_{m,1}$								Manufacturing Campaigns ($MC_{m,1}$)								Materials and Service Costs (kUSD) $MS_{m,1}$							
Function	S1	S2	S3	S4	S5	S6	S7	S8	S1	S2	S3	S4	S5	S6	S7	S8	S1	S2	S3	S4	S5	S6	S7	S8
Group A	15	25	20	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100	50	0	0	0	0	0
Group B	0	0	15	15	20	4	2	3	0	0	0	0	0	0	0	0	0	0	25	50	25	0	0	0
Group C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Group D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Flexible Site Activities - Launch Site

	$FTE_{m,LS}$								Manufacturing Campaigns ($MC_{m,LS}$)								Materials and Service Costs (kUSD) $MS_{m,LS}$							
Function	S1	S2	S3	S4	S5	S6	S7	S8	S1	S2	S3	S4	S5	S6	S7	S8	S1	S2	S3	S4	S5	S6	S7	S8
Group A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Group B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Group C	0	0	0	0	5	10	15	30	0	0	0	0	0	1	2	3	0	0	0	0	0	300	200	100
Group D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

In the top input section, the user identifies the tasks and resources required along the eight development stages (S1-S8), and assigns each task to a functional group and site. For the purpose of this demonstration, the Functional Groups A, B, C, and D and Sites 1 and 2 are used. The site assignment can be either a fixed site that always performs that task for all projects of that type, or the site assignment can be flexible if multiple sites have the ability to perform that task. The fixed site activities are assigned directly to the site responsible for the task. The

flexible site activities are classified as activities that occur at the Competence Center (CC), Development Site (DS) or Launch Site (LS).

The fixed site activities will always be allocated to the site responsible for that activity. However, the flexible site activities are assigned based on a site allocation factor, $x_{p,c,s}$, where p=product type, c=center, and s=site. Table D2 shows an example of the flexible activity site allocation factors which are defined by the user.

Table D2 - Example of Flexible Activity Site Allocation Factors

Flexible Activities Site Allocation		Site 1	Site 2
Cell Culture	CC	1	0
Cell Culture	DS	0.5	0.5
Cell Culture	LS	0	1
Microbial	CC	1	0
Microbial	DS	0.5	0.5
Microbial	LS	0	1
Biosimilar Cell Culture	CC	1	0
Biosimilar Cell Culture	DS	0.5	0.5
Biosimilar Cell Culture	LS	0.25	0.75
Biosimilar Microbial	CC	0	1
Biosimilar Microbial	DS	0.25	0.75
Biosimilar Microbial	LS	1	0

For instance, 100% of the cell culture competence center work is performed at Site 1 ($x_{cc,CC,1} = 1$), while 50% of the cell culture development site work is performed at Site 1 and the other 50% is performed at Site 2 ($x_{cc,DS,1} = x_{cc,DS,2} = 0.5$).