AN INVESTIGATION OF CLINICAL TRIAL SUPPLY CHAINS

BY ADAM FLEISCHHACKER

A dissertation submitted to the Graduate School—Newark Rutgers, The State University of New Jersey in partial fulfillment of the requirements for the degree of Doctor of Philosophy Ph.D. in Management Written under the direction of Dr. Yao Zhao

and approved by

Newark, New Jersey October, 2009

ABSTRACT OF THE DISSERTATION

An Investigation of Clinical Trial Supply Chains

by Adam Fleischhacker Dissertation Director: Dr. Yao Zhao

This dissertation investigates production and inventory decisions made within clinical trial supply chains in order to reduce drug supply costs. By investigating the SEC filings of public companies, we find that drug supply costs frequently account for a significant portion of pharmaceutical companies' R&D spending. To unlock value tied up in clinical trial supply chains, three unique aspects of clinical trial supply chains are explored and associated supply chain decisions are optimized. The first unique factor that differentiates the supply chains for clinical trials is the risk of failure, meaning that the investigational drug is proven unsafe or ineffective during human testing. Upon failure, any unused inventory is essentially wasted and needs to be destroyed. We explore the effect of this failure on production planning decisions and find the planner's decision to be a balancing act between waste and destruction costs versus production inefficiency. To optimally achieve this balance, we generalize the Wagner-Whitin model (W-W model) to incorporate the risk of failure. A second unique aspect of clinical trials is that demand can go from being quite unpredictable to fully predictable during the course of a trial. To take advantage of this demand learning, intra-trial batches can be produced, but at the expense of scale economies. Using various learning curves, we study this balance between learning and economies of scale in a finite horizon inventory model with fixed production costs and two production options: the pre-trial batch and the intra-trial batch. We characterize the optimal policy for both production batches in regards to optimally scheduling and sizing production. Lastly, we analyze the distribution networks of global clinical trial supply chains. Unique to these networks is their temporary existence; trials are ceased after patient enrollment goals are met. To manage these networks, we present a new class of multi-echelon inventory models to make stock positioning decisions, develop algorithms to identify lower and upper bounds on the optimal objective function for this new class, and leverage those algorithms to provide insights into optimal supply chain configurations.

Acknowledgements

I owe the deepest of thanks to my advisor, Dr. Yao Zhao, whose dedication to supporting me has been unwavering. His brilliance is astounding, his course is a highlight of my experience, and his generosity of time and mind are second to none. To my dissertation committee, I express great thanks: to Dr. Xiaowei Xu for his vast knowledge and insightful comments, to Dr. Mahesh Kumar for mentoring me as both teacher and researcher, to Dr. Michael Katehakis for showing me how to make the most difficult of problems look easy, and to Dr. Indranil Nandi whose encouragement and experience are much appreciated.

I also extend special thanks to my family: to my wife, Erin, with whom I share a love that most think only exists in fairy tales and who unconditionally supported me throughout my education; to my two great boys, Aidan and Tristan, who sacrificed hours of Daddy time so that I may type and think in peace and quiet; and to my mother, Fran, for being the most supportive mother and grandmother.

I also thank all the professors who put their time into the Ph.D. program and make it such a rewarding experience. Special thanks to Dr. Lei Lei whose commitment to teaching and bridging the gap between industry and academia I can only hope to emulate in my career. And thank you to both Dr. Zachary Stoumbos and Dr. Michael Rothkopf who both had so much more to contribute to the world. I was fortunate enough to cross their paths.

Dedication

In loving memory of my father, whose wisdom continues to guide me, whose music fills my soul, and whose support has enabled me to pursue my dreams.

Table of Contents

| A | bstra | ct | i |
|----|-------|---|---|
| A | cknov | wledgements | V |
| De | edica | $tion \ldots \ldots$ | V |
| Li | st of | Tables | i |
| Li | st of | Figures | ζ |
| 1. | Intr | oduction | L |
| | 1.1. | Clinical Trials | L |
| | 1.2. | Clinical Trial Supply Chains | 3 |
| 2. | Clir | ical Trial Supply Chain Spending | 7 |
| 3. | Pro | duction Planning Under Failure Risk | L |
| | 3.1. | Literature Review | 1 |
| | 3.2. | The Model |) |
| | 3.3. | Extensions | 3 |
| | | 3.3.1. Concave Cost Functions | 3 |
| | | 3.3.2. Additional Constraints | 7 |
| | 3.4. | An Illustrative Example | 3 |
| | 3.5. | Numerical Study of Potential Savings |) |
| | 3.6. | Summary Remarks | 7 |

| 4. | Ada | ptive Clinical Trials: Balancing Learning and Economies of Scale 3 | 9 |
|----|-------|--|----|
| | 4.1. | Literature Review | 1 |
| | 4.2. | The Two-Period Model | 5 |
| | | 4.2.1. Notation and Model | 5 |
| | | 4.2.2. Optimal Ordering Policy | 8 |
| | | 4.2.3. Demand Learning Model | 2 |
| | | 4.2.4. Impact of The Fixed Cost | 3 |
| | 4.3. | Numerical Analysis | 3 |
| | 4.4. | Summary Remarks | 4 |
| 5. | Inve | ntory Positioning in Clinical Trial Supply Chains | 6 |
| | 5.1. | Literature Review | 9 |
| | 5.2. | Two Special Cases | '1 |
| | 5.3. | General Clinical Trial Supply Chains | 6 |
| | | 5.3.1. Sub-models \ldots 75 | 9 |
| | | 5.3.2. Minimizing Inventory Investment in a Clinical Trial Supply Chain 84 | 4 |
| | | 5.3.3. Solving Mathematical Program (5.25) Using Graves Approximation 89 | 9 |
| | 5.4. | Numerical Analysis | 1 |
| | | 5.4.1. Solution Algorithm Effectiveness | 1 |
| | | 5.4.2. Optimal Supply Chain Configuration | 2 |
| | 5.5. | Summary Remarks | 4 |
| 6. | Con | cluding Remarks | 6 |
| Re | efere | \mathbf{ces} | 8 |
| Vi | ta . | | 6 |

List of Tables

| elated | 1. Percentage of Clinical Trial Expenses Spent of Manufacturing-Relate | 2.1. |
|---------|--|------|
| e 8 | Activities for Ariad Pharmaceutical's Deforolimus Drug Candidate | |
| elated | 2. Percentage of Clinical Trial Expenses Spent of Manufacturing-Relate | 2.2. |
| | Activities for Acushpere Incorporated's Imagify Drug Candidate | |
| Trial | 3. Percentage of Research and Development Spending on Clinical Tria | 2.3. |
| oduct | Manufacturing-Related Activities for Allos Therapeutics' Three Produc | |
| | Candidates | |
| 31 | 1. Parameters for Phase II and Phase III Simulations | 3.1. |
| Phase | 2. Maximum Observed % Savings Versus Number of Setups During Phas | 3.2. |
| 36 | III Simulation | |
| | 1. Baseline Parameter Values | 4.1. |
| s and | 2. Probability of Mid-Season Replenishment for Various Setup Costs an | 4.2. |
| 61 | Learning Rates | |
| 61 | 3. Table of Savings for Various Setup Costs and Learning Rates | 4.3. |
| lerage | 4. Expected Savings Over a Newsvendor Solution for Increasing Underag | 4.4. |
| 62 | Costs | |
| uction | 5. Expected Savings Over a Newsvendor Solution for Increasing Destructio | 4.5. |
| 63 | Costs | |
| oles 64 | 6. Effect of Parameter Increases on Performance and Decision Variables | 4.6. |
| 66 | 1. Growth in Number of Offshore Clinical Trial Sites | 5.1. |
| 92 | 2. Algorithm Performance for Various Supply Chain Configurations | 5.2. |

| 5.3. | Out of 729 test scenarios, frequency of configurations prescribing direct | |
|------|---|----|
| | shipment to sites for that country | 93 |
| 5.4. | Frequency of Optimal Supply Chain Configurations | 93 |
| 5.5. | Frequency of Shipping Direct in an Optimized Supply Chain | 94 |

List of Figures

| 1.1. | Medication dispensing and automated site inventory control using IVR . | 4 |
|------|--|----|
| 3.1. | Production Costs Excluding & Including the Risk of Failure | 29 |
| 3.2. | Expected Reduction in Costs Using F-R Algorithm in Phase II Simulations | 33 |
| 3.3. | Expected Reduction in Costs Using F-R Algorithm in Phase III Simulations | 35 |
| 4.1. | Expected Costs when $\kappa = \$0$ | 56 |
| 4.2. | Expected Costs when $\kappa = \$1,600$ | 57 |
| 4.3. | Expected Costs Versus Timing of Second Replenishment with Setup Costs | 58 |
| 4.4. | Expected Costs Versus Timing of Second Replenishment with Setup | |
| | Costs and Demand Learning | 59 |
| 4.5. | Optimal First Batch Size Versus Timing of Second Replenishment with | |
| | Setup Costs and Demand Learning | 60 |
| 5.1. | Timeline of Events | 71 |
| 5.2. | Picture of Model's Supply Chain Topology | 77 |
| 5.3. | Direct Shipment Supply Chain | 80 |
| 5.4. | Regional Warehouse Supply chain | 80 |

Chapter 1 Introduction

This dissertation investigates production-inventory decisions made within clinical trial supply chains in order to reduce drug supply costs. In this chapter, we describe current clinical trial practices and review the role that academics have played in studying the unique aspects of clinical trial supply chains. In the proceeding chapter, Chapter 2, we confirm that the costs of clinical trial supply chains are indeed an area worthy of study. By investigating the SEC filings of public pharmaceutical companies, we find that clinical drug supply costs account for a significant portion of the R&D spending. Given this large expense, we believe that this is a portion of drug development that is ripe for improvement and in Chapters 3-5 we improve upon existing supply chains. For these models to incorporate aspects that are unique to clinical trial supply chains. For these models, we are rewarded with both techniques and insights to reduce drug supply costs for clinical trials.

1.1. Clinical Trials

Clinical trials are one of the most important hurdles that a drug candidate must overcome prior to regulatory clearance. Before the U.S. Food and Drug Administration's (FDA) approval of a new drug, testing of a drug candidate in clinical trials is performed to ensure both the safety and effectiveness of the treatment in human subjects. Three phases of clinical trials are usually required and in each phase, the drug candidate is tested against either a placebo and/or an already commercialized medication. Typically, Phase I involves 50 to 100 healthy individuals and is testing the safety of the medicine in human subjects for the first time. Assuming Phase I is successful, Phase II recruits a few hundred potential patients and further tests safety while also looking at treatment efficacy. Lastly, upon success of Phase II, Phase III seeks to test the drug candidate in a few thousand patients to confirm both the efficacy and the safety of the medication. From initiating Phase I trials to awaiting final FDA approval after Phase III trials have been concluded, the development pipeline is filled with potential:

Today there are more than 2,900 medicines in the development pipeline. Researchers are armed with vast amounts of new information on the genetic and molecular underpinnings of disease, and they are working to translate this knowledge into treatments that can ease symptoms, slow progression and, ultimately, prevent or halt disease. (Pharmaceutical Research and Manufacturers of America, 2009)

For each of the 2,900 medicines referred to above, some of which hope to cure cancer or prevent heart disease, there is a long and expensive journey ahead of them towards either commercialization or failure. In efforts to find successful treatments, drug companies in the United States spend around 20% of their sales on research and development (R&D) activities. Worldwide, it is estimated that this spending exceeds \$65 billion per year with some estimates topping \$100 billion (Pharmaceutical Research and Manufacturers of America, 2009; Thomson CenterWatch, 2007). On average, 37% of the \$100 billion R&D spending by pharmaceutical companies is spent on the clinical trial process (Thomson CenterWatch, 2007). Despite these vast sums of money being invested, only about 20% of drug candidates that enter clinical trials actually succeed (DiMasi, Hansen, and Grabowski, 2003). Given these large sums of money being spent on research, the small percentage of successful drug candidates, and revenues that are proving harder to come by, it is natural that pressure exists to reduce costs in the industry (Shah, 2004). The most challenging issues in managing today's clinical trials are patient recruitment and securing drug supply. Patients are enrolled into trials through both doctors and hospitals that agree to aid the study. For a pharmaceutical company, enrollment is often considered the bottleneck that prevents timely completion of trials and 80% of clinical studies fail to meet their recruitment goals (Drennan, 2002). As a result of slow patient recruitment and also high patient costs in the United States, clinical trials are increasingly going global (Rowland, 2004). As a result of globalization, drug supply must now address the needs of multiple international regulatory bodies, multiple languages and cultures, and multiple investigative sites. Thus, it is not surprising to hear the sentiment that "Most current supply chains are entirely inadequate for the realities of global trials today." (Neuer, 2008).

1.2. Clinical Trial Supply Chains

Clinical trial supply chains of today are unique and challenging. While in the past pharmaceutical companies considered drug supplying costs (e.g., manufacturing and logistics costs) to be negligible, it is now recognized that these costs can absorb 20% of the final value of a drug (Papageorgiou, Rotstein, and Shah, 2001). And just as the attention of pharmaceutical executives has shifted from increasing revenues towards reducing costs, so has the attention of academics. Around the turn of the millenium, there is noticeable increase in attention devoted to drug supply issues during clinical trials (Bernstein and Hamrell, 1999, 2000).

Papageorgiou, Rotstein, and Shah (2001) is one of the first papers to advocate optimization methodologies for capacity planning that includes clinical trial manufacturing as part of the decision process. Their research presents a mixed integer linear program for modeling a pharmaceutical company's supply chain. In an offshoot of their work, Rogers, Gupta, and Maranas (2002) present a methodology to create an optimal portfolio of drug candidates. Other papers for choosing the right portfolio of candidates to enter or continue clinical trials followed (Levis and Papageorgiou, 2004; Maravelias and Grossmann, 2004). It is around the time of these articles that Nilay Shah notes "A great deal of research has been undertaken on facility location and design, inventory and distribution planning, capacity and production planning and detailed scheduling. Only a small proportion of this work directly addresses the issues faced in the pharmaceutical sector. On the other hand, this sector is very much ready for and in need of sophisticated supply chain optimisation techniques." (Shah, 2004)

Sophistication in the clinical trial supply chain has often been implemented because of enabling technologies. One such technology is Interactive Voice Response Systems (IVRS). IVRS is a system that enables real-time monitoring of inventory in the clinical trial supply chain. Leveraging Figure 1.1 which is presented in Byrom (2002), we are able to describe the operational aspects of an IVRS system within a clinical trial.



Figure 1.1: Medication dispensing and automated site inventory control using IVR

When a qualifying patient arrives to a clinical trial site, the participating physician calls into the IVRS system alerting the system to the new patient (step 1). With a database of site-level inventory, the IVRS system returns the kit number or pack number of the medication that is to be assigned to this patient (step 2). While neither doctor nor patient knows if this particular kit is a placebo or the actual drug candidate, the IVRS has allocated the most appropriate inventory to ensure a randomized trial (McEntegart and O'Gorman, 2005, present an overview of randomization using IVRS). Once inventory levels at the site fall below a pre-determined trigger point, a re-order is placed to a central or regional depot for resupply (steps 3-5). Finally, the order is shipped and upon arrival is properly inventoried for potential dispensing to future patients via the IVRS system (steps 6-8).

While IVRS provides a method of executing an operational plan, simulation is the enabling technology for creating the operational plan. This plan would have to answer a myriad of questions:

- How much drug will need to be supplied?
- How many clinical trial sites should I open and when should I open them?
- How and when should I change operational policies in response to demand?
- Should I ship direct to sites or should I open regional warehouses to cut leadtimes?

The above questions, which are purely tactical in nature (as opposed to strategic or operational), have not secured the much needed academic attention, but many in industry have advocated the use of simulation for answering these and other tactical questions (Peck, 1997; Dowlman, Lang, McEntegart, Nicholls, Bacon, and Star, 2004). In practice, simulation has been used for several decades as part of the clinical trial decision making process (Chow and Chang (2006)).

While technologies, like IVRS and simulation software, have certainly played a key role in advancing the efficiency of clinical trial supply chains, one idea is being promoted to enhance the effectiveness of clinical trials. Specifically, adaptive clinical trials are being proffered as a means to both increase successes and weed out failures earlier in the clinical trial process (Lowe, 2006). Instead of the traditional rigid approach to clinical trials where a dosage, a patient population, a length of treatment, and specific measures of success are chosen all prior to commencement of the trial, adaptive clinical trials allow for flexibility in how a trial is conducted (Lesko, 2007). For example, an adaptive approach allows for different dosages during the same trial and to drop dosages that prove toxic or non-therapeutic as trial data is collected and analyzed. While this flexible dosing paradigm does aim at increasing the success of the drug, it effectively creates one more unknown for the supply chain as doubling a doseage is essentially doubling demand for a drug. In addition, packaging and shipping considerations must now meet the needs of supporting this flexibility.

Given the dearth of research on tactical decision-support that fulfills the need for current sophistication in clinical trial supply chain and the added pressures of globalization and adaptive designs, the core of this dissertation (Chapters 3-5) provides tactical decision-support models for this changing industry. In answering a call for less waste and overage during clinical trials, we construct and analyze a production planning model in Chapter 3 that incorporates the risk of a trial being halted because the drug proves unsafe or ineffective. In Chapter 4, we model an adaptive clinical trial and study the effect of demand learning on initial batch sizes and replenishment timing. In Chapter 5, we address the issues of stock positioning and supply chain configuration to simultaneously support fast recruitment while ensuring appropriate fill rates at clinical sites. In summary, this dissertation is designed to support many of the key production-inventory decisions that a supply manager faces when supporting clinical trials.

Chapter 2 Clinical Trial Supply Chain Spending

In this chapter, we present empirical evidence that highlights the financial significance of the clinical trial supply chain. The clinical trial supply chain is an enormously complex process and the money flow within this chain lacks detailed breakdowns in the literature. We investigate the Securities and Exchange Commission filing's of publicly traded companies to identify supply chain costs associated with new drug development. We specifically look for smaller publicly traded companies to avoid the intentional ambiguity in larger company statements that makes it difficult to follow the costs associated with any one drug or specific clinical trial. In our search of public records, we found three companies' filings that allow for a more detailed look at clinical trial spending (Acusphere Inc., 2004–2007; Ariad Pharmaceuticals Inc., 2004–2007; Allos Therapeutics Inc., 1996–2007, see).

The first company is Ariad Pharmaceuticals Inc, whose lead product candidate, Deforolimus, remains in Phase III trials as of January 2008. The product is a small molecule compound for treating certain types of cancer. Manufacturing of the product is not enormously complex and the product is readily synthesized using conventional fermentation techniques. Although the manufacturing process was developed in-house, the company has relied on third-party manufacturers to supply its clinical trial material. Because of Ariad's reliance on third-party manufacturers, Ariad's 10-K's from fiscal

| | 2003 | 2004 | 2005 | 2006 | Total |
|---|----------|------------|----------|----------|---------|
| Clinical Trial Phase | Phase I | Phase II | Phase II | Phase II | — |
| Manufacturing Spend (in millions) | \$1.27** | \$5.87 | \$9.37 | \$5.27 | \$21.78 |
| Total Clinical Spending (in millions) | \$2.54 | \$11.54 | \$26.31 | \$15.58 | \$55.97 |
| Manufacturing Spend as % of Total | 50% | 51% | 36% | 34% | 39% |
| <u>++</u> <u>c</u> <u>i</u> <u>i</u> <u>i</u> <u>i</u> <u>i</u> | 1 1 1 7 | 207 61 1 1 | 1 1 | 1. 0 | 2000 |

** manufacturing spend is estimated at 50% of total clinical spending for 2003

Table 2.1: Percentage of Clinical Trial Expenses Spent of Manufacturing-Related Activities for Ariad Pharmaceutical's Deforolimus Drug Candidate.

years 2004 to 2006 specifically mentions changes to clinical trial expenses as a result of changes in manufacturing-related costs. For fiscal year 2003, a breakdown of clinical trial costs is not given so we estimate the manufacturing spend in this year. For 2007, clear breakdowns of manufacturing-related costs are no longer available due to a 2007 deal between Ariad and Merck which allows for the sharing of Deforolimus development costs. Using the available data, we estimate that approximately 39% of Ariad's overall clinical trial spending was spent on manufacturing-related activities (i.e. part of supply chain costs). The supporting data are shown in Table 2.1.

Similar to our analysis of Ariad's annual filings, we find cost information available from the 10-K filings of Acusphere, Inc. This company's lead product candidate, Imagify, is a cardiovascular drug that has completed Phase III clinical trials and filed a new drug application with the FDA in April 2008. Unlike Ariad's Deformilus, Acushpere's Imagify requires custom and proprietary manufacturing technology. Despite this difference, the percentage of clinical trial spending dedicated to manufacturing-related activities is similar to that of Ariad's. As shown in Table 2.2, around 40% of Acusphere's clinical trial spending has gone towards manufacturing-related activities for the fiscal years 2003 through 2006. ¹

In contrast to the previous two companies, Allos Therapeutics has multiple product candidates, nonetheless their annual filings provide usable detail on their clinical manufacturing costs. These costs are separated from all other research and development

 $^{^{1}}$ Acusphere went public in October 2003, thus earlier expense data relevant to our analysis is not available.

| | 2003 | 2004 | 2005 | 2006 | Total |
|---------------------------------------|--------------|-----------|-----------|-----------|--------|
| Clinical Trial Phase | Phase II/III | Phase III | Phase III | Phase III | — |
| Manufacturing Spend (in millions) | \$1.7 | \$7.9 | \$8.0 | \$11.4 | \$29.0 |
| Total Clinical Spending (in millions) | \$8.6 | \$18.5 | \$24.3 | \$21.9 | \$73.3 |
| Manufacturing Spend as % of Total | 20% | 43% | 33% | 52% | 40% |

Table 2.2: Percentage of Clinical Trial Expenses Spent of Manufacturing-Related Activities for Acushpere Incorporated's Imagify Drug Candidate.

| | 1995 - 1998 | 1999 - 2002 | 2003 - 2006 | Total |
|---------------------------------------|-------------|-------------|-------------|---------|
| Manufacturing Spend (in millions) | \$3.7 | \$11.5 | \$13.9 | \$29.1 |
| Total R&D Spending (in millions) | \$18.2 | \$56.7 | \$61.6 | \$136.4 |
| Manufacturing Spend as % of Total R&D | 20.4% | 20.3% | 22.6% | 21.3% |

Table 2.3: Percentage of Research and Development Spending on Clinical Trial Manufacturing-Related Activities for Allos Therapeutics' Three Product Candidates.

costs. Unfortunately, total clinical trial costs are lumped into Allos Therapeutics' R&D number, so a direct comparison of Allos' spending to our previous two examples is not possible. However, we do have data on twelve years of clinical trial manufacturing costs as a percentage of total research and development spending. Through analysis of these 12 years, we find that 21.3% of Allos Therapeutics' R&D spending from 1995-2006 has been spent on clinical trial manufacturing-related activities. See Table 2.3 for the yearly breakdown of spending:

It is worth noting that the two drug candidates in the first two examples are small molecule drugs. Discussions with pharmaceutical R&D researchers and industry consultants unanimously indicate that large molecule treatments, such as those developed by bio-tech companies using live organisms or their components, tend to have much more significant drug supply costs during clinical trials than small molecule chemical (nonbiologic) compounds. This is true because large molecule drugs typically require many more steps in production, have much lower yield, and require extra costs on shipping and storage. Our third example comes from a large molecule bio-tech company, and anecdotally, this supports the idea that large molecule manufacturing costs are more costly than small molecule manufacturing costs (this is true because the total R&D cost also covers drug discovery and pre-clinical trial costs, and the total clinical trial cost is about 40% of the total R&D cost).

Although the above analysis of clinical trial supply chain spending is limited to three data points, it is readily apparent that manufacturing-related activities can be a significant source of costs during clinical trials. In the first two examples, roughly 40% of clinical trial spending is attributable to supplying the investigational drugs. Given our previous estimate of 37% of research and development spending being attributable to clinical trials, a potentially unjustifiable extrapolation of our data points suggests that 14.8% of the \$100 billion spent annually on R&D is for clinical trial supply chain activities. Our analysis of Allos Therapeutics shows that they spent 21.3% of their R&D on clinical trial supply chain activities. Thus, even the very rough estimate of 14.8% of pharmaceutical R&D being spent on the clinical trial supply chain has some confirmatory evidence towards the significant magnitude of spending devoted to the clinical trial supply chain.

Chapter 3

Production Planning Under Failure Risk

For every new drug that reaches a pharmacy's shelf, roughly 5,000 to 10,000 other potential medicines have failed to achieve commercialization (Pharmaceutical Research and Manufacturers of America 2007). For a pharmaceutical or bio-tech company attempting to create a new medicine or treatment, failure is not a surprise, but rather an event to be planned for. In this chapter, we analyze the impact of failure during clinical trials on the production-inventory decisions for investigational drugs and discover that an extension of the Wagner-Whitin model (Wagner and Whitin, 1958) can greatly improve efficiency in the clinical trial supply chain.

One of the most important hurdles prior to the U.S. Food and Drug Administration's (FDA) approval of a new drug is the testing of a drug candidate in clinical trials. Three phases of clinical trials are usually required to test both safety and efficacy of a potential treatment in human subjects. Typically, Phase I involves 50 to 100 healthy individuals, Phase II recruits a few hundred potential patients, and Phase III seeks to test the drug candidate in a few thousand patients. While we may know how many patients are needed in each phase of the clinical trial, there is an inherent uncertainty associated with each trial: the risk of failure. Indeed, only 21.5% of drug candidates entering clinical trials actually achieve FDA approval (DiMasi, Hansen, and Grabowski, 2003). Many of these drug candidates that fail to pass through the clinical trial hurdle are well

documented in the financial press. Below is just one example from the *New York Times* (Berenson, 2006):

The news came to Pfizer's chief scientist, Dr. John L. LaMattina, as he was showering at 7 a.m. Saturday: the company's most promising experimental drug, intended to treat heart disease, actually caused an increase in deaths and heart problems. Eighty-two people had died so far in a clinical trial, versus 51 people in the same trial who had not taken it.

Within hours, Pfizer, the world's largest drug maker, told more than 100 trial investigators to stop giving patients the drug, called torcetrapib. Shortly after 9 p.m. Saturday, Pfizer announced that it had pulled the plug on the medicine entirely, turning the company's nearly \$1 billion investment in it into a total loss.

The small success rate of clinical trials is painful to a pharmaceutical company's balance sheet because of the enormous amounts of time, labor, and materials required to perform a clinical trial. Annual supply chain spending for drugs under clinical trials can be substantial, e.g., accounting for 40% or more of the total clinical trial spending (see Chapter 2). For just one drug candidate, a company can spend millions of dollars every quarter to produce supplies for the clinical trial. When failure in a clinical trial occurs, every dollar spent on manufacturing, packaging, and distribution of unused clinical trial supplies is wasted and in most cases, unused material must be returned to a proper disposal facility for destruction (English and Ma, 2007). For example, Cotherix Inc., estimated \$126,000 in destruction costs for an obsolete drug that was valued at \$1.5 million (Cotherix Inc., 2006).

It would be unfair of us to label all post-failure drug supply as waste. Inventory is needed to ensure that as patients are recruited to participate in the study, drug supply is available. Any delays in this phase of testing become one less day of patent protection available to the drug. According to Clemento (1999), every extra day of patent availability is worth \$1 million for a typical drug. Since patient recruitment is the typical bottleneck in conducting clinical trials, shortages of clinical drug is considered an unacceptable delay and our model assumes no backlogging of demand. That being said, one would usually be economically foolish to produce enough supply to support all three phases of a clinical trial at once.

Production of investigational drugs is typically characterized by high costs (both fixed and variable) due to the low demand volume, low yield and the premature manufacturing process. In addition, at each step in the synthesis of the chemical compounds, rigorous quality control procedures are required to ensure that investigational drugs "are consistently produced and controlled to the quality standards appropriate to their intended use." (George, 2005) Often, active ingredient production for a drug candidate is a costly process and may require unique manufacturing equipment and processes. Thus, both the fixed and variable production costs tend to be much higher for investigational drugs than approved drugs which have been scaled up for mass production.

In the remainder of this chapter, we present a mathematical model for production planning to balance the two opposing forces of 1) high fixed production costs pushing for large lot sizes and 2) high failure costs which favor smaller lot sizes. High fixed costs for production, in the form of both time and money, lend support to producing large lot sizes. Alternatively, the high risk of failure, the high production variable cost and inventory carrying cost argue for smaller lot sizes. Smaller lot sizes would avoid wasting unused clinical drug supplies as well as the significant cost of destroying the unused material, but can result in high costs due to multiple production setups and more numerous quality control activities. We accommodate this environment by generalizing the Wagner-Whitin (W-W) model (Wagner and Whitin, 1958) to incorporate a stochastic component, namely, the risk of failure, we will refer to this model as the *failure-risk model*. We make the following contributions:

• Every failure-risk model is equivalent to a corresponding deterministic W-W

model if one adjusts the cost parameters properly to reflect failure risk and destruction costs, so many classic results of the W-W model still apply. Most interestingly, the planning horizon theorem indicates that in the failure-risk model, learning (e.g., the failure probability) as the clinical trial proceeds does not affect optimal supply decisions under certain conditions.

• We conduct a comprehensive numerical study using various environments that clinical trial manufacturers may face. We show that the failure-risk model can lead to substantial costs savings as compared to using the W-W model which ignores the risk of failure.

The remainder of this chapter is organized as follows. We review the related literature in §3.1. The model and analysis are presented in §3.2, and their extensions are discussed in §3.3. The potential benefits of properly accounting for failure are shown in an illustrative example in §3.4. A more thorough numerical study to test the effectiveness of the model under real-world scenarios is performed in §3.5. Finally, we summarize the results in §3.6.

3.1. Literature Review

Because of the interdisciplinary nature of this work, we shall first review literature that relates the disciplines of production planning and clinical research. Then, we highlight papers on dynamic economic lot size models and stochastic inventory models. Finally, we turn our attention to literature on research and development (R&D) supply chains.

Investigations of integrating drug supply with the clinical trial process are found in the medical and pharmaceutical literature. For example, George (2005) presents common issues encountered during clinical trial supply management and proposes coordination and flexibility as keys to success. A thorough description of clinical material manufacturing practices is provided by Bernstein and Hamrell (2000). In their paper, the authors advocate coordinating the disciplines of manufacturing and clinical programs to achieve efficient execution of drug development. Their study is conceptual and qualitative.

Quantitative research on production planning and capacity expansion under clinical trial uncertainty has been conducted in the chemical engineering literature. Gatica, Papageorgiou, and Shah (2003) simultaneously determines the optimal capacity and production decisions for multiple clinical trial drugs in different stages of their lifecycle. The underlying problem is a large-scale multi-stage stochastic program with integer and continuous variables and is solved as a mixed-integer linear program. Shah (2004) provides a recent survey for this line of research and an article by Colvin and Maravelias (2008) highlights more recent advances.

The work contained here differs from the previous work on clinical trial supply chains by its focus. We study a simpler model with one drug candidate and aim at deriving structural results which provide managerial insights and enable efficient solution algorithms. Thus, our work is more closely related to the dynamic economic lot size (DEL) models and stochastic inventory models studied in the operations management literature.

There is a long lasting interest and huge body of literature on DEL models for production-inventory systems with time-varying but known demand. Wagner and Whitin (1958) proposes the basic model (referred to as the *W-W model* hereafter). The paper characterizes several important system properties and develops a polynomial solution algorithm. Since then, many extensions and variations of the model have been studied. For more efficient solution algorithms, see AggarwalL and Park (1993), Federgruen and Tzur (1991), and Wagelmans, Van Hoesel, and Kolen (1992). For DEL models with various capacity constraints, see, e.g., Florian, Lenstra, and Rinnooy Kan (1980) and Shaw and Wagelmans (1998). For more general cost functions, see Eppen, Gould, and Pashigian (1969), Veinott Jr (1963), and Zangwill (1969). More recently, Chu, Hsu, and Shen (2005) study a lot sizing problem with general economies of scale cost functions. Realizing the problem is NP-hard, they develop approximation solutions and perform a worst-case analysis. Zipkin (2000) provides a thorough review of models and solution techniques on this topic.

In §3.2, we extend the classical W-W model to include the risk of failure. This feature transforms the W-W model into a stochastic production-inventory model. The most related stochastic inventory models to this work are those on single-stage systems with world-driven demand. Iglehart and Karlin (1962) analyzes optimal inventory ordering policies for non-stationary stochastic demand. Johnson and Thompson (1975) models demand as mixed autoregressive-moving average time series. Song and Zipkin (1993) and Sethi and Cheng (1997) characterize the optimal inventory control policies for various inventory systems with Markov-modulated demand. Comprehensive reviews are provided by Zipkin (2000) and Porteus (2002).

The failure-risk model studied here can be regarded as a special case of the models with Markov-modulated demand. Here demand in each period is a Bernoulli random variable, and if demand ever becomes zero, it stays zero for the rest of the planning horizon. While it is known that under certain regularity conditions, the state-dependent (s, S) policy is optimal for such systems with fixed production costs, the special structure of the demand process in a clinical trial supply chain allows us to develop much stronger results (e.g., equivalence to W-W model) and new insights (e.g., impact of failure risk).

The demand structure we use is similar to those analyzed in the inventory models with "sudden death obsolescence". Brown, Lu, and Wolfson (1964) introduces the model under periodic-review where demand may cease at an uncertain future date. A Bayesian procedure is employed to update demand distribution and a dynamic program is proposed to find the optimal solution. Pierskalla (1969) considers a model with stochastic demand and convex cost functions, and shows that the base-stock policy is optimal. Song and Zipkin (1996) generalizes the model to treat Markov-modulated demand. Katok, Lathrop, Tarantino, and Xu (2001) considers a model similar to ours but with random demand. To derive simple heuristic solutions, the authors analyzed their model with deterministic demand and found that it is a variant of the W-W model. Both our study and a similar study of obsolescence by Jain and Silver (1994) prove only the zero-inventory property for the deterministic model and derive heuristic solutions to the stochastic problem based on this property. Katok and Xu (2001) provides more details on the mathematic model and technical development which expand the Katok, Lathrop, Tarantino, and Xu (2001). While we study a similar model (with some differences on the cost structure) as the previous three papers, we takes the analysis of the deterministic demand case further by proving the full equivalence of production planning in a demand failure environment to a re-parameterized Wagner-Whitin model. We also leverage this equivalence to show conditions under which savings may be achieved. Lastly, a few authors have studied sudden death obsolescence models in continuous time with deterministic demand and developed EOQ types of solutions, see, David, Greenshtein, and Mehrez (1997) and references therein.

To overcome the complexities of existing stochastic obsolescence models, we study failure in the supply chain by focusing on a particular type of demand uncertainty that we term *demand failure*. Demand failure is defined as the sudden ceasing of a deterministic non-stationary demand stream. While the point of failure is not known, we do assume that failure probabilities in each period are known (DiMasi, 2001). By employing the assumption of demand failure, we are able to yield both clean and insightful results. As Song and Zipkin (1996) note in their study of obsolescence, which assumes a stochastic demand stream with random lifetime, clean results are not forthcoming in fully stochastic models:

Generally, we find that obsolescence does (or should) have a substantial impact in the way inventories are managed. The nature of these effects, moreover, is fairly intricate. It appears that obsolescence cannot be captured in a simpler model through parameter adjustments. Leveraging the deterministic demand assumption, we can formulate the failure-risk model into the simpler W-W model where the adjusted cost parameters incorporate the costs of failure. This result connects the failure-risk model with the vast literature of the W-W models, and thus, many results of the latter directly apply here. In addition, adjusting parameters of the W-W model is a simple way to include failure into production planning and thus, is more likely to be implemented than more complex obsolescence models. Lastly, we believe the demand failure assumption to be tenable to practitioners when they can accurately predict demand. According to a recent survey of clinical supply managers conducted by Bearing Point, 75% of Phase I and roughly 50% of Phases II-III SKU-level clinical supply forecasts are within 10% of actual demand (Kumar, 2008).

Our work is also related to the literature on R&D supply chains. Most of this literature focuses on supply chain design to support a product entering the market for the first time. However, much less attention has been devoted to the actual development supply chain (Krishnan and Ulrich, 2001; Pisano, 1997). At a pharmaceutical company, both the supply chain design for production ramp-up and the material supply during the development stage are important decisions. The focus of this study is on creating a model for the latter. More recently, there is a growing interest in combining R&D and supply chain decisions. Tomas and Hult (2003) provides a conceptual framework to analyze the interdependencies of product development and supply chain activities. Specific to the pharmaceutical world, Pisano (1997) presents strategic guidelines for effectively linking manufacturing strategy with the highly uncertain world of drug candidate development. Lastly, allocating scarce resources to an R&D pipeline of promising drug candidates is taken up as a portfolio problem (Girotra, Terwiesch, and Ulrich, 2007; Blau, Pekny, Varma, and Bunch, 2004).

3.2. The Model

Consider an investigational drug in a clinical trial over a finite time horizon with periods ranging from t = 1, 2, ..., N. We assume that demand is known for the drug in all periods (see justifications in §3.1). Demand and costs in each period are nonnegative. If the trial succeeds at the end of period t, we make production decisions and move to next period. Otherwise, we stop and all remaining inventory is wasted and is recycled or destroyed. The known demand must be satisfied and no backorders are allowed. Because the production cycle time is often much shorter than a clinical trial duration, we assume zero lead-time for production.

The system has the following state variables at the beginning of period t:

- I: inventory level.
- θ : system status indicator, success ($\theta = 1$), failure ($\theta = 0$).

The system has the following parameters,

- h_t : holding cost for inventory carried from period t to period t + 1.
- s_t : fixed production cost at period t if a production is initiated.
- α_t : failure probability at the end of period t.
- $\beta_t \equiv 1 \alpha_t$: success probability at the end of period t.
- d_t : demand in period t.
- c_t : production variable cost at period t.
- r_t : recycle/destruction cost at period t for any inventory un-used.

The estimates of failure probabilities in various therapeutic classes are readily available from the literature (Gatica, Papageorgiou, and Shah, 2003). It is possible that the failure probability of a trial does not depend on the results of previous trials if they are testing on different criteria, e.g., efficacy vs. safety. In this case, α_t is the unconditional probability of failure in the trial. It is also possible that the failure probabilities depend on the results of previous tests. For instance, during multiple trials for effectiveness, success in early trials can provide a strong indicator for success in on-going trials. In this case, α_t is effectively the failure probability conditioning on successes to date.

The action at period t is to produce $x_t \ge 0$. Let initial inventory level $I_0 = 0$. Define $f_t(\theta, I)$ to be the minimum expected cost for period t through N with initial inventory I and system status θ . Let $\delta(x_t)$ be the indicator function of $x_t > 0$, and $h_0 = 0$. The dynamic programming recursion can be written as follows,

$$f_t(0,I) = r_t I, \quad 1 \le t \le N \tag{3.1}$$

$$f_t(1,I) = \min_{\{x_t \ge 0, \ I+x_t \ge d_t\}} \{h_{t-1}I + \delta(x_t)s_t + c_t x_t + \alpha_t f_{t+1}(0, I + x_t - d_t) + \beta_t f_{t+1}(1, I + x_t - d_t)\}, \quad t = 1, 2, \dots, N-1 \quad (3.2)$$

$$f_N(1,I) = \min_{\{x_N \ge 0, \ I+x_N = d_N\}} \{h_{N-1}I + \delta(x_N)s_N + c_N x_N\}.$$
(3.3)

Combining Eqs. (3.1)-(3.2), and noting that $I + x_t - d_t$ is the inventory at the beginning of period t + 1, we can make the following transformation,

$$g_t(I) = \frac{\alpha_{t-1}r_t}{\beta_{t-1}}I + f_t(1,I), \forall t = 1, 2, \dots, N,$$
(3.4)

where $\alpha_0 = 0$. Then, $g_t(I)$ satisfies the following recursive equations,

$$g_{t}(I) = \min_{\{x_{t} \geq 0, \ I+x_{t} \geq d_{t}\}} \{ \frac{\alpha_{t-1}r_{t} + \beta_{t-1}h_{t-1}}{\beta_{t-1}}I + \delta(x_{t})s_{t} + c_{t}x_{t} + \beta_{t}g_{t+1}(I + x_{t} - d_{t})\},$$

$$t = 1, 2, \dots, N - 1$$

$$g_{N}(I) = \min_{\{x_{N} \geq 0, \ I+x_{N} = d_{N}\}} \{ \frac{\alpha_{N-1}r_{N} + \beta_{N-1}h_{N-1}}{\beta_{N-1}}I + \delta(x_{N})s_{N} + c_{N}x_{N}\}.$$

Note that this formulation is identical to the W-W model with modified inventory

holding cost and a time discount factor β_t at period t. One can adjust the cost parameters at each period, and by doing so, the dynamic program reduces to the Wagner-Whitin model with variable production costs. Let $h'_0 = 0$, and define the *effective* production costs and holding costs as follows,

$$s_{1}' = s_{1}$$

$$s_{t}' = s_{t} \cdot \prod_{j=1}^{t-1} \beta_{j}, \quad 1 < t \leq N$$

$$c_{1}' = c_{1}$$

$$c_{t}' = c_{t} \cdot \prod_{j=1}^{t-1} \beta_{j}, \quad 1 < t \leq N$$

$$h_{1}' = \alpha_{1} r_{2} + \beta_{1} h_{1}$$

$$h_{t}' = (\alpha_{t} r_{t+1} + \beta_{t} h_{t}) \cdot \prod_{j=1}^{t-1} \beta_{j}, \quad 1 < t < N.$$

Hence,

$$g'_{t}(I) = \min_{\{x_{t} \ge 0, \ I+x_{t} \ge d_{t}\}} \{h'_{t-1}I + \delta(x_{t})s'_{t} + c'_{t}x_{t} + g'_{t+1}(I + x_{t} - d_{t})\}, \ t = 1, 2, \dots, N - (3.5)$$

$$g'_{N}(I) = \min_{\{x_{N} \ge 0, \ I+x_{N} = d_{N}\}} \{h'_{N-1}I + \delta(x_{N})s'_{N} + c'_{N}x_{N}\}.$$

$$(3.6)$$

Eqs. (3.5)-(3.6) show that one can transform the stochastic failure-risk model to an equivalent deterministic W-W model with properly adjusted production and inventory holding costs. Note that the adjusted (or effective) inventory holding cost is the weighted average of the destruction cost and the regular inventory holding cost which is discounted by the success probabilities to date.

Because all cost parameters defined in Eqs. (3.5)-(3.6) are nonnegative, by Zipkin (2000, §4.3.3), the "zero-inventory property" holds. Specifically, let I_t be initial inventory level at period t, and we can formally state the "zero-inventory property".

Theorem 1 (The Zero Inventory Property) For the dynamic program defined in Eqs. (3.1)-(3.3), the following claims hold.

- 1. For each period t, $I_t \cdot x_t = 0$.
- 2. $x_t = 0$ or $x_t = \sum_{j=t}^k d_j$.
- 3. If d_t is satisfied by some x_{τ} for $\tau < t$, then d_j , $j = \tau + 1, \ldots, t 1$ is also satisfied by x_{τ} .
- 4. Given that $I_t = 0$ for period t, it is optimal to consider periods 1 through t 1independent of other periods.

For ease of analysis, we further transform the dynamic program into the W-W model without variable production costs. Note that $c'_t x_t = c'_t (I + x_t - d_t) - c'_t (I - d_t)$ for t = 1, 2, ..., N.

$$g'_{t}(I) = \min_{\{x_{t} \ge 0, \ I+x_{t} \ge d_{t}\}} \{(h'_{t-1} - c'_{t})I + \delta(x_{t})s'_{t} + c'_{t}(I + x_{t} - d_{t}) + c'_{t}d_{t} + g'_{t+1}(I + x_{t} - d_{t})\},$$

$$t = 1, 2, \dots, N - 1$$

$$g'_{N}(I) = \min_{\{x_{N} \ge 0, \ I+x_{N} = d_{N}\}} \{(h'_{N-1} - c'_{N})I + \delta(x_{N})s'_{N} + c'_{N}d_{N}\}.$$

To remove the constants $c_t d_t$ and combine terms which are functions of $I + x_t - d_t$, we define,

$$G_t(I) = c'_{t-1}I + g'_t(I) - [c'_t d_t + \sum_{n=t+1}^N (c'_n d_n \cdot \prod_{j=t}^{n-1} \beta_j)], \quad t = 1, 2, \dots, N-1$$

$$G_N(I) = c'_{N-1}I + g'_N(I) - c'_N d_N,$$

where $c'_0 = 0$. The recursion for G_t is as follows,

$$G_t(I) = \min_{\{x_t \ge 0, \ I+x_t \ge d_t\}} \{H_{t-1}I + \delta(x_t)S_t + G_{t+1}(I + x_t - d_t)\}, \quad t = 1, 2, \dots, N(3.7)$$

$$G_N(I) = \min_{\{x_N \ge 0, \ I+x_N = d_N\}} \{ H_{N-1}I + \delta(x_N)S_N \},$$
(3.8)

where

$$S_t = s'_t, \quad 1 \le t \le N$$

$$H_1 = c_1 - c_2 + \alpha_1(c_2 + r_2) + \beta_1 h_1$$
(3.9)

$$H_t = [c_t - c_{t+1} + \alpha_t (c_{t+1} + r_{t+1}) + \beta_t h_t] \cdot \prod_{j=1}^{t-1} \beta_j, \quad 1 < t < N.$$
(3.10)

Note that H_t consists of two parts: the first part is the difference between production costs in two successive periods; the second part is the weighted average of the total loss due to failure (including the production and destruction costs, referred to as the *failure cost*) and the regular inventory holding cost.

Define F(j, i) to be the minimum cost to cover all demands in periods j, j + 1, ..., iwith $I_j = 0$ and $I_{i+1} = 0$ if $j \le i$; let F(j, i) be zero otherwise. The forward formulation to compute F(j, i) is as follows.

$$F(j,i) = \min\{\min_{j \le k < i} \{S_k + \sum_{n=k}^{i-1} H_n \sum_{l=n+1}^{i} d_l + F(j,k-1)\}, S_i + F(j,i-1)\}, \quad j < i. (3.11)$$

The backward formulation works as follows,

$$F(j,i) = \min\{S_j + F(j+1,i), \min_{j < k \le i} \{S_j + \sum_{n=j}^{k-1} H_n \sum_{l=n+1}^k d_l + F(k+1,i)\}\}, \quad j < i. (3.12)$$

To compute the optimal solution and optimal cost functions, one can use the well known algorithms of Wagner and Whitin (1958), Federgruen and Tzur (1991) and Wagelmans, Van Hoesel, and Kolen (1992).

Because the failure probability α_t only affects the holding costs H_j for $j \ge t$, it follows from the forward formulation, Eq. (3.11), that the Planning horizon Theorem of Wagner-Whitin can be applied and interpreted in our model as follows.

Theorem 2 (The Planning Horizon Theorem) If $H_t \ge 0$ for all $1 \le t < N$, then

- If the optimal solution for F(1,t) in Eq. (3.11) is t* ≤ t, then to solve F(1,τ) with τ > t, one only needs to consider F(t*,τ). In other words, if it is optimal to incur a set-up cost at period t* when periods 1 through t are considered alone, then it is optimal to incur a set-up cost at period t* in any τ-period model.
- 2. The optimal solution for periods 1 to t^* does not change even if we can update α_j for $j \ge t^*$ along with time.

If $H_t < 0$, Theorem 2 may not hold, see Eppen, Gould, and Pashigian (1969) for more discussion. Due to the high destruction cost and failure risk, H_t will be positive in clinical trial supply chains. Hence, we assume $H_t \ge 0$ for all $1 \le t < N$ for the remainder of this chapter.

One question that could be asked is, does there exist a threshold on α_k , the failure probability, so that above which, it is optimal to produce in each period? The following Theorem answers this question.

Theorem 3 (The High Failure Risk Property) If $\beta_j < (c_j + r_{j+1})/[s_{j+1}/d_{j+1} + c_{j+1} + r_{j+1} - h_j]$ for all j = 1, 2, ..., N - 1, then it is optimal to produce in each period from 1 to N.

Proof. By Theorem 2, it suffices to consider F(j, j + 1) for j = 1, 2, ..., N - 1.

$$F(j, j+1) = \min\{S_j + F(j+1, j+1), S_j + H_j \cdot d_{j+1}\}$$
(3.13)

$$= \min\{S_j + S_{j+1}, S_j + H_j \cdot d_{j+1}\}.$$
(3.14)

Clearly, if $S_{j+1} < H_j \cdot d_{j+1}$, then it is optimal to produce in both periods j and j + 1. Simple derivation shows that the condition $S_{j+1} < H_j \cdot d_{j+1}$ is equivalent to $\beta_j < (c_j + r_{j+1})/[s_{j+1}/d_{j+1} + c_{j+1} + r_{j+1} - h_j]$.

To interpret Theorem 3, let $c_j = c$ and $r_j = r$ for all j. If $s_{j+1}/d_{j+1} < h_j$ for all j, it is optimal to produce at each period even if $\alpha_j = 0$ for all j. Otherwise, if $s_{j+1}/d_{j+1} > h_j$ for all j, then the condition reduces to $\beta_j < 1/[(s_{j+1}/d_{j+1} - h_j)/(c+r) + 1]$. Clearly, if the production cost, the recycle cost or the demand quantity increases, the likelihood of producing in each period increases.

Finally, we study the impact of failure risk on the optimal expected total cost, C^* .

$$C^* = \sum_{t=1}^{N-1} h'_t \cdot I^*_{t+1} + \sum_{t=1}^N \delta(x^*_t) \cdot s'_t + \sum_{t=1}^N c'_t x^*_t, \qquad (3.15)$$

where x_t^* and I_t^* are the optimal production and inventory decisions.

Proposition 1 C^* is a piecewise linear concave function for each α_t , t = 1, 2, ..., N. In addition, $C^*(\alpha_t^2) \leq (1 + \alpha_t)C^*(\alpha_t)$ for each t.

Proof. The proof of the first statement follows that of Zipkin (2000) problem 4.6. Briefly, for any fixed sequence of production times, we note that h'_j , s'_j and c'_j are either independent of α_t or linear functions of α_t . Therefore, given the sequence of production times, C^* is linear in every α_t . C^* is concave in α_t because C^* is the minimum cost over all possible sequence of production times.

To prove the second statement, we consider α_1 as a special case. For α_1^2 , $s_1' = s_1$ and $c_1' = c_1$,

$$s'_{t} = s_{t} \cdot \prod_{j=1}^{t-1} \beta_{j} (1 + \alpha_{1}), \ t > 1$$

$$c'_{t} = c_{t} \cdot \prod_{j=1}^{t-1} \beta_{j} (1 + \alpha_{1}), \ t > 1$$

$$h'_{1} = \alpha_{1}^{2} (r_{2} - h_{1}) + h_{1} \leq [\alpha_{1} (r_{2} - h_{1}) + h_{1}] (1 + \alpha_{1})$$

$$h'_{t} = (\alpha_{t} r_{t+1} + \beta_{t} h_{t}) \cdot \prod_{j=1}^{t-1} \beta_{j} (1 + \alpha_{1}), \ t > 1.$$

Suppose that the optimal sequence of production times remains the same for both α_1 and α_1^2 . Then $C^*(\alpha_t^2) \leq (1 + \alpha_t)C^*(\alpha_t)$. Otherwise, the same inequality also holds because $C^*(\alpha_t^2)$ becomes even smaller. The same proof applies to all α_t for t > 1. \Box

 $C^*(\overline{\alpha})$ is generally not a monotonic function of α_t . Consider the special case of $\alpha_t \to 1$ for all t. C^* effectively reduces to a single-period cost function, which is clearly

less than the multi-period cost function as $\alpha_t \to 0$ for all t. Proposition 1 gives a upper bound on the diminishing rate for C^* as α_t increases. On the other hand, if the optimal sequence of production times is to produce at period N-1 to cover demand in both N-1 and N, then $C^*(\alpha_{N-1} + \Delta) > C^*(\alpha_{N-1})$ can hold for sufficiently small Δ if $r_N > h_{N-1}$.

3.3. Extensions

In this section, we consider two extensions of the model in §3.2 to incorporate real-world situations: general concave cost functions and production/storage constraints.

3.3.1 Concave Cost Functions

Let $c_t(x)$ be the production cost function, $h_t(I)$ be the inventory cost function, and $r_t(I)$ be the destruction/recycle cost function. In line with economies of scale, we assume that $c_t(x)$, $h_t(I)$ and $r_t(I)$ are concave and increasing.

Under these cost functions, the dynamic program recursion, Eqs. (3.1)-(3.3), can be written as follows,

$$f_{t}(0, I) = r_{t}(I), \quad 1 \le t \le N$$

$$f_{t}(1, I) = \min_{\{x_{t} \ge 0, \ I+x_{t} \ge d_{t}\}} \{h_{t-1}(I) + c_{t}(x_{t}) + \alpha_{t} f_{t+1}(0, I + x_{t} - d_{t}) + \beta_{t} f_{t+1}(1, I + x_{t} - d_{t})\}, \quad t < N$$

$$(3.1)$$

$$f_N(1,I) = \min_{\{x_N \ge 0, \ I+x_N = d_N\}} \{h_{N-1}(I) + c_N(x_N)\}.$$
(3.3)

Similar to §3.2, define $g_t(I) = \frac{\alpha_{t-1}}{\beta_{t-1}} r_t(I) + f_t(1, I)$ for $t = 1, 2, \dots, N$, where $\alpha_0 = 0$. Then

$$g_{t}(I) = \min_{\{x_{t} \ge 0, \ I+x_{t} \ge d_{t}\}} \{ \frac{\alpha_{t-1}}{\beta_{t-1}} r_{t}(I) + h_{t-1}(I) + c_{t}(x_{t}) + \beta_{t}g_{t+1}(I + x_{t} - d_{t}) \}, \ t < \mathfrak{(S.4)}$$

$$g_{N}(I) = \min_{\{x_{N} \ge 0, \ I+x_{N} = d_{N}\}} \{ \frac{\alpha_{N-1}}{\beta_{N-1}} r_{N}(I) + h_{N-1}(I) + c_{N}(x_{N}) \}.$$

$$(3.5)$$
Discounting cost functions in each period by β_t , we define,

$$\begin{aligned} c_1'(x) &= c_1(x) \\ c_t'(x) &= c_t(x) \cdot \prod_{j=1}^{t-1} \beta_j, \quad t > 1 \\ h_1'(I) &= \alpha_1 r_2(I) + \beta_1 h_1(I) \\ h_t'(I) &= [\alpha_t r_{t+1}(I) + \beta_t h_t(I)] \cdot \prod_{j=1}^{t-1} \beta_j, \quad 1 < t < N \end{aligned}$$

Finally,

$$g'_t(I) = \min_{\{x_t \ge 0, \ I+x_t \ge d_t\}} \{h'_{t-1}(I) + c'_t(x_t) + g'_{t+1}(I+x_t - d_t)\}, \quad t < N$$
(3.6)

$$g'_{N}(I) = \min_{\{x_{N} \ge 0, \ I+x_{N}=d_{N}\}} \{h'_{N-1}(I) + c'_{N}(x_{N})\}.$$
(3.7)

Note that the *effective* cost functions, $h'_{t-1}(I)$ and $c'_t(x)$, are still concave and increasing. By Eqs. (3.6)-(3.7), the stochastic failure-risk model is equivalent to the deterministic W-W model with general concave and increasing cost functions. By Zipkin (2000, Sections 4.3.3 and 4.4.6), The "Zero Inventory Property" (Theorem 1) holds here. One can derive the forward and backward formulations in a similar way as Eqs. (3.11)-(3.12), for brevity, we omit the details. As Veinott Jr (1963) and AggarwalL and Park (1993) point out, the W-W model with general concave cost functions can be solved using the forward formulation with complexity $O(N^2)$. However, Theorem 2 does not hold because of the general form of the concave production cost functions.

3.3.2 Additional Constraints

We discuss three types of constraints: the production capacity constraint, the inventory shelf-life constraint and the storage capacity constraint. The inventory shelf-life constraint specifies the number of periods that a unit can be carried in inventory, which limits the number of future periods that can be covered by a production batch. Thus, it is effectively a production capacity constraint. It is easily seen that with any subset of these constraints, one can use the same technique in §3.2 to reduce the stochastic failure-risk model to an equivalent deterministic W-W model with adjusted cost parameters and the same set of constraints. For the W-W model with production capacity, inventory shelf-life and storage capacity constraints, one can find the solution using well established algorithms, see, e.g., Shaw and Wagelmans (1998).

3.4. An Illustrative Example

In this section, we demonstrate that accounting for demand failure when planning a production schedule can lead to substantial cost savings over using the Wagner-Whitin model ignoring the failure risk. To develop insight, we consider a special case of Phase II clinical trials with stationary data where $c_t = \$75$, $r_t = \$25$, $h_t = \$5$, $d_t = 250$ and $s_t = \$50,000 \ (\forall t \in 1, 2, ..., 12)$. Note that both h_t and d_t are defined per period where a period equals two months here. We consider a 12-period (two years) planning horizon and a 7% probability of failure in each period (i.e. approximately a 42% chance of success for phase II trials). Further justification of the parameters chosen here is provided in the next section.

For stationary production variable costs, it makes sense to utilize Eqs. (3.7)-(3.8), which ignore the production variable costs except as included in the failure costs. We first consider the classic W-W model ignoring the risk of failure. Figure 3.1a shows the production-inventory costs, excluding the risk of failure, as a function of production schedule. From Figure 3.1a, we see that satisfying 12 periods of demand with just one production run, is both the optimal plan as calculated by the W-W model and also represents a typical heuristic of pharmaceutical manufacturers (Shah, 2004).

Figure 3.1b shows the same example, but now the costs account for the 7% probability of failure in each period. The optimal plan is now to produce batches which satisfy six periods of demand. This plan, which minimizes total expected costs, calls for two batches during the planning horizon as compared with the one batch that is



Figure 3.1: Production Costs Excluding & Including the Risk of Failure

prescribed by the W-W model. From Figure 3.1b, one easily sees that using the plan of satisfying all 12 periods of demand with one production run as prescribed by the W-W model would lead to very high failure costs. This is a direct consequence from carrying large amounts of inventory that will likely be wasted due to demand failure. In fact, the optimal schedule generated by the failure-risk model is expected to be 28% less costly than the optimal plan of the W-W model. Even with the high fixed costs of this example, failure costs lead to reducing the optimal lot size and scheduling more frequent production runs.

3.5. Numerical Study of Potential Savings

In this section, we conduct a comprehensive numerical study to gauge the potential savings of incorporating failure risk into production planning by solving various environments that clinical trial manufacturers may face. From our discussions with industry professionals, most clinical supply managers plan for success despite knowing that failure is both likely and costly. Our objective is to quantify the savings and identify conditions under which the savings of incorporating failure into production planning are likely to be substantial. We vary our key parameters, namely, production costs, holding costs, and failure probability based on our observations of the industry. Please note that we do not explicitly model destruction costs because including them in production costs is mathematically equivalent when production variable costs are constant (see Eqs. 3.7-3.10).

In our estimates of production costs, we estimate the variable cost of the active ingredient, the cost of packaging, the cost of distribution and tracking, and the cost of destruction as a percentage of the fixed set-up cost of production. In addition, the cost of manufacturing a placebo drug must also be included in the cost of each treatment. As all of the above costs can vary widely based upon the production of active ingredient, packaging requirements, formulation requirements, and the dosing schedule, we employ a wide range of production costs. To get the magnitude of our estimates, we look to vaccine production, which, like clinical trial production, is less standardized and less predictable than typical commercial drug production (Institute of Medicine 2004). As a rough proxy for clinical trial production, we expect the total variable cost to fixed set-up cost ratio to be 3:5.¹ Assuming that we obtain this ratio when we are producing 10³ treatments per batch, we get variable production costs per treatment of approximately 0.06% of the setup cost per batch. Using this estimate as a reference, we employ a wide range for simulating variable production costs per treatment of between 0.01% and 1.25% of the fixed set-up costs.

Holding costs in the clinical trial supply chain are most likely higher than incurred in typical pharmaceutical supply chains due to costly tracking and auditing of inventory levels. In addition, bio-tech molecules often require controlled storage environments which may add to the cost. For a lower bound on annual holding costs, we take the conservative estimate of DiMasi, Hansen, and Grabowski (2003) of 11% as the pharmaceutical industry's real cost-of-capital for money tied up in inventory. We choose an upper bound of 80% which may better reflect the potentially high costs of storing and tracking each treatment and the corresponding placebo.

Failure probability in drug development is well documented in the literature. If we ignore Phase I trials because of the relatively small drug supply that is required,

¹From the 2002 Mercer Management study we see that variable costs are 15% of total production costs and batch costs are 25% of total production costs. Thus, a 3:5 ratio seems appropriate.

| Parameter | Average Value | Lower Bound | Upper Bound | |
|-------------------------------|------------------|-------------|-------------|--|
| Production Cost Per Treatment | 0.62% | 0.01% | 1.25% | |
| (as $\%$ of Set-Up Cost) | 0.0370 | 0.0170 | | |
| Annual Holding Cost | 45 50% | 110% | 800% | |
| (as $\%$ of Prod. Cost) | 40.070 | 11/0 | 8070 | |
| Phase II Failure Probability | 59% | 45% | 73% | |
| Phase III Failure Probability | 21.5% | 13% | 30% | |
| Duration of Phase II Trial | 2 Years | — | _ | |
| Duration of Phase III Trial | 3 Years | — | _ | |
| Annual Phase II Demand | 150 Treatments | — | — | |
| Annual Phase III Demand | 1,200 Treatments | — | _ | |
| Planning Period | Two Months | _ | _ | |

Table 3.1: Parameters for Phase II and Phase III Simulations

we apply our analysis to the failure probability that is present in Phase II and Phase III of clinical trials. Typically, failure is mostly likely to occur in Phase II where the phase II attrition rate (i.e. probability of failure) is around 58.8%. Phase III performs better with an average failure rate of 21.5% (DiMasi, 2001). Based on these numbers and noting that the average length of Phase II and Phase III trials for new chemical entities are 26.0 months and 33.8 months, respectively (DiMasi and Grabowski, 2007), or roughly 2 and 3 years respectively, we perform our parametric study around these industry averages as shown in Table 3.1.

The Phase II Simulation: We randomly generated 1,000 scenarios for Phase II clinical trails with parameters being uniformly distributed over the range of possible values. For each scenario, two optimal production plans were generated: 1) a Wagner-Whitin production plan and 2) a failure-risk production plan. Since our goal is to understand under what circumstances the failure-risk model is likely to outperform the Wagner-Whitin model, we calculate, for each scenario, the percentage cost reduction by the failure-risk model relative to the Wagner-Whitin model. We then plot the percentage savings against various system parameters to gain insight. The results are shown in the four graphs in Figure 3.2 with each diamond on the graphs representing one of the 1,000 scenarios. In 419 of the scenarios, the failure-risk model led to savings with an average of 11%.

Since the traditional Wagner-Whitin model, like current industry practice, does not incorporate demand failure, it is quite intuitive that an increase in the likelihood of failure would reduce the W-W model's effectiveness. As shown in Figure 3.2a, the failure-risk model's maximum potential savings over the W-W model does increase in failure probability. Nonetheless, the presence of many scenarios with 0% savings at all failure probabilities demonstrates that even high failure probabilities do not always lead to different solutions by the two planning models. The solutions also depend on other system parameters, such as production costs and inventory holding cost.

Figure 3.2b shows the impact of the variable production cost to fixed setup cost ratio in our Phase II simulation. A threshold appears where the variable production costs, which in our model include destruction costs, need to be sufficiently high for savings to be realized. In particular, to achieve savings in our test scenarios, variable production costs must exceed 0.39% of the fixed set-up costs. This is true because when variable production costs are low, the costs of failure in the form of wasted production are also low. As the variable to fixed cost ratio increases beyond the threshold, the increased cost of failure leads to an increase in expected magnitude of savings, because the production frequency under the failure-risk model increases faster than that of the W-W model.

The impact of holding costs are studied in Figure 3.2c. It is interesting to note that in our test scenarios, to achieve any savings, the ratio of holding costs to set-up costs must be within a range of 0.017% to 0.111%. To see what is behind this observation, we present the production frequency information in Figure 3.2d. On the secondary vertical axis, we present the planned number of setups that each model recommends for each of the 1,000 scenarios. Combining the information of production frequency and the percentage savings, we make the following observations about this Phase II simulation:

• Both models call for either one or two setups during the 2-year planning horizon.



Figure 3.2: Expected Reduction in Costs Using F-R Algorithm in Phase II Simulations

- The failure-risk model is most beneficial when holding costs are not too low. When holding costs are too low, both the failure-risk and W-W models will both call for producing just one big lot.
- The failure-risk model is most beneficial when holding costs are not too high. When holding costs are too high, both models have the same plan of producing two times over the planning horizon.
- The range of holding to fixed cost ratio observed above (where savings are possible) approximately corresponds to the range within which the failure-risk model plans for two setups while the Wagner-Whitin model calls for one setup. In this range of scenarios, the potential benefits of the failure-risk model increase with increasing holding costs. This is a simple reflection of the larger average inventory level, some of which is more likely to be wasted, under the plan of the W-W model.

Phase III Simulation: Phase III clinical trials differ from Phase II trials due to their longer duration, higher demand for treatments, but lower probability of failure. Our Phase III simulation adjusts these parameters accordingly. One result of these changes is that out of the 1,000 scenarios investigated, savings were achieved in 55.4% of the trials. This is more frequent than the 41.9% frequency in which savings were achieved in the Phase II simulation. However, Phase III savings, when they occurred, averaged about 2.85% which is significantly less than the average savings of 11.0% observed in Phase II. The first difference is due to the longer planning horizon, because with more periods, it is more likely that the two models yield different solutions. The second difference is largely due to the lower failure rate.

Similar to our Phase II investigation, we look to a graphical representation of the percentage savings against certain key parameters of the model. These graphs are shown in Figure 3.3. As seen in Phase II, the failure-risk model in Phase III has larger potential benefits as the probability of failure increases (see Figure 3.3a). However, in contrast to our Phase II results, Phase III has some noteworthy differences in regards



Figure 3.3: Expected Reduction in Costs Using F-R Algorithm in Phase III Simulations

to the potential for savings against other parameters.

First, as shown in Figure 3.3b, the larger demand for treatments and longer planning horizon substantially reduce the minimum production cost threshold. We see that savings are achievable at almost any level of production cost. We also see a saw-blade pattern in the diagram of these costs which is best explained by the data shown in Table 3.2. We see that the maximum potential savings are achieved when the Failure Risk (F-R) algorithm plans two setups while the W-W model plans only one setup. While production costs are not directly responsible for the drop in savings that we see when production costs reach about 0.47% of setup costs, they directly affect holding costs, which have been defined as a percentage of production costs. At production costs of about 0.47%, holding costs are driven sufficiently high such that the W-W model will perform a minimum of at least two setups over the planning horizon. As seen in Table 3.2, once the W-W model plans for more than one setup, the maximum potential

| Planned Number of Setups Using W-W Algorithm | Planned Number of Setups Using F-R Algorithm | Maximum % Sav- ings Observed | |
|--|--|---------------------------------|--|
| 1 | 2 | 26.7% | |
| 2 | 3 | 12.1% | |
| | 4 | 16.7% | |
| 3 | 4 | 5.5% | |
| | 4 | 0.2% | |
| 4 | 5 | 3.1% | |
| | 6 | 4.5% | |
| 5 | 5 | 0.1% | |
| ~ | 6 | 2.3% | |

Table 3.2: Maximum Observed % Savings Versus Number of Setups During Phase III Simulation

savings drops significantly.

In the Phase II simulation, we saw that holding costs were required to be within a certain range for savings to occur. From our Phase III chart of holding costs (Figure 3.3c), we see several ranges of holding costs that have different effects on the savings achieved. In Figure 3.3d, we overlay the frequency of production that the two models call for on the second vertical axis. As we increase the holding costs, we notice a transition in the failure-risk production plans from one of less frequent setups to one of more frequent setups. This transition is then followed by a similar transition of the W-W plan to one of more frequent setups. As shown in both Table 3.2 and Figure 3.3d, savings are achieved every time the failure-risk model makes the jump to a production plan that has more frequent setups to the same schedule that the failure-risk model calls for. **Industry Notes**: Mapping the results of our analysis to industry, we expect the failurerisk model to have the most significant impact for drugs that have a high probability of failure, sufficiently high production costs and relatively low inventory holding costs. Since it is hard to characterize holding and production costs for a certain clinical trial environment, we comment only on the probability of failure that is seen during clinical trials. In pharmaceutical and bio-tech industries, we see below-average success probabilities for drugs in the following therapeutic classes: antineoplastic, cardiovascular, central nervous system, immunologic, and respiratory medicines (DiMasi, 2001; DiMasi, Grabowski, and Vernon, 2004).

3.6. Summary Remarks

This chapter applies operations management models to clinical trial drug supply chains and demonstrates their potential impact. Specifically, we consider a class of dynamic economic lot size models under the risk of demand failure – the failure-risk models. We show that the stochastic failure-risk models can be transformed to corresponding W-W models where only the cost parameters need to be adjusted according to the failure risk and destruction cost. Therefore, many of the classic results for W-W models directly apply here. Most interestingly, the planning horizon theorem (Theorem 2) indicates that learning during clinical trials does not affect supply decisions under certain conditions. Our numerical study (based on our observation of the industry) reveals that while the failure-risk model does not always call for a production plan different from the W-W model, certain combinations of holding, production, and setup costs lead to substantive savings.

The model and insights developed in this research indicate ways to improve the current practice of clinical trial supply chains. Often, pharmaceutical/bio-tech companies employ different teams to plan for clinical trial activities and clinical drug supplies, where each team reports to its own Vice President. There is little connection between the teams beyond the direct supply and demand relationship, and the supply team typically plans for success (i.e., ignores failure in planning). This study shows that proper communication between the teams about the failure probabilities and properly accounting for failures in drug supplies can help the supply team substantially reduce drug manufacturing cost without harming service. We should point out that the models and results developed here also apply to other business practices where demand may cease to exist at a uncertain future date.

Chapter 4

Adaptive Clinical Trials: Balancing Learning and Economies of Scale

Before an investigational drug can obtain FDA approval, the drug must be proven both safe and effective in humans. Unfortunately, only one in five drugs that enter this human testing hurdle of the FDA approval process, also known as clinical trials, actually obtains approval (DiMasi, 2003). This failure to obtain approval does not always mean that the drug is truly unsafe or ineffective; it may simply be the test was not setup for success. Possibly, a different dosage or a different length of treatment may have led to a better outcome for the clinical trial. In an effort to achieve more successes, the pharmaceutical industry is moving towards adaptive clinical trials which allow for flexibility in how a trial is conducted (Lesko, 2007). For example, an adaptive approach allows for different dosages during the same trial and to weed out the dosages that prove toxic or non-therapeutic as trial data is collected and analyzed. This is in stark contrast to the traditional, more rigid approach to clinical trials where a dosage, a patient population, a length of treatment, and specific measures of success are chosen all prior to commencement of the trial.

Of course, the purely rigid approach is a supply manager's dream because demand is more certain. However, in an adaptive trial, more uncertainty about demand is introduced. For example, if during an adaptive trial the recommended dosage of an investigational treatment goes from 10mg to 20mg, the demand has just been effectively doubled. As a result, we see that the increased probability of a successful trial for FDA approval comes at the expense of the supply chain. Many of the key parameters used for forecasting supply are now subject to change during the trial.

To ensure supply for the start of a trial, clinical supply managers are forced to "forecast based on a limited number of variables before the cost of the supply chain overshadows the risk of excluded variables impacting supplies." (Hall, 2008). Given more time, all of the variables impacting clinical supply requirements become less fluid and a better forecast of demand can be made. Given the need to start clinical production before demand is known leads to two possible strategies. If we assume a sufficient shelf life of the investigational drug, one supply strategy is to produce enough material in the initial production run to accommodate any potential scenario for clinical trial demand. Alternatively, a supply manager can break production into two production runs. The first production run is made to ensure enough supply is available to start the trial. The second production run ensures enough supply to end the trial and is made with greater precision in forecasted demand. For example, a clinical trial, with globally distributed testing centers, will start with enough supply to satisfy a launch of the trial in the United States. Once launched in the United States, international testing centers are opened and most demand for these sites will be satisfied using supply from a second production run.

Supply from a second production run is committed to with greater certainty as to dosage and study enrollment parameters. A trial that begins with a range of possible dosages, also called treatment arms, also begins with high demand uncertainty. However, as the trial continues patient allocation to dosages becomes concentrated on the most promising of the initial dosages (see e.g. Krams, Lees, Hacke, Grieve, Orgogozo, and Ford, 2003). Thus, the longer the trial continues, the longer the supply manager learns about demand and the more certain he can be about the remaining supply requirements.

Unfortunately, reducing the costs of supply-demand mismatches can only be realized

by increasing the number of production runs. The additional runs, while costly, can be planned for using a more accurate forecast of trial demand. By incurring multiple setups, the high fixed costs of physical drug production and the associated quality control activities are amortized over less inventory and economies of scale are sacrificed. The essence of this research is to understand this sacrifice of economies of scale and to optimally balance the benefits of learning against higher costs for physical supply.

In this research, we construct a model to plan drug supply for adaptive clinical trials. While minimizing inventory costs, the model explicitly incorporates the learning that takes place during a clinical trial. At the beginning of a clinical trial, both dosage and patient recruitment requirements are very uncertain. As a trial is underway and continues, the ideal dosage and the desired patient enrollment are zeroed in and a more accurate forecast can be made. Various rates of demand learning are considered through the use of learning curves that are a function of time. In addition to demand learning, the model incorporates setup costs (i.e., fixed costs in production) for both pre-trial and intra-trial production. Thus, the model enables us to weigh the benefit of demand uncertainty reduction against the cost of incurring a second setup cost for production. The objective is to study the impact of fixed costs and demand learning rates on the value and the optimal timing of demand learning.

We organize the rest of this chapter as follows. We review the related literature in §4.1. The model and analysis are presented in §4.2. A numerical study is presented in §4.3. Finally, we summarize the study in §4.4.

4.1. Literature Review

While our model is motivated by challenges of the pharmaceutical industry, it is both applicable and similar to the problems faced by production planners of many short lifecycle products with uncertain demand. In one of the most cited works on production planning for short lifecycle products, Fisher and Raman (1996) pioneered an approach to reduce stock-out and markdown costs for a set of seasonal fashion products based on the idea that sufficiently short manufacturing lead times allow for a portion of production to occur after some initial demand is observed. This initial demand could then be used to better forecast actual demand for a set of new products. They described this logic as follows:

The dramatic improvement in forecast accuracy after observing only 20% of initial demand suggests a strategy for reducing the cost of too much or too little inventory: commit to a modest amount of initial inventory for each product, observe initial demand, and then produce an additional amount of each product based on improved forecasts.

Consistent with the above logic, the authors present a two-stage model that incorporates second stage forecast uncertainty reduction, which we will also refer to as *demand learning*, through the correlation of demand between observed first stage demand and total expected demand. Through this correlation, the authors captured how the second-stage production enables a more accurate match of supply and demand. As a result of decreased supply-demand mismatch costs, it is shown that use of the authors' methodology at Sport Obermeyer can potentially quadruple profits.

As our work in this chapter is highly motivated by the vein of research in Fisher and Raman's seminal work, we highlight three key distinctions between our model and theirs. First, we allow the second production run to occur at different time points and thus, the scheduling of the second production run is also a decision variable. Second, we include setup costs directly whereas their work includes minimum lot size requirements as a proxy for these costs. Lastly, we study the impact of learning rate on the system performance by incorporate learning curves in the demand learning process. This enable us to connect the amount of learning with the amount of time we have available to learn. Combined, these three distinctions allow us to weigh the benefits of demand learning against the fixed costs of introducing a second production run. If we abstract from adaptive clinical trials, the general idea we seek to model is that forecasts and planning decisions can be improved based on observation of a full or partial season's demand. Within this abstraction, we can find papers that address a similar problem to ours. For example, Parlar and Weng (1997) research a two-period production decision where a second production run is possible. However, this second production run occurs after the realization of demand and thus, full learning has occurred. As an example of partial learning, Eppen and Iyer (1997) study a large catalog retailer's decisions of how much of a women's fashion product to order and then how much to divert to outlet stores upon observation of a portion of demand. They continue a long-line of studies that utilize bayesian updating for forecast revision. A much earlier paper recognizing the importance of adaptively revising forecasts is given by Murray and Silver (1966). They employ a Bayesian methodology for updating an unknown sales probability of an item based on a known amount of potential customers.

In our work, we employ a learning curve approach to study the effect that different rates of demand learning have on the optimal first period batch size and second period production quantity. As detailed in the survey by Yelle (1979), learning curve applications have extended far beyond the more traditional applications of modeling the decrease in per unit manufacturing costs or the increase in labor productivity due to organizational experience. To our knowledge, this study is among the first to apply a learning curve model that predicts forecast uncertainty as decreasing with the log of time allotted to observe demand. While the application is new, previous studies suggest the applicability of its use. For example, Bitran, Haas, and Matsuo (1986)'s study of production planning at a consumer electronics company notes the reduction in the coefficient of variation (CV) between forecasts that are made in January (CV = 1), April (CV = 0.5), and October (CV = 0.2) and actual sales for the Christmas season. As opposed to using a learning curve to model this reduction in forecast uncertainty, the author's simply assume forecast error in each period is normally distributed with a known and decreasing standard deviation over time. Even though our use of learning curves to model forecast uncertainty reduction is unique, the use of learning curves to understand the benefits of learning in production planning is not. Terwiesch and Bohn (2001) study whether learning to ramp up the yield of a supply process and thus, delaying time to market and foregoing early demand, is preferable to producing first to satisfy early market demand and then investing in learning to achieve greater yield. Using a dynamic programming approach, the authors prescribe that manager's experiment early during production ramp up to increase learning early on. Despite this early learning taking time from production when prices are at their highest, the authors argue that this is the right time to devote production capacity towards engineering trials and efforts to improve yield and production rates.

In another interesting study of learning and production, Kornish and Keeney (2008) study the tradeoff of learning against capacity. Their model, motivated by the annual decision of which strains of flu to vaccinate against, address "a trade-off between quantity (producing more) and quality (produce a more effective vaccine because you know more)." By delaying the commitment to which flu strains will be targeted by the produced vaccine, a more effective vaccine can be made, but there is less time for production. The optimal policy is discussed and its implications for choosing to commit to production or wait for more information are detailed in their study.

The inclusion of a second production run, or mid-season replenishment option if in a retail setting, for short lifecycle products is a logical extension to the classical newsboy model. Lau and Lau (1997) study a mid-season replenishment possibility, but do not include set-up costs and only uniformly distributed demand is addressed in their model. While they consider making the time of the replenishment a decision variable prior to the selling season, their use of modeling each period's demand as a uniform distribution restricts this possibility. Milner and Rosenblatt (2002) study of the buyer's perspective when making a supply contract for a short life-cycle product is more amenable to making the timing of the second production run a decision variable. While the focus of their two-period model is on the contract form, they have a secondary contribution

which is to consider the duration of the first and second period as a decision variable. Fisher, Rajaram, and Raman (2001) explicitly consider time as a decision variable when planning mid-season replenishment and use a heuristic solution to help a catalog retailer. More recently, Li, Chand, Dada, and Mehta (2009) have relaxed many of the assumptions in Fisher et al. (2001) and yield more structural results on the form of the optimal policy. Our work differs from all of these mid-season replenishment models through the inclusion of setup costs and learning curves in our model.

The inclusion of setup costs in a production planning model for short lifecycle products with forecast updating has also been addressed in the literature. For example, the previously mentioned study by Bitran, Haas, and Matsuo (1986) includes the effect of setup costs for a family of products in their model. In a more recent study, Weng (2004) includes setup costs in his model to study the effects of those costs on coordination of ordering quantities between manufacturer and retailer. They find that as setup costs increase, the importance of a coordinating contract also increases. Our inclusion of both setup costs and learning allows us to find a balance between sacrificing economies of scale with additional production and benefiting from production made with less demand uncertainty.

4.2. The Two-Period Model

In this section, we first introduce the notation and the model in Subsection 4.2.1. We then characterize the optimal ordering policy in Subsection 4.2.2. Next we introduce the demand learning model in Subsection 4.2.3 and finally we study the impact of fixed cost in Subsection 4.2.4.

4.2.1 Notation and Model

Let the planning horizon be [0, T] and assume zero lead time. We will use the following notation throughout this chapter:

- t: Length of the learning period (i.e. the first period) where $0 \le t < T$. In the case of one production batch, t = 0.
- D: Total demand in period [0, T].
- D_1 : Demand in the 1st period, [0, t].
- D_2 : Demand in the 2nd period, [t, T].
- $D_2|_{D_1=\xi}$: Demand in the 2nd period [t,T] given $D_1 = \xi$ where ξ is the realization.
- x_1 : Production quantity for the first period and made available at time 0.
- x_2 : Production quantity for the second period and made available at time t.
- κ : Setup cost of production.
- π_b : Backorder penalty per unit short after the first period that is ultimately satisfied.
- π_s : Shortage penalty per unit of unmet demand at the end of the time horizon.
- r: Overage penalty (destruction/recycle cost) per unit leftover item at the end of the time horizon.
- y_2 : Order up to level for the second period.

To avoid trivial cases and ensure a realistic model, we make the following assumption.

Assumption 1 $\pi_s > \pi_b > c \ge 0$, $\pi_s - \pi_b > c$ and demand is non-negative.

Assuming zero initial inventory, we let $f_1(t)$ be the optimal expected inventory cost for the two-period problem with t being the duration of the learning period and let $\delta(x_t)$ be the indicator function of $x_t > 0$. Then, the optimal cost is expressed as

$$f_1(t) = \min_{x_1 \ge 0} \left\{ \delta(x_1)\kappa + cx_1 + \pi_b E_{D_1} \left[(D_1 - x_1)^+ \right] + E_{D_1} \left[f_2 \left(x_1 - D_1, t \right) |_{D_1} \right] \right\}, \quad (4.1)$$

where $f_2(I,t)|_{D_1}$ represents the optimal expected inventory costs for the second period conditioning on D_1 , given a starting inventory of $I = x_1 - D_1$ and the length of the first period, t.

$$f_{2}(I,t)|_{D_{1}=\xi} = \min_{x_{2}\geq 0} \left\{ \delta(x_{2})\kappa + cx_{2} - \pi_{b} \left(-I - x_{2}\right)^{+} + \pi_{s} E_{D_{2}|_{D_{1}=\xi}} \left[\left(D_{2}|_{D_{1}=\xi} - I - x_{2}\right)^{+} \right] + r E_{D_{2}|_{D_{1}=\xi}} \left[\left(I + x_{2} - D_{2}|_{D_{1}=\xi}\right)^{+} \right] \right\}.$$

$$(4.2)$$

Thus, the optimal first period cost is the sum of the first period setup costs, first period variable production costs, first period backorder penalty costs, and the expected second period costs. The second period costs consist of a setup cost for a second period production, variable production costs for the second production run, a rebate on backorder penalties charged in the first period that turn out to be lost sales (i.e. second period production does not satisfy the unmet demand of the first period), a lost sales penalty, and destruction costs. Note that in adding an additional replenishment option, we also must introduce in intra-period shortage penalty. We consider this a backorder penalty which is much less costly than the shortage penalty charged at the end of the horizon. The second period shortage penalty, since there is no additional recourse for additional replenishment, is analogous to a lost sales penalty.

In our analysis of the timing of the first period, three points are worthy of mention. First, our model explicitly excludes lead time in the consideration. Second, our model requires that the timing of the second production run be scheduled in advance of the season. Third, our model excludes the substantial risk of the trial being halted prior to the end of the time horizon. This risk, which we call failure risk, is due to the possibility of a trial showing that a drug is unsafe or ineffective prior to the conclusion of the trial.

For certain types of clinical drug supply the zero lead time assumption may be untenable. However, for other types this assumption is valid. For example, when NeoRx Corporation outsourced clinical trial supply to International Isotopes Inc., purchase orders were only placed one week in advance and rolling forecasts were provided for 3 months of future demand (NeoRX, 10-Q Filed on May 9, 2000.). These durations are much shorter than the overall duration of the trial, which may take as long as 2-3 years. Please note that these purchase orders are for batches produced after the first batch. The lead time on the first batch may still be lengthy as manufacturing facilities are configured for initial production. Once production facilities and processes are in place, lead time on additional batches can be much shorter than that of the first batch.

The timing of the intra-trial batch being determined prior to commencement of the actual trial is consistent with outsourcing contracts for clinical supply where the availability of manufacturing capacity is reserved in advance. It is also worth noting that Li, Chand, Dada, and Mehta (2009) have found little value to dynamically determining the timing of the second production run. We do not consider the effects of cancellation fees on the supply manager's decisions, although we believe this might be an interesting area for future research.

Lastly, the risk of failure risk, although important, is excluded from our analysis and reflects the typical supply philosophy of planning for success during a clinical trial. If failure were to occur during time (0, t), it is as if the second production option would go unutilized. By applying a discount factor to the cost of producing in the second period, this aspect of clinical trials could be captured. To keep our analysis to the balancing of economies of scale and uncertainty reduction, we propose that inclusion of failure risk may be an interesting area of future research.

4.2.2 Optimal Ordering Policy

It is more convenient to use $y_2 = I + x_2$ and thus Eq. (4.2) becomes,

$$f_2(I,t)|_{D_1=\xi} = -cI + \min_{y_2 \ge I} \left\{ \delta(y_2 - I)\kappa + cy_2 - \pi_b \left(-y_2\right)^+ + L_{\xi}(y_2,t) \right\},$$
(4.3)

where $L_{\xi}(y,t) = \pi_s E_{D_2|_{D_1=\xi}} \left[(D_2|_{D_1=\xi} - y)^+ \right] + r E_{D_2|_{D_1=\xi}} \left[(y - D_2|_{D_1=\xi})^+ \right].$

To further analyze Eq. (4.3), we note that we can either produce (i.e. $x_2 > 0$) or

not produce,

$$f_{2}(I,t)|_{D_{1}=\xi} = -cI + \min\left\{\min_{y_{2}>I}\left\{\kappa + cy_{2} - \pi_{b}\left(-y_{2}\right)^{+} + L_{\xi}(y_{2},t)\right\}, \quad (4.4)$$
$$cI - \pi_{b}\left(-I\right)^{+} + L_{\xi}(I,t)\right\}.$$

Observation 1 If $y_2 > I$, then the optimal order-up-to level for the second period $y_2^* \ge 0$.

Proof. If $I \ge 0$, $y_2^* \ge 0$ by definition. If I < 0, consider a $y_2 \le 0$. By Assumption 1 and Eq. (4.3),

$$\kappa + cy_2 - \pi_b (-y_2)^+ + L_{\xi}(y_2, t) = \kappa + cy_2 - \pi_b (-y_2) + \pi_s E_{D_2|_{D_1 = \xi}} [D_2|_{D_1 = \xi} - y_2]$$

= $\kappa + cy_2 + E_{D_2|_{D_1 = \xi}} [\pi_s D_2|_{D_1 = \xi} + (\pi_s - \pi_b)(-y_2)].$

By Assumption 1, the cost function is decreasing in y_2 for $y_2 \le 0$. Thus $y_2^* \ge 0$. \Box By Observation 1, Eq. (4.4) can be reduced to,

$$f_2(I,t)|_{D_1=\xi} = -cI + \min\left\{\min_{y_2>I}\left\{\kappa + cy_2 + L_{\xi}(y_2,t)\right\}, cI - \pi_b\left(-I\right)^+ + L_{\xi}(I,t)\right\}.$$
(4.5)

The following observation shows that the second period cost function is not convex and thus we cannot directly apply the classical result of (s, S) policy (e.g., see Zipkin (2000, Section 9.5)) to this problem.

Observation 2 The second period cost function $cI - \pi_b (-I)^+ + L_{\xi}(I, t)$ is not convex in *I*, but it is unimodular in *I* and approaches infinity as $I \to \pm \infty$.

Proof. First, we note that $-\pi_b (-I)^+$ is concave in *I*. For I < 0, it follows by Assumption 1 that the second period cost function reduces to

$$cI - \pi_b(-I) + \pi_s E_{D_2|_{D_1=\xi}}[D_2|_{D_1=\xi} - I] = \pi_s E_{D_2|_{D_1=\xi}}[D_2|_{D_1=\xi}] - (\pi_s - \pi_b - c)I,$$

which is clearly convex in I. For $I \ge 0$, the second period cost function reduces to

 $cI + L_{\xi}(I, t)$, which is also convex in I. However, the left derivative of the cost function at I = 0 equals $-(\pi_s - \pi_b - c) < 0$ (by Assumption 1), which is greater than the right derivative of the cost function at $I = 0, -(\pi_s - c)$. Thus, the cost function is not convex in I for $I \in (-\infty, \infty)$.

To show the cost function is unimodular, we note that it is convex and decreasing in I for $I \in (-\infty, 0]$. Because it is also convex in I for $I \in [0, \infty)$, it must be unimodular.

Finally, as $I \to -\infty$, the slope of the cost function is $-(\pi_s - \pi_b - c)$; as $I \to \infty$, the slope approaches c + r. The proof is now completed.

Now we are ready to identify the optimal ordering policy for the second period given $D_1 = \xi$. Let $S_2(\xi)$ be the smallest global minimizer of $cy + L_{\xi}(y,t)$, and $s_2(\xi)$ be the largest I (but smaller than $S_2(\xi)$) such that $cI - \pi_b(-I)^+ + L_{\xi}(I,t) = \kappa + cS_2(\xi) + L_{\xi}(S_2(\xi),t)$. Indeed, $S_2(\xi) = \Phi_{D_2|_{D_1=\xi}}^{-1} \left(\frac{\pi_s - c}{\pi_s + c}\right) \ge 0$, where $\Phi_{D_2|_{D_1=\xi}}(\cdot)$ is the probability density function of $D_2|_{D_1=\xi}$. $s_2(\xi)$ must exist by the unimodularity and asymptotic properties shown in Observation 2.

Theorem 4 The optimal ordering policy for the second period is a (s, S) type of policy depending on $D_1 = \xi$, where $s = s_2(\xi)$ and $S = S_2(\xi)$. In other words, if the beginning inventory position $I < s_2(\xi)$, we order up to $S_2(\xi)$; otherwise, we do not order.

Proof. The proof follows directly from the definition of $s_2(\xi), S_2(\xi)$ and Observation 2.

Note that $s_2(\xi)$ and $S_2(\xi)$ are dependent on $D_1 = \xi$ but independent of second period starting inventory I.

By Theorem 4, we can write $f_2(I,t)|_{D_1=\xi}$ as follows,

$$f_2(I,t)|_{D_1=\xi} = -cI + \begin{cases} \kappa + cS_2(\xi) + L_{\xi}(S_2(\xi),t), & I \le s_2(\xi) \\ cI - \pi_b(-I)^+ + L_{\xi}(I,t), & I > s_2(\xi). \end{cases}$$
(4.6)

We now show $f_2(I,t)|_{D_1=\xi}$ is κ -convex for any ξ . By Zipkin (2000, Section 9.5), we have the following definition.

Definition 1 We call a function f(x) k-convex if for any x, and nonnegative u and v, f(x) satisfies

$$f(x) + v \frac{f(x) - f(x - u)}{u} \le f(x + v) + k.$$

Lemma 1 $f_2(I,t)|_{D_1=\xi}$ is k-convex in I for any ξ .

Proof. For simplicity, we drop ξ from our notation without causing confusion. We also define $\tilde{f}(I,t) = f_2(I,t) + cI$. Clearly if $\tilde{f}(I,t)$ is k-convex, $f_2(I,t)$ is also k-convex.

If $s_2 \ge 0$, then

$$\tilde{f}_2(I,t) = \begin{cases} k + cS_2 + L(S_2,t), & I \le s_2 \\ cI + L(I,t), & I > s_2. \end{cases}$$

By Zipkin (2000, Section 9.5), $\tilde{f}_2(I, t)$ and thus $f_2(I, t)$ is k-convex.

If $s_2 < 0$, we first consider any $I \le s_2$. By the definition of the (s_2, S_2) policy, we must have $\tilde{f}_2(I,t) + v \frac{\tilde{f}_2(I,t) - \tilde{f}_2(I-u,t)}{u} = \tilde{f}_2(I,t) = \tilde{f}_2(S_2,t) + k \le \tilde{f}_2(x,t) + k$ for all x. Next we consider $I \ge S_2$. The k-convexity inequality must hold because $\tilde{f}_2(I,t)$ is convex and increasing. Finally, we consider $s_2 < I \le S_2$. Note that $\tilde{f}_2(I,t)$ is decreasing for $I < S_2$ by Observation 2, thus $\tilde{f}_2(I,t) + v \frac{\tilde{f}_2(I,t) - \tilde{f}_2(I-u,t)}{u} \le \tilde{f}_2(I,t) \le \tilde{f}_2(S_2) + k \le \tilde{f}_2(I+v) + k$. The proof is now completed. \Box

Theorem 5 The optimal ordering policy for the first period is a (s, S) type of policy.

Proof. By Lemma 1, $f_2(I,t)|_{D_1=\xi}$ is k-convex. By Lemma 9.5.1 of Zipkin (2000), $E_{D_1}[f_2(x_1 - D_1, t)|_{D_1}]$ is also k-convex, and so is $cx_1 + \pi_b E_{D_1} \left[(D_1 - x_1)^+ \right] + E_{D_1} \left[f_2 (x_1 - D_1, t) |_{D_1} \right]$. By Theorem 9.5.2 of Zipkin (2000), the proof is completed.

Let (s_1, S_1) be the optimal (s, S) policy for the first period. Thus, if $s_1 > 0$, then we produce up to S_1 in the first period. Otherwise, we do not produce.

4.2.3 Demand Learning Model

The initial belief of the total demand during the planning horizon is that D is normally distributed with mean μ and standard deviation σ . We model demand learning by the dependent demand in periods 1-2: D_1 and D_2 ($D = D_1 + D_2$). Our demand learning model follows that of Fisher and Raman (1996) and Fisher, et al. (2001). Specifically, we assume (D_1, D_2) follows a bivariate normal distribution with correlation coefficient $\rho(t)$, where $\rho(t)$ depends on the amount of learning that can take place by time t. In Section 3.5, we borrow methodology from the learning curves literature to model this function. The marginal distribution D_1 is also normal with mean μ_1 and standard deviation σ_1 . We assume that for $t \in [0, T]$,

$$E(D_1) = \mu_1 = \alpha(t)\mu, \qquad \sigma^2(D_1) = \sigma_1^2 = \beta(t)\sigma^2,$$

where $\alpha(t)$ and $\beta(t)$ are fractions increasing from 0 to 1 as t increases from 0 to T. For example, $\alpha(t) = t/T$ and $\beta(t) = t/T$. Then the marginal distribution of D_2 is normal with

$$E(D_2) = \mu_2 = (1 - \alpha(t))\mu, \quad \sigma(D_2) = \sigma_2 = -\rho(t)\sigma_1 + \sqrt{\rho^2(t)\sigma_1^2 + \sigma^2 - \sigma_1^2}.$$

Conditioning on $D_1 = \xi$, D_2 follows a normal distribution with the following parameters (Fisher and Raman 1996):

$$\mu_2(\xi) = \mu_2 + \rho(t)\sigma_2 \frac{\xi - \mu_1}{\sigma_1}, \quad \sigma_2(\xi) = \sigma_2 \sqrt{1 - \rho^2(t)}.$$

It is easy to see that given t, as ρ increases, σ_2 decreases and thus $\sigma_2(\xi)$ decreases.

4.2.4 Impact of The Fixed Cost

Intuitively, as the fixed cost, κ , increases, the two-period model reduces to the newsvendor model. This intuition is confirmed by the following proposition.

Proposition 2 S_1 tends to the newsvendor quantity as $\kappa \to \infty$.

Proof. For simplicity, we drop the dependence on D_1 for (s_2, S_2) without causing confusion. Suppose we produce in the first period, the total cost function can be expressed as follows,

$$\kappa + \min_{x_1>0} \left\{ cx_1 + \pi_b E_{D_1} [(D_1 - x_1)^+] + E_{D_1} [-c(x_1 - D_1) + \kappa + cS_2 + L_{D_1}(S_2, t) | D_1 \ge x_1 - s_2] + E_{D_1} [-\pi_b (D_1 - x_1)^+ + L_{D_1} (x_1 - D_1, t) | D_1 < x_1 - s_2] \right\}.$$

As $\kappa \to \infty$, $s_2 \to -\infty$ (by Observation 2) for each realization of D_1 . Thus, the cost function tends to

$$\kappa + \min_{x_1>0} \Big\{ cx_1 + E_{D_1} [L_{D_1}(x_1 - D_1, t)] \Big\},$$

where $E_{D_1}[L_{D_1}(x_1 - D_1, t)] = E_{D_1}[E_{D_2}[\pi_s(D_1 + D_2 - x_1)^+ + r(x_1 - D_1 - D_2)^+ |D_1]] = E_{D_1+D_2}[\pi_s(D_1 + D_2 - x_1)^+ + r(x_1 - D_1 - D_2)^+]$. The last equality comes from the definition of conditional expectation. Note that $E_{D_1+D_2}[\pi_s(D_1 + D_2 - x_1)^+ + r(x_1 - D_1 - D_2)^+]$ represents the cost function of the newsvendor model without the second period, the proof is now completed.

In general, as the fixed cost κ increases, S_1 will be more likely used to cover both D_1 and D_2 , and thus S_1 typically increases.

4.3. Numerical Analysis

The objective of this section is to quantify the effects of setup costs, learning rates, and penalty costs on the value of the second production option, the optimal timing of learning, and the optimal first batch size. The penalty costs include the overage penalty cost, r + c, and underage penalty costs, $\pi_s - c$ and $\pi_b - c$.

As a starting point for the analysis, a baseline problem is created where we assume mean demand over the time horizon is 1,000 units and the standard deviation of demand is 300 units. We assume demand for the time horizon is normally distributed and since this distribution is divisible, we can mathematically divide demand between two periods. Thus, the fraction, $\frac{t}{T}$, represents the percentage of total demand expected to occur in the first period. The parameters of our baseline model, in the absence of demand learning (i.e. $\gamma = 0$), are shown in Table 4.1.

| Parameter | Description | Baseline Value |
|-----------|--------------------------------------|--|
| t | Length of first (learning) period. | 5 |
| T | Length of planning horizon. | 10 |
| γ | Rate of demand learning. | 0 |
| D(0,t) | First Period Demand $N(\mu,\sigma)$ | $N(1000 * \frac{5}{10}, \sqrt{300^2 * \frac{5}{10}})$ |
| D(t,T) | Second Period Demand $N(\mu,\sigma)$ | $N(1000 * \frac{10-5}{10}, \sqrt{300^2 * (1 - \frac{5}{10})})$ |
| κ | Setup cost of a production run | 0 |
| π_b | Backorder penalty | 20 |
| π_s | End of horizon shortage penalty | 50 |
| С | Variable production cost | 2 |
| r | Destruction Cost | 1 |

 Table 4.1: Baseline Parameter Values

In Figure 4.1, we compare the expected costs of our baseline model with the expected costs of a newsvendor model (i.e. a single production run at t = 0). For our baseline model, we have arbitrarily scheduled an additional replenishment option midway through the planning horizon. As can be seen from the graph, the additional replenishment opportunity leads to a greater than 10% reduction in costs. It is also interesting to note that the expected costs of the baseline model are less sensitive than the newsvendor model to first period order quantity. At this point, one might conclude that a mid-season replenishment can reduce expected costs and lead to a decision that is less sensitive to model parameters. However, the comparison we made ignores several key components that constitute the motivation for this study. First, setup costs are zero and when producing clinical trial drug supply, large setup costs are a reality that must be accommodated. And second, our comparison fails to account for demand learning and the freeing of the variable, t, so that this additional production can be optimally scheduled.

We now introduce setup costs into our analysis without considering demand learning. From the proof of Proposition 2, we know that as setup costs increase the two-period model reduces to the newsvendor model. We can see this effect by comparing Figure 4.2 which includes a setup cost ($\kappa = \$1,600$) to Figure 4.1 which assumes there are zero fixed costs when producing. With setup costs introduced, we can see that the value of an additional replenishment midway through the planning horizon yields minimal savings of 2% of the newsvendor solution's inventory costs.

Counteracting this decrease in value from our baseline model, we can free the scheduling of the second replenishment and show the effects of this scheduling on expected costs. To analyze this, we graph expected costs as a function of the first period length (t) as seen in Figure 4.3. For every choice of first period length, t, the optimal first period order quantity has been numerically determined. We see the optimal first period length is close to the entire planning horizon $(t \approx 8.5)$ and not the arbitrarily chosen mid-horizon production (t = 5). As might be expected, the mid-horizon production case with setup costs leads to reduced savings over the newsvendor solution as compared to an available second production run without setup costs. Interestingly, we see that when t = T, the solution still outperforms the newsvendor solution. In this case, the value of the additional replenishment (when optimally planned) is purely derived from replacing the expected end of horizon shortage penalty with a backorder penalty in cases of high demand. In contrast, when t = 8.5, some of the value provided

by this model as compared to the neswvendor solutions is from the ability to effectively match supply and demand for a portion of the horizon.



Figure 4.1: Expected Costs when $\kappa =$

While setup costs have reduced some of the benefit of the intra-season replenishment option, demand learning creates greater incentives to plan an additional replenishment and counter-balance the costs imposed by an additional setup. In Section 4.2.3, we introduced the notion that demand learning will be modeled through correlation between first and second period demand. Through correlation, a fraction of the variance in the second period's demand is explained by the realization of the first period's demand. Mathematically, this fraction is simply $\rho(t)^2$ and the fraction of variance that remains unexplained in the second period is $1 - \rho(t)^2$. Consistent with this mathematical interpretation, we will model demand learning as a reduction in unexplained variance. Intuitively, the fraction of unexplained variance in second period demand should be close to one early in the time horizon and closer to zero at the end of the horizon. To study different rates of learning, we will assume that learning, more specifically the reduction in uncertainty surrounding second period demand, $1 - \rho(t)^2$, follows a power law form that was introduced as a learning curve model by Wright (1936). Modifying



Figure 4.2: Expected Costs when $\kappa = \$1,600$

Wright's learning curve to represent a supply manager's ability to remove uncertainty in the variance of second period demand, we have $1 - \rho(t)^2 = \left(\frac{T-t}{T}\right)^{\gamma}$ and thus:

$$\rho(t) = \sqrt{1 - \left(\frac{T-t}{T}\right)^{\gamma}} \tag{4.7}$$

where $\gamma > 0$ is the shape parameter of the learning curve. When $\gamma = 1$ the amount of learning is linear in time, when $\gamma < 1$, learning is slow and the uncertainty parameter, $1 - \rho(t)^2$, will be a concave function of time. Lastly, when $\gamma > 1$, learning occurs more rapidly and $1 - \rho(t)^2$ is a convex function of time.

Intuitively, faster rate of demand learning encourages earlier scheduling of the potential second production run. To see this in our example, we now analyze the baseline model with setup costs for various rates of demand learning. We pick various values of our learning parameter, γ , to represent different rates of uncertainty reduction and plot the expected costs of our baseline model with setup costs ($\kappa = \$1,600$) and learning in Figure 4.4. We see from this graph that faster learning leads to both an earlier scheduling for the second production run and larger cost reductions versus the single



Figure 4.3: Expected Costs Versus Timing of Second Replenishment with Setup Costs

newsvendor production. More importantly, the cost benefit of learning has effectively nullified the substantial setup cost. We see that when the rate of learning is simply linear in time ($\gamma = 1$), we can achieve savings of greater than 10% over the single production solution. This linear learning rate would yield a correlation coefficient of 0.82 at the optimal duration of the first period ($t \approx 6.75$) which is consistent with the correlations used in (Fisher, Rajaram, and Raman, 2001).

While the scheduling decision is important, it is not made in isolation. The optimal supply strategy will simultaneously consider the timing of the second production and the sizing of the first batch. In Figure 4.5, we analyze the interplay of production scheduling and optimal first batch size for various setup costs and linear learning ($\gamma =$ 1). In the absence of setup costs (i.e. $\kappa = 0$) and when replenishment is planned after observing a small fraction (roughly 10%) of demand, we observe that the optimal first period batch is less than half the newsvendor batch size. However, with even modest setup costs of \$100, the optimal batch sizing this early in the season is much closer to the newsvendor quantity. This is a key observation that the suggestion of a "modest amount of initial inventory" (Fisher and Raman (1996)) is less appropriate when setup



Figure 4.4: Expected Costs Versus Timing of Second Replenishment with Setup Costs and Demand Learning

costs are factored into the decision making. The benefits of ample inventory, including avoiding additional setup costs and first period backorder costs, outweigh the benefits of uncertainty reduction afforded by a second production.

In the absence of setup costs and in the presence of demand learning, a second production run is a likely event. For our baseline model with linear learning and zero setup costs, the second production run is optimally scheduled at around $t \approx 5.5$ and it is expected that 71% of the time the production run will be utilized. The other 29%, demand is so low in the first 55% of the planning horizon that a second production is not needed. Even though overage risk is present, this risk is offset by having enough inventory to avoid intra-period backorder costs.

As soon as we introduce setup costs, we also introduce a notion of economies of scale in production. A manager's expectation of producing more than once reflects his willingness to sacrifice scale economies to achieve savings. The tradeoff between sacrificing scale economies to better match supply and demand is summarized in Table 4.2. We see from this table that setup costs significantly decrease the probability of a second production. For example, in the case of linear learning ($\gamma = 1$), the introduction of



Figure 4.5: Optimal First Batch Size Versus Timing of Second Replenishment with Setup Costs and Demand Learning

setup costs of \$1,600 reduces the likelihood of producing a second time from 71% to 21%. Further increases in setup costs drastically reduce the likelihood of producing a second time. From a planning perspective, mid-season replenishment in the presence of high setup costs is really an emergency supply option for cases of extremely high demand.

Even though the likelihood of producing a second time can be small, the value of this option remains significant in the presence of learning. This can be seen in Table 4.3 which shows the expected savings over the newsvendor model when optimally scheduling potential replenishment. From this table, we can see that with linear learning and setup costs of \$1,600, a 10.7% reduction in costs can be expected by just having a resupply option available. From Table 4.2, we know that this resupply option will only be exercised about 21% of the time. Digging deeper into Table 4.3, we see that even with higher levels of setup costs (e.g. \$3,200), savings of greater than 5% are achievable. While in a pharmaceutical setting, these cost reductions are significant, in a retail setting Fisher et al. (2001) show how much smaller cost savings can translate into big gains in profitability.

| | Learning Rate (γ) | | | | |
|------------|--------------------------|-------|-------|----------|-------|
| κ | 0 | 0.5 | 1 | 2 | 4 |
| - | 66.7% | 66.3% | 70.6% | 74.1% | 77.4% |
| 100 | 50.2% | 59.4% | 64.5% | 69.1% | 73.6% |
| 200 | 41.1% | 52.1% | 57.5% | 63.0% | 68.1% |
| 400 | 30.5% | 43.3% | 48.8% | 54.4% | 60.6% |
| 800 | 19.8% | 31.9% | 36.8% | 40.0% | 47.5% |
| $1,\!600$ | 10.4% | 18.7% | 21.4% | 24.4% | 27.2% |
| 3,200 | 4.8% | 9.2% | 10.2% | 11.0% | 11.9% |
| $6,\!400$ | 1.6% | 3.4% | 3.6% | 3.8% | 4.0% |
| $12,\!800$ | 0.1% | 0.7% | 0.7% | 0.8% | 0.8% |

Table 4.2: Probability of Mid-Season Replenishment for Various Setup Costs and Learning Rates

| | Learning Rate (γ) | | | | |
|------------|--------------------------|-------|-------|----------|-------|
| κ | 0 | 0.5 | 1 | 2 | 4 |
| - | 16.5% | 28.8% | 31.7% | 34.8% | 38.2% |
| 100 | 14.7% | 26.5% | 29.1% | 32.1% | 35.3% |
| 200 | 13.1% | 24.4% | 26.9% | 29.7% | 32.7% |
| 400 | 10.9% | 20.9% | 23.1% | 25.5% | 28.0% |
| 800 | 7.8% | 15.8% | 17.4% | 19.1% | 20.8% |
| $1,\!600$ | 4.6% | 9.8% | 10.7% | 11.6% | 12.4% |
| $3,\!200$ | 1.9% | 4.5% | 4.9% | 5.2% | 5.5% |
| $6,\!400$ | 0.4% | 1.3% | 1.4% | 1.5% | 1.6% |
| $12,\!800$ | 0.0% | 0.2% | 0.2% | 0.2% | 0.2% |

Table 4.3: Table of Savings for Various Setup Costs and Learning Rates

Another consideration in this mid-season replenishment environment is how penalty costs (i.e. overage, lost sales, and intra-period backorder penalties) impact our decisions of replenishment timing and first period batch size. And even more importantly, how do changes in these parameters affect the magnitude of savings over a simpler newsvendor solution? The clinical trial supply environment is driven by a fear of delaying a trial due to insufficient supply and intuitively, one would think increasing underage penalties (π_s or π_b would lead to greater expected savings of a second production. In studying this numerically, we surprisingly find the advantage of having an intra-season replenishment option is not dramatically improved by dramatically increased underage penalties. For example, our numerical study has found that doubling the two underage penalties of our baseline model with setup costs ($\kappa = \$1, 600, \pi_b = 40, \pi_s = 100$) only increases expected savings over the newsvendor solution an additional 2.8% from 10.7% to 13.5%. Further increases to these underage penalties, as shown in Table 4.4, yield similarly modest results with the reason being that avoiding these underage penalties is relatively inexpensive; overage costs are only \$3 which is small in comparison to the end of horizon shortage penalty of \$50 of our baseline model. Basically, it is cheap to hedge against having too little inventory by simply producing more.

| π_b / π_s | % Savings |
|-----------------|-----------|
| 20 / 50 | 10.7 |
| 40 / 100 | 13.5 |
| 80 / 200 | 15.8 |
| 160 / 400 | 17.7 |
| 320 / 800 | 19.35 |

Table 4.4: Expected Savings Over a Newsvendor Solution for Increasing Underage Costs

In studying the effect of changes to the overage penalty, we find that increasing the overage penalty leads to greater jumps in savings magnitude than increasing underage costs. For example, if we look at the case where $\pi_b = 40, \pi_s = 100$, and $\kappa = \$1,600$, we find that savings of greater than 50% over the newsvendor solution can be expected when overage costs are 64% of the end of horizon shortage costs. A selection of overage
cost penalty values and expected savings are shown in Table 4.5. Less dramatic than the changes in savings, both optimal batch sizes and optimal timing values fall within small ranges as the overage penalty is adjusted. Optimal batch sizes fall between 71% and 79% of their respective newsvendor optimal order sizes and optimal timing for the second production run is between 62% and 68% of the planning horizon.

While increases to both overage and underage penalties will always increase savings over a newsvendor solution, the increase in percentage savings can be both small and large. Percentage savings increases per dollar of increased penalty cost are fastest when overage and underage costs are highly unbalanced and increases are made to the lower of the two costs. Conversely, when increases are made to the higher of the two costs, only marginal benefits will be realized. In a pharmaceutical setting where underage costs far exceed overage costs, the observation on the effects of increasing overage penalties suggests that increases in variable production costs, which effectively reduce the lost sales penalty and increase the overage penalty, will greatly increase the attractiveness of intra-season replenishment.

| r | % Savings |
|-----|-----------|
| 1 | 13.5 |
| 2 | 16.9 |
| 4 | 22.4 |
| 8 | 29.9 |
| 16 | 38.1 |
| 32 | 45.4 |
| 64 | 50.7 |
| 128 | 53.9 |

 Table 4.5: Expected Savings Over a Newsvendor Solution for Increasing Destruction

 Costs

In summary, our numerical study pursued an understanding of the value of an intraseason replenishment over the newsvendor solution. The value that is created depends on two key decisions in planning for this intra-season option. First, how is our first batch size decision affected by the presence of a resupply option and second, when should the potential resupply be planned for. These two decisions, and the potential to achieve meaningful savings are affected by a multitude of parameters and for simplicity, we now summarize our findings in Table 4.6.

| relative to newsvendor solution | | | | | | | |
|---------------------------------|---|--------------------------------------|----------------|-----------------------|--|--|--|
| Parameter † | % Savings | 1st Batch Size (%) | 1st Batch Size | 1st Period Length | | | |
| | $\left(\frac{f_1(0) - f_1(t_{opt})}{f_1(0)}\right)$ | $\left(\frac{x_1}{x_{newsv}}\right)$ | (x_1) | (t) | | | |
| Learning (γ) | \uparrow | \checkmark | \checkmark | \checkmark | | | |
| Setup Costs (κ) | \checkmark | 1 | \uparrow | $\checkmark \uparrow$ | | | |
| Backorder Penalty (π_b) | \checkmark | \checkmark | \checkmark | \checkmark | | | |
| Shortage Penalty (π_s) | 1 | $\wedge \uparrow$ | \uparrow | ↑ | | | |
| Destruction Costs (r) | 1 | $\wedge \uparrow$ | \checkmark | \checkmark | | | |

Table 4.6: Effect of Parameter Increases on Performance and Decision Variables

4.4. Summary Remarks

The inclusion of setup costs and learning curves in our study leads us to many conclusions that add to the body of literature dealing with an additional replenishment options for products with short lifecycles. When setup costs are present, the first batch remains large as compared to a newsvendor batch. As opposed to a smaller batch, the inventory helps to potentially avoid incurring setup costs a second time and avoiding intra-season backorder penalties.

Selection of learning period length is driven by both changes in setup costs and learning. We observe that increasing setup costs will initially increase optimal period length and then decrease it. At lower levels of setup costs, when these costs increase, it is advantageous to have a longer learning period to permit greater observation of demand and a more certain second period forecast. Eventually, further increases to setup costs decrease the learning period to avoid backorder costs in the cases of extremely high demand that would actually warrant incurring a second setup.

Our study is the first to look at the tradeoff between sacrificing economies of scale by planning for multiple batches and benefiting from demand learning so that a better match of supply and demand can be made. We have found that the ideal conditions for consideration of an additional production run are when setup costs are low, learning is fast, and both overage and underage penalties are significant. In certain examples, we find savings to exceed 50%. Admittedly, these high-value examples are less applicable to clinical trials where underage costs far exceed overage costs. However, these high value examples are realistic when extending this model to a fashion environment or other short lifecycle product where underage and overage costs are not so lop-sided.

Chapter 5

Inventory Positioning in Clinical Trial Supply Chains

Every second of delay in a clinical trial costs Bristol-Meyers Squibb \$17.¹

At a cost of \$17 per second, clinical trial delays are to be avoided. Unfortunately, avoiding clinical trial delays is easier said than done. The biggest stumbling block is often patient recruitment. According to Getz and de Bruin (2000), 80% of clinical trials fail to meet their patient recruitment deadlines. As a result of slow patient recruitment and also high patient costs in the United States, clinical trials are increasingly going global (Rowland, 2004). In quantifying this globalization, Thiers, Sinskey, and E.R. (2008) report the following data on the growth in the number of clinical trial sites for various countries:

| Country | Annual Growth Rate (%) | | | | |
|----------------|------------------------|--|--|--|--|
| China | 47.0 | | | | |
| Russia | 33.0 | | | | |
| Argentina | 26.9 | | | | |
| Czech Republic | 24.6 | | | | |
| Mexico | 22.1 | | | | |
| United States | -6.5 | | | | |

Table 5.1: Growth in Number of Offshore Clinical Trial Sites

¹Paul Loveday (CEO, ClinStar) at the Clinical Research in Emerging Countries Third Annual Marcus Evans Conference 21-22 July 2008, Washington DC, USA

We see from Table 5, that clinical trial growth in emerging markets far outpaces that in the United States. In fact, the United States is experiencing declines in the number of trial sites being used. In addition to the relocation of trials, more clinical trials are seeking participants in multiple countries simultaneously. In 2005, 7.8% of the trials reported in three top medical journals were being conducted in 10 or more countries while in 1995, none of the articles in those journals reported such high levels of globalization (Glickman, McHutchison, Peterson, Cairns, Harrington, Califf, and Schulman, 2009).

Unfortunately, while recruitment efforts benefit from a larger global footprint, the supply chain has to work much harder at getting the right drug supply, usually in the form of patient kits, to the right place, at the right time. Subjected to sets of local regulations and various levels of supporting infrastructure, a supply chain for global clinical trials becomes much more complex than the supply chain of a one country trial. Supported by results of a BearingPoint and AMR Research survey which found that only 13% of clinical trial products are received on time at investigative sites, one senior industry consultant comments that "Most current supply chains are entirely inadequate for the realities of global trials today." (Neuer, 2008).

In addition to the complexity of a global supply chain, simply spreading out demand over numerous sites increases the amount of inventory required. Let's give a simple example to demonstrate this. Let's assume we have a goal of recruiting 612 patients for our trial (this example is stylized from the 612 patient, 45 site trials described in Le Chevalier, Brisgand, Douillard, Pujol, Alberola, Monnier, Riviere, Lianes, Chomy, and Cigolari (1994)). If one were to recruit all 612 patients from one clinical trial site, then we would simply send 612 patient kits to the site. Now, let's assume we open two sites and in an effort to get to 612 patients as fast as possible, we aim for a 99% service level (non-rejection) at each of the sites. Assuming the sites are identical in their patient recruiting rates, i.e., there is a 50% chance that the first patient goes to site one and a 50% chance that this patient goes to site two, and so on. Extending this 50/50 logic to our first 612 patients, the distribution of the number of patients at site one is a binomial distribution with 612 trials and a 50% probability of success. To maintain a 99% service level, we would ship 335 patient kits to site one. Since, the sites are identical, we would also ship 335 patient kits to site two. This is a somewhat palatable increase in inventory of 9.5% over the one site case. Now let's extend this logic to 45 identical sites. In this example, we now must increase inventory by 69.1% over the one site trial and 423 patient kits will ultimately be unused overage.

It now becomes clear that both the demand chain and the supply chain for clinical trials are in flux. The demand chain, in its effort to increase patient recruitment rates, is going global with an increasing number of investigative sites. The supply chain, which may have been adequate for domestic trials, is now struggling with meeting the increased demands imposed by a global footprint and an increased number of investigative sites. Simply assuming that supply can match globally dispersed demand is not a good strategy. The supply chain will inevitably fall short of these expectations. The right strategy will shape demand (through the opening and closing of sites) to match a realistic supply strategy. One might now think that all supply strategies are possible, but the supply of clinical material is limited by capacity, drug expiry, and cost among other factors.

In this chapter, we commence a stream of academic literature to address some of the main challenges in clinical trial supply chains: How many sites are needed and where to locate them? How to balance the trade-off between faster patient recruiting rate and more excessive drug supply requirements (thus longer production time)? Given a configuration of clinical trial supply chain, how to place inventory in the central distribution center (CDC), the regional warehouse (RWH) and sites? As a step in towards answering all of these questions, we develop mathematical models to aid decision making in stock positioning and site selection in the clinical trial supply chain. We derive the following results: (1) We first present a new class of multi-echelon inventory models to make stock positioning decisions, (2) we develop algorithms to identify lower and

upper bounds on the optimal objective function for this new class, and (3) we leverage the algorithms to provide insights into optimal supply chain configurations.

The remainder of this chapter is organized as follows: After reviewing the relevant literature in §5.1, we introduce basic modeling assumptions and study two extreme cases in §5.2. The two cases include one with a central warehouse and one without warehouses. In §5.3, we consider the general clinical trial supply chain and develop solution algorithms for stock positioning. In §5.4, we present a numerical study to test the effectiveness of the solution and to quantify the impact of various supply chain configurations and parameters. §5.5 concludes the chapter.

5.1. Literature Review

A good introduction to the challenges in managing a global clinical trial supply chain can be found in Lis, Gourley, Wilson, and Page (June 1, 2009) where it is succinctly noted that "the key challenge clinical trial supply chain (CTSC) managers face in global distribution is ensuring that supplies arrive at the trial sites on time and in good condition." To be on time, the inventory must not only be produced in sufficient quantity to meet demand, but must also be positioned in the supply chain to satisfy demand as it is realized.

The management of inventory in these supply chains hasn't attracted much attention within industry literature, and has attracted even less attention from academics. In industry literature, it is advocated that operational policies in the clinical trial supply chain are usually created and managed using simulation (Peterson, Byrom, Dowlman, and McEntegart, 2004). In addition, it is often assumed within these simulations that an integrated voice response system (IVRS) is available so that inventory can be monitored continuously (McEntegart and O'Gorman, 2005) and as inventory at a location falls below a specified trigger, more inventory is ordered. Despite the availability of more sophisticated alternatives, inventory policies in the clinical trial supply chain are still often set using experience and by looking at the patterns from previous clinical trials.

To supplement the utility of experience, academics have created more sophisticated models for managing inventory in distribution networks (e.g., for spare parts) similar to the one we model for a clinical trial supply chain. However, these models have yet to find their way to the problems faced by clinical trials and yet, it is noted that these models are needed. Shah (2004) points out that many of the key challenges faced by pharmaceutical supply chains in general and surveys the literature that addresses those challenges. Of particular interest is the recent focus of academics on capacity planning for clinical trial supply. Our work differs from this surveyed work in that we assume the capacity decision has already taken place and we now focus on the more executional/tactical policy details of managing inventory for a given supply chain.

Within the context of multiechelon inventory research, our models are most closely related to work done for service parts where end-user demand is low and one-for-one base stock policies are employed (see reviews by Zipkin 2000, Simchi-Levi and Zhao 2006, and Muckstadt 2005). One of the seminal works in this stream of literature is Sherbrooke (1968). He approximated the distribution of the backorder at the depot by its first moment in a two-echelon supply chain when one-for-one ordering policies are used. Improving on this approximation is the approximation by Graves (1985) who shows how to effectively approximate backorder and lead time demand by a negative binomial distribution. We leverage the work of Graves (1985) to optimize certain subsystems within a three-echelon supply chain that includes one central warehouse, multiple regional depots and multiple sites, where sites can be either supplied directly by the central warehouse or indirectly by a regional warehouse. Svoronos and Zipkin (1991) refines the approximation by Graves (1985) and extends it to evaluate multi-echelon distribution systems.

Graves (1985) and Axsater (1990) provide means to exactly evaluate the distribution of net inventory levels in a multi-echelon supply chain, but these methods require the convolution of multiple probability distributions and thus are computationally intensive. Simchi-Levi and Zhao (2005) extends the exact approach to evaluate tree structure supply chains subject to fill rate constraints, but in making the stock positioning decisions, approximations in line with Graves (1985) and Svoronos and Zipkin (1991) are utilized. For other recent work on evaluation and/or optimization of stock positioning in distribution systems for service parts, we refer to Caglar, Li, and Simchi-levi (2004) and Caggiano, Jackson, Muckstadt, and Rappold (2007).

The stock positioning problem in clinical trial supply chains represents a new variation of the classical multi-echelon inventory models because it differs from the literature in a key aspect; specifically, system performance concerns are only relevant until an adequate number of patients are recruited. In addition, clinical trial supply chain cannot afford to reject patients due to supply shortages and therefore have to place enough stock in the system to satisfy all recruited patients up to the pre-determined limit on the number of subjects needed to complete the trial. This difference results in non-trivial modification of the objective and in different solution algorithms and new insights.

5.2. Two Special Cases

For this work, we assume the following timeline:



Figure 5.1: Timeline of Events

• t: Time of material shipment.

- t + l: Time of trial site opening after receiving shipped drug supply.
- λ : Rate of patient recruitment at each site (assuming Poisson arrivals).
- C(s): Per patient recruiting costs when number of sites is equal to s. Assume convex and increasing.
- ρ : Rate of production given in kits per unit time.
- ζ : The service level of not rejecting patients due to drug supply shortage.
- ζ': The service level of not backordering patients due to drug supply shortage,
 i.e., the immediate fill-rate of demand upon its arrival.
- n: Number of sites to open.
- S: Patient horizon or number of subjects needed to complete the trial.
- *I*: Site level inventory.
- T: Time at which patient recruitment is done.

In this section, we focus on sites with identical patient/demand arrival rates, lead times and cost structures. Throughout this paper, we assume independent Poisson arrivals of demand at each site. Demand cannot be transferred among different sites. Excessive drug inventory at different locations cannot be transferred.

Inventory at Sites Only.

In this case, we assume that all material is distributed once to each site without opportunity for replenishment. To guarantee fulfillment of all demand, the system has to set I = S at each site. The problem becomes,

$$\min_{n>0} \frac{S}{n\lambda} + \frac{nS}{\rho} \tag{5.1}$$

The above equation is reminiscent of an EOQ type of model with an even simpler result of $n_{direct}^* = \sqrt{\frac{\rho}{\lambda}}$ and $T_{direct}^* = \frac{2S}{\sqrt{\lambda\rho}}$.

Clearly, stocking enough drug supply at each site to never miss a potential demand is very costly. Alternatively, one can stock each site enough inventory to guarantee a certain service level (ζ). In clinical trials, it is common to have adequate safety stock to support service levels of upwards of 99%. This requirement is driven by the need to finish a trial as fast as possible. The required inventory level, I(n), at each site is given as follows:

$$I(n) = \min_{I>0} \Pr\{I \ge D_S\} \ge \zeta,$$

where D_S is the demand faced by one site until the system-wide total reaches S. Since all sites are identical, D_S follows a binomial distribution with parameters (S, 1/n). Thus,

$$I(n) = \min_{I>0} I \quad s.t. \quad \zeta \le \sum_{i=0}^{\lfloor I \rfloor} {\binom{S}{i}} \left(\frac{1}{n}\right)^i \left(1 - \frac{1}{n}\right)^{S-i}.$$
(5.2)

Given that the overall patient arrival to the trial is a Poisson process with rate $n\lambda$, the waiting time until the S^{th} arrival, assuming no stockouts, would follow an Erlang distribution with shape parameter, S, and rate parameter, $n\lambda$. Thus, the expected time to complete recruitment is simply $\frac{S}{n\lambda}$. Our mathematical program for minimizing the total production and recruitment completion time is therefore:

$$\min_{n>0} \frac{S}{n\lambda} + \frac{nI(n)}{\rho} \tag{5.3}$$

Note Eq. (5.3) is an approximation because a site can stockout, and thus some demand may not be satisfied. Given that some patients may be rejected due to drug shortage, it may take longer to get S patients recruited than indicated by the above formula. However, if ζ is very close to 1, as would be expected for clinical trials, this approximation can be very accurate. To yield tractable results, we approximate the discrete distribution of Eq. (5.2) using a normal approximation. This is justified given patient horizons in the hundreds. For any one site, the number of patients is the sum of S identical Bernoulli trials with success probability $\frac{1}{n}$. Thus,

$$I(n) = \Phi^{-1}(\zeta) \quad where \quad \Phi \sim N\left(\frac{S}{n}, \frac{S}{n}\left(1 - \frac{1}{n}\right)\right)$$
(5.4)

For any given ζ , we can now numerically approximate I(n). For example, if $\zeta = 0.99$, then:

$$I(n) = \frac{S}{n} + 2.32635 \sqrt{\frac{S}{n} \left(1 - \frac{1}{n}\right)}$$
(5.5)

where z = 2.32625 is the z-score associated with service level ζ .

The minimization problem (5.3) now becomes:

$$\min_{n>0} \frac{S}{n\lambda} + \frac{n}{\rho} \left(\frac{S}{n} + z \sqrt{\frac{S}{n} \left(1 - \frac{1}{n} \right)} \right) = \min_{n>0} \frac{S}{n\lambda} + \frac{S}{\rho} + \frac{z \sqrt{S(n-1)}}{\rho}.$$
 (5.6)

For simplicity, assuming that $n \ge 1$ is a continuous variable. Then the first derivative is $\frac{S}{\lambda}\left(-\frac{1}{n^2} + \frac{z\lambda}{2\rho\sqrt{S}}\frac{1}{\sqrt{n-1}}\right)$, which leads to the first order condition as,

$$\frac{S}{n^2\lambda} = \frac{z\sqrt{S}}{2\rho} \frac{1}{\sqrt{n-1}}$$

Let $a = \left(\frac{2\rho\sqrt{S}}{z\lambda}\right)^2$, we have $n^4 - a(n-1) = 0.$

It is easy to see that we have two cases: Case 1, there is no solution or one solution to the above equation. Then the first derivative must be non-negative for all n, so the optimal solution $n^* = 1$. Case 2, there are two solutions to the above equation. Then the first solution is a local maximum while the second is a local minimum. We can compare n = 1 and the second solution to identify the global minimum.

One Central Warehouse.

Now, we consider another special case with a central warehouse. In this case, we hold inventory at two locations: 1) at the clinical trial sites, and 2) at the warehouse. We assume that each clinical trial site will now initially stock lead-time demand plus lead-time safety stock. Each trial site will follow an (S - 1, S) policy or equivalently a continuous-time base-stock policy. The warehouse needs to carry enough stock so that no patient will be rejected from any site because the warehouse and the site runs out of stock (i.e., $\zeta = 100\%$).

Given that demand at each site is Poisson, we calculate the required initial stock level at each site as follows:

$$I = \Phi^{-1}(\zeta') \quad where \quad \Phi \sim Poisson(\lambda l) \tag{5.7}$$

$$I = \min_{I>0} I \quad s.t. \quad \Pr\{I \ge D_l\} \ge \zeta',\tag{5.8}$$

where D_l is the lead time demand and ζ' is the backorder service level during lead time (not the lost-sales service level ζ). Due to Poisson distributed demand, the service level constraint can be written as $\sum_{i=0}^{\lfloor I \rfloor} \frac{e^{-\lambda l} (\lambda l)^i}{i!} \geq \zeta'$.

For tractability, we approximate the discrete Poisson distribution using a normal approximation and we now have that:

$$I = \Phi^{-1}(\zeta') = \lambda l + z' \sqrt{\lambda l} \quad where \quad \Phi \sim N(\lambda l, \lambda l), \qquad (5.9)$$

where z' is the z-score of ζ' . For simplicity, let $\theta = \lambda l + z' \sqrt{\lambda l}$.

Analogous to Eq. (5.6), we seek to minimize the sum of stock production time and patient recruitment time (100% ζ service level at warehouse):

$$\min_{n>0} \frac{S}{n\lambda} + \frac{n}{\rho}\theta + \frac{S-\theta}{\rho} = \min_{n>0} \frac{S}{n\lambda} + \frac{n-1}{\rho}\theta + \frac{S}{\rho},$$
(5.10)

where the warehouse stock is chosen to fulfill each of the first S demand from sites. In this case, clearly, it is S - I. Because when the warehouse runs out of stock, each site still has I units inventory and there are I patients remain to be served.

Taking the derivative and setting it equal to zero yields (2nd derivative is always positive):

$$n_{warehouse}^* = \sqrt{\frac{S\rho}{\lambda\theta}}.$$
(5.11)

The optimal expected completion time is

$$T_{warehouse}^{*} = \frac{S}{\rho} - \frac{\theta}{\rho} + 2\sqrt{\frac{S\theta}{\lambda\rho}}.$$
(5.12)

Comparing the first case (Eq. 5.1) and the second case (Eq. 5.10), clearly if $S > \theta$, then the first case requires more stock and takes longer time than the second case. However, the second case may have temporary stockout ($\zeta < 100\%$) while the first case has no stockout at all ($\zeta = 100\%$).

5.3. General Clinical Trial Supply Chains

The special cases in §5.2 represent extreme cases of possible clinical-trial drug supply chain configurations. In this section, we consider a supply chain of general distribution topology, pictured in Figure 5.2, where a central warehouse (CW), indexed as 0, supplies multiple regional warehouses (RWH), indexed by i = 1, 2, ..., and some sites directly, which are indexed by 0j where j = 1, 2, ... Each regional warehouse, i, supplies a set of sites indexed by ij where j = 1, 2, ... Keeping the assumption of independent Poisson demand and constant lead times as in §5.2 but relaxing the assumption of identical sites, our objective is to minimize the system-wide inventory investment subject to 100% fulfillment, ζ , and high immediate satisfaction (ζ'), by setting the stock levels at each location appropriately.

For large global studies, the topology pictured in Figure 5.2 closely resembles that of



Figure 5.2: Picture of Model's Supply Chain Topology

actual physical supply chains. Often, because of required regulatory clearance time and the time required to simply transport supplies over large distances, regional warehouses or depots are used to ensure more reliable and timely shipments to clinical trial sites. An article in the trade magazine *Applied Clinical Trials* (June 1, 2009) offered this rule of thumb:

In countries such as Canada, for example, depot delivery is unnecessary, as the domestic delivery system is similar to that in the United States and there are simplified customs procedures between the two countries. But in other countries, such as Argentina, Russia, China, and India, a depot is preferable because of sheer distance, and these countries require considerable time to clear materials. (Lis, Gourley, Wilson, and Page, June 1, 2009)

In addition to topology, we make the assumption of a continuous-review base-stock policy at both the sites and the regional warehouses. This too closely resembles the more modernized clinical trial supply chains. Through the use of integrated voice response systems (IVRS), all doctors administering clinical trial drugs to patients are required to call in to a system for instructions on which of the patient kits on site are to be administered to the patient. Through this system, both real-time inventory information is maintained and automated shipments to replenish site-level inventory are triggered (Byrom (2002)). While sometimes it will take multiple demands at a site to trigger an order, we make the equally realistic assumption that an order is triggered each time a demand occurs. Since clinical trial production is often associated with long lead times and is done in large batches, we assume the central warehouse does not have available resupply.

To represent the above inventory policy in our model, we use decision variables: s_0, s_i, s_{ij}, s_{0j} , where s_i, s_{ij}, s_{0j} are base-stock levels at locations i, ij and 0j respectively and s_0 is the initial stock at the central warehouse (which does not have the availability of additional replenishment). Since there is already inventory allocated to sites and regional warehouses downstream of the central warehouse, s_0 can be less than the required number of patient recruits, S. Planning for the worst casescenario of demand only arriving through one site and given s_i, s_{ij} and s_{0j} , we must set $s_0 = S - \min\{\min_i \{s_i + \min_j s_{ij}\}, \min_j s_{0j}\}$ to guarantee 100% recruitment supply (i.e., $\zeta = 100\%$). Given the above definitions, the problem of minimizing inventory investment subject to service level constraints (ζ') can be formulated as follows,

$$\min\{s_0 + \sum_i [s_i + \sum_j s_{ij}] + \sum_j s_{0j}\}$$
s.t. Fill rate at site $ij \ge \zeta'$, $\forall ij$,

$$\Pr\{s_{0j} \ge \text{lead time demand at site } 0j\} \ge \zeta', \quad \forall 0j, \quad (5.13)$$

$$s_0 = S - \min\{\min_i \{s_i + \min_j s_{ij}\}, \min_j s_{0j}\},$$

$$s_0, s_i, s_{ij}, s_{0j} \ge 0, \quad \forall i, j.$$

Clearly, this model differs from the classical multi-echelon model of inventory investment minimization subject to service level constraints due to the constraint on s_0 , which comes from the finite patient horizon. The constraint on s_0 effectively connects different regional warehouse systems and the problem cannot be decomposed.

5.3.1 Sub-models

To solve the problem in Eq. (5.13), we first make the following observations for two specific sub-models:

- **Observation 1:** Each direct ship distribution system, as shown in Figure 5.3, is an example of the second special case in §5.2 with the exception that the sites no longer need to have identical demands and lead times.
- **Observation 2:** The performance of each regional distribution system the subsystem of the regional warehouse i and the sites ij as shown in Figure 5.4, can be evaluated as a classical two-echelon distribution system in steady state. This holds because of the large number of patients required for the trial.

By Observation 2, we can write the fill rate constraints in Eq. (5.13) as follows,

$$\Pr\{s_{ij} - B_j^i \ge \text{lead time demand at site } ij\} \ge \zeta', \quad \forall ij,$$

where B_j^i is the steady-state backordered demand from site j at RWH i.

We first focus on the sub-model noted in Observation 1. If we choose to minimize inventory investment for this topology (Figure 5.3), the minimization of Eq. (5.13) is simplified to:

$$\min\{s_0 + \sum_j s_{0j}\}$$
s.t. $\Pr\{s_{0j} \ge \text{lead time demand at site } 0j\} \ge \zeta', \quad \forall 0j$

$$s_0 = S - \min_j s_{0j}$$

$$s_0, s_{0j} \ge 0, \quad \forall j.$$
(5.14)





Figure 5.4: Regional Warehouse Supply chain

Replacing s_0 in the objective function by its constraint yields,

Figure 5.3: Direct Shipment Supply Chain

$$\min\{S + \sum_{j} s_{0j} - \min_{j} s_{0j}\}$$

$$s.t. \Pr\{s_{0j} \ge \text{lead time demand at site } 0j\} \ge \zeta', \qquad \forall 0j \qquad (5.15)$$

$$s_{0}, s_{0j} \ge 0, \qquad \forall j.$$

Clearly, the optimal solution is to choose the smallest s_{0j} that achieves the desired fill rate at site 0j for all j.

Now, we look at the sub-model of Observation 2. In this sub-model, the objective is to minimize inventory investment for the topology shown in Figure 5.4. For this sub-model, the minimization of Eq. (5.13) is reduced to:

$$\min\{s_0 + s_i + \sum_j s_{ij}\}$$
s.t.
$$\Pr\{s_{ij} - B_j^i \ge \text{lead time demand at site } ij\} \ge \zeta', \qquad \forall j$$

$$s_0 = S - s_i - \min_j s_{ij}$$

$$s_0, s_i, s_{ij} \ge 0, \qquad \forall j.$$
(5.16)

To simplify further, we place the no lost demand constraint, $s_0 = S - s_i + \min_j s_{ij}$, into the objective function, then we have an equivalent problem:

$$\min\{S + \sum_{j} s_{ij} - \min_{j} s_{ij}\}$$

s.t. $\Pr\{s_{ij} - B_{j}^{i} \ge \text{lead time demand at site } ij\} \ge \zeta', \qquad \forall j \neq \tilde{j} \quad (5.17)$
 $s_{0}, s_{i}, s_{ij} \ge 0, \qquad \forall j.$

From Eq. (5.17), the optimal inventory policy is revealed. For both identical and nonidentical sites, the objective function reduces to minimizing site level inventory. This result is summarized in the following proposition.

Proposition 3 The solution to the following problem,

$$\min\{S + \sum_{j} s_{ij}\}$$
s.t. $\Pr\{s_{ij} - B_j^i \ge lead \ time \ demand \ at \ site \ ij\} \ge \zeta', \qquad \forall j \neq \tilde{j} \quad (5.18)$

$$s_0, s_i, s_{ij} \ge 0, \qquad \forall j,$$

is also the solution to Problem (5.17).

Proof. We first note that given s_i , s_{ij} are uniquely determined by the service levels in both Problem (5.17) and (5.18). In addition, s_{ij} is decreasing in s_i .

Denote the solution of Problem (5.18) to be \mathbf{s}' . Suppose the proposition is not true, then the solution of Problem (5.17), \mathbf{s}'' , must not minimize $\sum_j s_{ij}$ (subject to service level constraints). Thus \mathbf{s}'' must have a smaller s_i than \mathbf{s}' , and $s''_{ij} \ge s'_{ij}$ for all j. We discuss two cases:

- Case 1: argmin_j{s''_{ij}} = argmin_j{s'_{ij}}. Then ∑_j s''_{ij} min_j s''_{ij} ≥ ∑_j s'_{ij} min_j s'_{ij}. If > holds, then this creates a contradiction to the statement that s'' is the optimal solution to Problem (5.17). If = holds, then s' is also optimal in Problem (5.17) which contradicts to the assumption that this proposition is not true.
- Case 2: $\operatorname{argmin}_{j}\{s''_{ij}\} \neq \operatorname{argmin}_{j}\{s'_{ij}\}$. Let $\tilde{j} = \operatorname{argmin}_{j}s'_{ij}$, then $s''_{i\tilde{j}} > s'_{i\tilde{j}}$. Since $s''_{ij} \geq s'_{ij}$ for all j, thus $\sum_{j}s''_{ij} \min_{j}s''_{ij} > \sum_{j}s'_{ij} \min_{j}s'_{ij}$ which creates a contradiction.

Intuitively, Proposition 3 indicates that one should stock inventory at the regional warehouse i as much as possible to reduce the needed stock at sites. Effectively, one can increase s_i as long as s_0 is non-negative.

Expanding the one warehouse case to include multiple warehouses (but not direct shipment), Eq. (5.13) becomes:

$$\min\{s_0 + \sum_i [s_i + \sum_j s_{ij}]\}$$
s.t. $\Pr\{s_{ij} - B_j^i \ge \text{lead time demand at site } ij\} \ge \zeta', \quad \forall ij$

$$s_0 = S - \min_i \{s_i + \min_j s_{ij}\}$$

$$s_0, s_i, s_{ij} \ge 0, \quad \forall i, j.$$
(5.19)

Replacing s_0 by its constraint in the objective function, and defining $\tilde{i} = \operatorname{argmin}_i \{s_i + \min_j s_{ij}\}$ and $\tilde{j} = \operatorname{argmin}_j \{s_{\tilde{i}j}\}$, the above becomes:

$$\min\{S + \sum_{i \neq \tilde{i}} [s_i + \sum_j s_{ij}] + \sum_{j \neq \tilde{j}} s_{\tilde{i}j}\}$$

s.t. $\Pr\{s_{ij} - B_j^i \ge \text{lead time demand at site } ij\} \ge \zeta', \qquad \forall ij \qquad (5.20)$
 $s_0, s_i, s_{ij} \ge 0, \qquad \forall i, j.$

Inspired by this formulation, we design the following algorithm to identify heuristic solutions which yield lower and upper bounds for the optimal objective function.

Lower Bound Solution

- 1. For i = 1, 2, ..., set it to be \tilde{i} .
- 2. For $i \neq \tilde{i}$, minimize total inventory investment in the regional warehouse subsystem (Figure 5.4) subject to the fill rate constraints.
- 3. For $i = \tilde{i}$, minimize inventory investment at the sites and push inventory to warehouse \tilde{i} until either s_0 becomes negative or increasing s_i no longer leads to lower s_{ij} for any j.
- 4. Enumerate over all i = 1, 2, ... to find the one with minimum objective function.

This solution clearly provides a lower bound on the optimal objective function. This is true because if the condition $\tilde{i} = \operatorname{argmin}_i \{s_i + \min_j s_{ij}\}$ always holds, one cannot do better than this solution. However, the solution found may violate this condition. When this condition $\tilde{i} = \operatorname{argmin}_i \{s_i + \min_j s_{ij}\}$ holds in the solution found, then solution is also feasible and thus indeed optimal.

Slightly revising the lower bound solution results in an upper bound solution.

Upper Bound Solution

1. For i = 1, 2, ..., set it to be \tilde{i} .

- 2. For $i \neq \tilde{i}$, minimize total inventory investment in the regional warehouse subsystem (Figure 5.4) subject to the fill rate constraints.
- 3. For $i = \tilde{i}$, minimize inventory investment at the sites and push inventory to warehouse \tilde{i} as long as $\tilde{i} = \operatorname{argmin}_i \{s_i + \min_j s_{ij}\}$ holds until either s_0 becomes negative or increasing s_i no longer leads to lower s_{ij} for any j.
- 4. Enumerate over all i = 1, 2, ... to find the one with minimum objective function.

This solution may not be optimal but it is indeed feasible and thus provides a upper bound on the optimal objective function.

5.3.2 Minimizing Inventory Investment in a Clinical Trial Supply Chain

We now combine the results of our observations above into a completely generalized topology as pictured in Figure 5.2. We now rewrite Eq. (5.13) which governs the optimal inventory investment for the general system as follows:

$$\min\{s_0 + \sum_i [s_i + \sum_j s_{ij}] + \sum_j s_{0j}\}$$
s.t.
$$\Pr\{s_{ij} - B_j^i \ge \text{lead time demand at site } ij\} \ge \zeta', \qquad \forall ij$$

$$\Pr\{s_{0j} \ge \text{lead time demand at site } 0j\} \ge \zeta', \qquad \forall 0j$$

$$s_0 = S - \min\{\min_i \{s_i + \min_j s_{ij}\}, \min_j s_{0j}\}$$

$$s_0, s_i, s_{ij}, s_{0j} \ge 0, \qquad \forall i, j.$$

Suppose in the above, $\min\{\min_i\{s_i + \min_j s_{ij}\}, \min_j s_{0j}\} = \min_j s_{0j}$, then we can

reduce Eq. (5.13) to the following:

$$\min\{S - \min_{j} s_{0j} + \sum_{i} [s_{i} + \sum_{j} s_{ij}] + \sum_{j} s_{0j}\}$$
s.t. $\Pr\{s_{ij} - B_{j}^{i} \ge \text{lead time demand at site } ij\} \ge \zeta', \qquad \forall ij$
 $\Pr\{s_{0j} \ge \text{lead time demand at site } 0j\} \ge \zeta', \qquad \forall 0j$
 $s_{0}, s_{i}, s_{ij}, s_{0j} \ge 0, \qquad \forall i, j.$

$$(5.21)$$

In this case, we can separate the above problem into two minimization problems:

1. The minimization shown in Eq. (5.15):

$$\min\{\sum_{j} s_{0j} - \min_{j} s_{0j}\}$$

$$s.t. \Pr\{s_{0j} \ge \text{lead time demand at site } 0j\} \ge \zeta', \qquad (5.22)$$

2. A minimization problem of regional warehouse sub-systems:

$$\min\{\sum_{i} [s_i + \sum_{j} s_{ij}]\}$$
s.t. $\Pr\{s_{ij} - B_j^i \ge \text{lead time demand at site } ij\} \ge \zeta', \quad \forall ij$

$$s_i, s_{ij} \ge 0, \quad \forall i, j.$$
(5.23)

Note that this problem can be separated into multiple problems, one for each regional warehouse sub-system.

From the two separate minimization problems, our inventory strategy is revealed. For the direct shipment sites, the base stock policy at the sites is set at the minimum amount that is sufficient to satisfy the fill rate constraint. For the regional warehouse sites, solve each regional warehouse sub-systems to minimize its total inventory investment $s_i + \sum_j s_{ij}$. Alternatively, if the $\min\{\min_i\{s_i + \min_j s_{ij}\}, \min_j s_{0j}\} = \min_i\{s_i + \min_j s_{ij}\}$, then we simplify Eq. (5.13) as follows:

$$\min\{S - \min_{i}\{s_{i} + \min_{j} s_{ij}\} + \sum_{i}[s_{i} + \sum_{j} s_{ij}] + \sum_{j} s_{0j}\}$$

$$s.t. \Pr\{s_{ij} - B_{j}^{i} \ge \text{lead time demand at site } ij\} \ge \zeta', \qquad \forall ij \qquad (5.24)$$

$$\Pr\{s_{0j} \ge \text{lead time demand at site } 0j\} \ge \zeta', \qquad \forall 0j$$

$$s_{0}, s_{i}, s_{ij}, s_{0j} \ge 0, \qquad \forall i, j.$$

This equation separates into Eq. (5.20) and a minimization of site inventories for direct shipment. Intuitively, our strategy now is to do the following:

- For all regional warehouses except for \tilde{i} , we minimize inventory investment in the regional system (i.e. minimize $s_i + \sum_j s_{ij}$).
- For the regional warehouse, \tilde{i} , minimize inventory at the sites.
- For all direct shipment sites, indexed 0j, inventory is also minimized at the sites.

Assuming that S is greater than the minimum total inventory investment of any regional warehouse sub-system, we now summarize the inventory strategies discovered in this section by presenting the following algorithm for a general clinical trial supply chain.

Lower Bound Solution

- 1. Let $s_{0j} = \Phi_{0j}^{-1}(\zeta'), \forall j$, where $\Phi_{0j}(\cdot)$ is the cdf function of lead time demand at site 0j.
- 2. Solve s_i and s_{ij} by the following mathematical programming problem for all regional warehouses, *i*:

$$\min\{s_i + \sum_j s_{ij}\}$$

s.t. $\Pr\{s_{ij} - B_j^i \ge \text{lead time demand at site } ij\} \ge \zeta', \qquad \forall j \quad (5.25)$
 $s_i, s_{ij} \ge 0, \qquad \forall i, j.$

- 3. Suppose $\min_j s_{0j} = \min\{\min_i\{s_i + \min_j s_{ij}\}, \min_j s_{0j}\}$, then let s_i and s_{ij} be found by the mathematical programming problem (5.25).
- 4. Suppose $\min_{i} \{s_{i} + \min_{j} s_{ij}\} = \min \{\min_{i} \{s_{i} + \min_{j} s_{ij}\}, \min_{j} s_{0j}\}$. Set i = 1, 2, ...to be \tilde{i} , then
 - (a) For $i \neq \tilde{i}$, let s_i and s_{ij} be found by the mathematical programming problem (5.25).
 - (b) For $i = \tilde{i}$, minimize inventory investment at the sites and push inventory to warehouse \tilde{i} until either s_0 becomes negative or increasing s_i no longer leads to lower s_{ij} for any j.
 - (c) Enumerate over all feasible i = 1, 2, ... to find the one with minimum objective function:
 - Note on Feasibility: To tighten the lower bound, we eliminate certain regional warehouses that are unable to optimally satisfy the condition min_i{s_i + min_j s_{ij}} = min{min_i{s_i + min_j s_{ij}}, min_j s_{0j}}. Specifically, for each i = 1, 2, ..., we find the maximum possible s_i + min_j s_{ij} by increasing s_i until all s_{ij} achieve their minimal values assuming infinite supply is available from the warehouse. We then compare s_i + min_j s_{ij} to the calculated maximums. If s_i + min_j s_{ij} is not lower than all of the calculated maximums, the i regional warehouse subsystem cannot deviate from the solution to 5.25 for the network to achieve its optimum.
- 5. Compare the solution found in Steps 3 and 4 to find the minimum.

Clearly, this algorithm generates a solution that provides a lower bound on the optimal objective function. The solution is indeed optimal if $\min_j s_{0j} = \min\{\min_i\{s_i + \min_j s_{ij}\}, \min_j s_{0j}\}$ holds when Step 3 generates the solution or if $\min_i\{s_i + \min_j s_{ij}\} = \min\{\min_i\{s_i + \min_j s_{ij}\}, \min_j s_{0j}\}$ holds when Step 4 generates the solution. The algorithm runs in polynomial time with a complexity at most of O(Sn) where S is the patient horizon and n is the total number of locations in the system.

We can also construct a upper bound solution as follows.

Upper Bound Solution

- 1. Same as Lower Bound Solution.
- 2. Same as Lower Bound Solution.
- 3. Suppose $\min_j s_{0j} = \min\{\min_i\{s_i + \min_j s_{ij}\}, \min_j s_{0j}\}$, then let s_i and s_{ij} be found by the mathematical programming problem (5.25) with an additional constraint of $s_i + \min_j s_{ij} \ge \min_j s_{0j}$.
- 4. Suppose $\min_{i} \{s_{i} + \min_{j} s_{ij}\} = \min \{\min_{i} \{s_{i} + \min_{j} s_{ij}\}, \min_{j} s_{0j}\}$. Set i = 1, 2, ...to be \tilde{i} , then
 - (a) For $i \neq \tilde{i}$, let s_i and s_{ij} be found by the mathematical programming problem (5.25).
 - (b) For i = i, minimize inventory investment at the sites and push inventory to warehouse i as long as s_i+min_j s_{ij} = min{min_i{s_i+min_j s_{ij}}, min_j s_{0j}} holds until either s₀ becomes negative or increasing s_i no longer leads to lower s_{ij} for any j..
 - (c) Enumerate over all i = 1, 2, ... to find the one with minimum objective function.
- 5. Compare the solution found in Steps 3 and 4 to find the minimum.

This solution may not be optimal but is feasible and thus provides a upper bound on the optimal objective function. In the next section we test the effectiveness of this algorithm. In lieu of calculating exact distributions of net inventory levels in solving the mathematical programming problem (5.25), we employ Graves (1985)'s negative binomial approximation of backorders and lead time demand which has been shown to perform well in a variety of multi-echelon settings (Caggiano, Jackson, Muckstadt, and Rappold, 2009). In addition, Caggiano, Jackson, Muckstadt, and Rappold (2009) points out the computational simplicity of calculating fill rates using Graves algorithm as compared to using exact methods which require the convolution of multiple probability distributions.

5.3.3 Solving Mathematical Program (5.25) Using Graves Approximation

Our objective in mathematical programming problem (5.25) is to minimize inventory investment in the regional warehouse subsystem subject to site level fill rate constraints. In this section, we provide details on how to leverage the work of Graves (1985) to solve this problem in a computationally efficient manner. To that end, we define Q_{ij} to be the outstanding orders at site ij, Q_i to be the outstanding orders at regional warehouse i, and $B(s_i)$ to be the regional warehouse backorder level assuming s_i units are stocked at the regional warehouse. Note that any site-level outstanding orders will always be in-transit from the regional warehouse or backordered at the regional warehouse.

The key algorithmic insight of Graves (1985) is that the distribution of outstanding orders for a site (Q_{ij}) can be approximated using a negative binomial distribution (assuming independent Poisson demand at the sites, regional warehouses fill demand on a first-come first-serve basis, and deterministic leadtimes). This two-parameter distribution can then be used to ensure compliance with any fill rate constraint, mathematically represented as $\Pr{Q_{ij}(t) \leq s_{ij}} \geq \zeta'$. Recognizing that the distribution of outstanding orders for a site (Q_{ij}) will be a function of the distribution of backorders at the regional warehouse, we first characterize, through a recursive calculation, the warehouse level backorder distribution by its mean and variance:

$$E[B(s_i)] = E[B(s_i - 1)] - \Pr\{Q_i \ge s_i\}$$

$$Var[B(s_i)] = Var[B(s_i - 1)] - [E[B(s_i)] + E[B(s_i - 1)]] \cdot [1 - \Pr\{Q_i \ge s_i\}]$$
(5.26)

where starting values of the recursion at $s_i = 0$ are the first two moments of the sum of all site level demands (which are Poisson and thus sum to, λ_i , a Poisson distributed variable). Using (5.27), we now find approximations for the first two moments of Q_{ij} where we assume leadtimes from the warehouse to each of the sites, L_{ij} , are all equivalent (this assumption can be relaxed with some minor modifications):

$$E[Q_{ij}] = \frac{\lambda_{ij}}{\lambda_i} E[B(s_i)] + \lambda_{ij} L_{ij}$$

$$Var[Q_{ij}] = \frac{\lambda_{ij}}{\lambda_i}^2 Var[B(s_i)] + \frac{\lambda_i - \lambda_{ij}}{\lambda_i} \frac{\lambda_{ij}}{\lambda_i} E[B(s_i)] + \lambda_{ij} L_{ij}.$$
(5.27)

The key idea is to approximate Q_{ij} by a negative binomial random variable. Given the first two moments of Q_{ij} , we then solve for the corresponding shape and scale parameters of the target negative binomial distributed variable. From this point, we use the inverse cumulative distribution of a negative binomial variable to calculate sitelevel inventory stocking policies. Note that for each policy s_i , all site level policies must be recalculated.

The final step to find the minimal possible inventory in the system is to calculate for each possible warehouse stocking level the corresponding site-level inventories. For each potential policy, record the sum of warehouse plus site level inventories and pick the lowest value over all possibilities as the solution. For this last step, choose the smallest possible value of s_i to be zero as negative inventory is not possible and iterate by one extra unit of inventory in the warehouse. This iteration continues until either: 1) all calculated site level inventories are equal to their minimal quantity as calculated assuming infinite supply at the warehouse or 2) the warehouse stocking level is equal to S which because of the structure of the problem would guarantee enough supply at the regional warehouse to complete the trial. To break ties, we chose the solution with the largest s_i as this solution would present more flexibility should demand rates change.

5.4. Numerical Analysis

The objective of this section is two-fold: (1) To test the effectiveness of the solution algorithms (2) To study the effect of leadtimes and recruiting rates on optimal supply chain configurations.

5.4.1 Solution Algorithm Effectiveness

To test the effectiveness of our solution algorithm, we examine the scenario of a threecountry trial. In each of the three countries, we consider putting a regional depot or shipping direct. In each country the arrival rate of patients, λ , in that country is disaggregated among four sites such that $\lambda 1/\lambda = 0.1$, $\lambda 2/\lambda = 0.2$, $\lambda 3/\lambda = 0.3$, and $\lambda 4/\lambda = 0.4$ (similar to the test scenario of Graves (1985)). The patient arrival rate, λ , of the three countries will be 3, 5, and 7 patient arrivals per day respectively(i.e. one slow recruiting rate country like the United States, one medium recruiting rate country, and one fast recruiting rate country like India, Russia, or China). Lead time to the warehouse, L_i , may be 1, 3, 7, or 14 days with lead time from warehouse to site, L_{ij} being 1, 3, or 7 days, but always equal to or less than L_i (9 cases per country when using a warehouse). In examining direct shipment, we assume that the direct shipment lead-time to the warehouse is equal to the direct shipment lead time to the site. As such, direct shipment times may be 1, 3, 7, or 14 days (4 cases if direct shipment is used). In other words, we assume that the lead times of getting shipments from the central warehouse into a country's regional warehouse or a country's site is identical. In summary, each country may have one of 13 possible shipping/lead time configurations and thus, 2197 (i.e. 13^3) possible test cases are examined.

In summary, the results based on the above test problems are encouraging. The algorithm provides a feasible and optimal solution in 1,000 of the 2,197 scenarios or 45.5% of the cases. As summarized in Table 5.2, the average performance gap, i.e., the difference in objective function between upper bound and lower bound over lower bound objective function (which is the measure of potential percentage increase in the objective function over the lower bound) is 11.2%.

| # of Regional Warehouses | % Gap | | # of Optimal Solutions | # of Scenarios |
|--------------------------|-------|-------|------------------------|----------------|
| 0 | 0.0% | | 64 | 64 |
| 1 | 1.0% | | 360 | 432 |
| 2 | 4.7% | | 519 | 972 |
| 3 | 26.9% | | 57 | 729 |
| Average | 11.2% | Total | 1000 | 2197 |

Table 5.2: Algorithm Performance for Various Supply Chain Configurations

5.4.2 Optimal Supply Chain Configuration

Given the effectiveness of the algorithm in finding the optimal solution, we now leverage those solutions to develop insight into optimal configuration decisions. Specifically, we focus on the impact of recruitment rates and lead times on the supply chain configuration (i.e., using a warehouse or shipping directly). Using the same parameters as used in §5.4.1 we look at the 9 possible warehouse/site leadtime combinations that are possible for each country. Given our example of three countries, we have $9^3 = 729$ possibilities. In addition, for each country, we make a decision as to whether to use a regional warehouse in that country or to ship direct to the sites. Thus, for each country/warehouse leadtime/site leadtime combination, we compare eight (i.e. 2^3) possible warehouse versus direct combinations and choose the best configuration determined as the one having the lowest upper bound. The results are summarized in Table 5.3.

| | Scenarios Using Direct Shipping | | | | | |
|-------------------|---------------------------------|----------------|--|--|--|--|
| Country | # of Scenarios | % of Scenarios | | | | |
| Slow Recruiting | 303 | 41.6% | | | | |
| Medium Recruiting | 438 | 60.1% | | | | |
| Fast Recruiting | 473 | 64.9% | | | | |

Table 5.3: Out of 729 test scenarios, frequency of configurations prescribing direct shipment to sites for that country

From Table 5.3, we see that in the fastest recruiting rate country, there appears to be the smallest need for aggregating demand through a regional warehouse. Only in instances where the warehouse leadtime is high (i.e., 7 or 14) and the site level leadtime from the warehouse is very low (i.e. 1 day or in some rare instances 3 days) do we find it optimal to introduce a regional warehouse in the fastest recruiting rate country. More generally, we show the frequency of the optimal supply chain configurations for our test scenarios as shown in Table 5.4.

| Country1 | Country2 | Country3 | Frequency | % |
|-----------------|-----------------|-----------------|-----------|------|
| $\lambda_1 = 3$ | $\lambda_2 = 5$ | $\lambda_3 = 7$ | | |
| Direct | Direct | Direct | 198 | 27% |
| Warehouse | Warehouse | Warehouse | 156 | 21% |
| Warehouse | Direct | Direct | 145 | 20% |
| Warehouse | Warehouse | Direct | 80 | 11% |
| Direct | Warehouse | Direct | 50 | 7% |
| Direct | Direct | Warehouse | 50 | 7% |
| Warehouse | Direct | Warehouse | 45 | 6% |
| Direct | Warehouse | Warehouse | 5 | < 1% |
| | | Total | 729 | 100% |

Table 5.4: Frequency of Optimal Supply Chain Configurations

The impact of recruiting rate on choosing a supply chain configuration seems somewhat counter-intuitive. In the slow recruiting country, a warehouse is recommended for a majority of scenarios if we exclude cases where all countries are going to receive direct shipment. For the fastest recruiting country, the opposite is true; a warehouse is rarely recommended in the optimal configuration. In practice, it is often recommended to consider putting a regional warehouse in countries with long lead times and the reason these countries are even considered to be included in the trial in the first place is often because of their fast recruitment. So interestingly, in countries where we'd expect a warehouse to almost definitely be included as part of the optimal configuration, the numerical study indicates otherwise.

Lastly, we examine how lead times affect the optimal supply chain configuration. Table 5.5 shows us the frequency with which it is optimal to ship direct or to use a warehouse for various leadtimes to both regional warehouse and site. As we examine this table, it is worth remembering the assumption that when we ship direct, it is the lead time to the warehouse that is used for shipping time directly to the sites. Interestingly, it appears that regional warehouses are only to be used if they offer a shipping time advantage over the direct shipment case. For example, all cases where warehouse leadtime is 7 days and site leadtime is also 7 days should be handled via direct shipment regardless of the country's recruiting rate. Consistent to Table 5.3, Table 5.5 implies that a regional warehouse is more likely to be optimal when the lead time from regional warehouse to sites is much shorter than the lead time from the central warehouse to the regional warehouse.

| LT To Regional Whse | 1 | | 3 | | 7 | | | 14 | |
|----------------------|-----|-----|-----|----|-----|------|----|-----|------|
| LT From Whse to Site | 1 | 1 | 3 | 1 | 3 | 7 | 1 | 3 | 7 |
| Country1 | 56% | 0% | 90% | 0% | 54% | 95% | 0% | 0% | 79% |
| Country2 | 63% | 51% | 93% | 0% | 74% | 100% | 0% | 68% | 93% |
| Country3 | 85% | 49% | 95% | 0% | 81% | 100% | 0% | 73% | 100% |

Table 5.5: Frequency of Shipping Direct in an Optimized Supply Chain

5.5. Summary Remarks

In this chapter, we have illuminated the supply issues that spreading clinical trial demand over multiple sites can cause. Specifically, getting the right inventory, to the right place, at the right time becomes increasingly difficult as both countries and sites are added to the study. We have constructed a model for the general clinical trial supply chain that can optimally position inventory in support of the trial's service level requirements. To aid implementation of this model, we presented an heuristic algorithm that reduces the complexity of the exact model without greatly sacrificing the performance of the model. Using this algorithm and a representative example, we obtain some insights into configuring the supply chain.

Chapter 6 Concluding Remarks

In the introductory chapter of this dissertation, we have motivated the need for providing tactical decision support to managers of clinical trial supply chains. Through the analysis of SEC filings, we see the enormity of expense devoted to clinical trial supply chains and the opportunity which exists to create value for these companies. That opportunity, as is shown in Chapters 3-5, can be seized by tailoring supply chain models to the unique needs of clinical trial supply chains. Chapter 3 looks at the unique aspect of having a deterministic demand stream that may simply disappear and its effect on batch sizing and production frequency. Chapter 4 addresses the effects of learning during a trial and incorporates this learning into production planning decisions. Effectively, the model of Chapter 4 balances economies of scale against the benefits of more accurate forecasts. In Chapter 5, we examine how the presence of a definitive trial ending after enrolling the desired number of patients affects the optimal supply chain configuration and we use the model to prescribe optimal inventory policies for a given configuration.

Going forward beyond the scope of the dissertation, clinical trial supply chain research promises to be both fruitful to practitioners because of the need for tailored models and of interest to academicians due to the unique components of clinical trial supply chains. The huge potential in optimizing the drug supply chain has recently been recognized both in academia and in industry. While there is ample work to be done, we suggest the following specific research directions:

- 1. An Empirical Study: This dissertation provides some empirical evidence of the magnitude of spending needed to support clinical trial supply chains. A more comprehensive empirical study is needed to verify the financial significance and determine the impacting factors.
- 2. Multi-Product/Multi-echelon Optimization: Drug supply chains often consist of multiple manufacturing steps that are done in geographically dispersed facilities, e.g., the active pharmaceutical ingredients (API) manufacturing, formulation and packaging (Bernstein and Hamrell, 2000). Furthermore, companies may have multiple investigational drugs in clinical trials simultaneously. Thus, it is important to generalize the model of multi-echelon clinical trial supply chains to coordinate multiple drugs.
- 3. Outsourcing Contracts: While many large pharmaceutical companies produce investigational drugs in-house, most smaller companies outsource production to 3rd party manufacturers. Given the potential failure risk and the large costs of production, constructing efficient and fair outsourcing contracts is important to both clinical trial suppliers and pharmaceutical companies.
- 4. Applications of Models: While the models presented in this dissertation illustrate enormous opportunities to provide value for clinical trial supply chains, applying these models in the real-world practice would help refine and validate them.
- 5. *Stochastic Lead Times*: For all of the models presented, finding mechanisms to incorporate stochastic leadtimes would be a valuable way to expand the scope of their utility.
- 6. *Tightening bounds for the multi-echelon model*: While an approximately 11% performance gap of our algorithms to support clinical-trial inventory positioning decisions is not bad, ways to further tighten this gap should be explored.
- 7. Batch ordering policy in the multi-echelon model: In this thesis, we analyze the

clinical trial supply chain under the assumption of one-for-one ordering policies. Expanding these models to include batch ordering policies is one other area worthy of future study.
Bibliography

- Acusphere Inc. 2004–2007. Form 10-K for Fiscal Years 2003 2006. Retrieved November 8, 2007 from http://investor.acuspher.com/.
- AggarwalL, A., J.K. Park. 1993. Improved algorithms for economic lot size problems. Operations research 41(3) 549–571.
- Allos Therapeutics Inc. 1996–2007. Form 10-K for Fiscal Years 1995 2006.
- Ariad Pharmaceuticals Inc. 2004–2007. Form 10-K for Fiscal Years 2004 2006. Retrieved November 8, 2007 from http://www.araid.com/.
- Axsater, Sven. 1990. Simple solution procedures for a class of two-echelon inventory problems. *Operations Research* **38**(1) 64–69.
- Berenson, A. 2006. End of Drug Trial is a big loss for Pfizer. The New York Times.
- Bernstein, D.F., M.R. Hamrell. 1999. Improve Communication With The Clinical Supplies Department To Speed Clinical Studies. *Applied clinical trials* 8(9).
- Bernstein, D.F., M.R. Hamrell. 2000. Integrating drug supply issues with strategic preclinical and clinical development. *Drug Information Journal* **34**(3) 909–918.
- Bitran, Gabriel R., Elizabeth A. Haas, Hirofumi Matsuo. 1986. Production planning of style goods with high setup costs and forecast revisions. *Operations Research* **34**(2) 226–236.
- Blau, G.E., J.F. Pekny, V.A. Varma, P.R. Bunch. 2004. Managing a Portfolio of Interdependent New Product Candidates in the Pharmaceutical Industry. *Journal of Product Innovation Management* 21(4) 227–245.
- Brown, G.W., J.Y. Lu, R.J. Wolfson. 1964. Dynamic Modeling of Inventories Subject to Obsolescence. *Management Science* **11**(1) 51–63.
- Byrom, Bill. 2002. Using ivrs in clinical trial management. *Applied Clinical Trials* October.
- Caggiano, Kathryn E., Peter L. Jackson, John A. Muckstadt, James A. Rappold. 2007. Optimizing Service Parts Inventory in a Multiechelon, Multi-Item Supply Chain with

Time-Based Customer Service-Level Agreements. *OPERATIONS RESEARCH* **55**(2) 303–318. doi:10.1287/opre.1060.0345.

- Caggiano, Kathryn E., Peter L. Jackson, John A. Muckstadt, James A. Rappold. 2009. Efficient computation of time-based customer service levels in a multi-item, multiechelon supply chain: A practical approach for inventory optimization. *European Journal of Operational Research* 199(3) 744 – 749.
- Caglar, Deniz, Chung-Lun Li, David Simchi-levi. 2004. Two-echelon spare parts inventory system subject to a service constraint. *IIE Transactions* **36**(7) 655 – 666.
- Chow, S.C., M. Chang. 2006. Adaptive design methods in clinical trials. CRC Press.
- Chu, L.Y., V.N. Hsu, Z.J.M. Shen. 2005. An economic lot-sizing problem with perishable inventory and economies of scale costs: Approximation solutions and worst case analysis. *Naval Research Logistics* 52(6) 536–548.
- Clemento, A. 1999. New and integrated approaches to successful accelerated drug development. *Drug information journal* **33**(3) 699–710.
- Colvin, M., C.T. Maravelias. 2008. A stochastic programming approach for clinical trial planning in new drug development. *Computers and Chemical Engineering* **32**(11) 2626–2642.
- Cotherix Inc. 2006. Form 10-K for Fiscal Year Ended December 31, 2005. Retrieved November 8, 2007 from Lexis-Nexis Academic Database.
- David, I., E. Greenshtein, A. Mehrez. 1997. A dynamic-programming approach to continuous-review obsolescent inventory problems. Naval Research Logistics 44(8) 757-774.
- DiMasi, J. A. 2003. The price of innovation: new estimates of drug development costs. Journal of Health Economics 22(4–5) 469–479.
- DiMasi, J.A. 2001. Risks in new drug development: approval success rates for investigational drugs. *Clin Pharmacol Ther* **69**(5) 297–307.
- DiMasi, J.A., H.G. Grabowski. 2007. The cost of biopharmaceutical R&D: Is biotech different? *Managerial and Decision Economics* 28(4-5) 469–479.
- DiMasi, J.A., H.G. Grabowski, J. Vernon. 2004. R&D costs and returns by therapeutic category. *Drug Information Journal* **38**(3) 211–223.
- DiMasi, J.A., R.W. Hansen, H.G. Grabowski. 2003. The price of innovation: new estimates of drug development costs. *Journal of Health Economics* **22**(2) 151–185.

Dowlman, N., M. Lang, D. McEntegart, G. Nicholls, S. Bacon, J. Star. 2004. Optimizing

the Clinical Supply Chain Through Trial Simulation. Applied Clinical Trials 40–46.

- Drennan, K.B. 2002. Patient recruitment: the costly and growing bottleneck in drug development. Drug discovery today 7(3) 167–170.
- English, M., A. Ma. 2007. From Cradle to Grave: Tracking the life cycle of a drug is a complex process that early planning and IVR technology can help ease. *Applied Clinical Trials* **16**(8) 50–54.
- Eppen, Gary D., Ananth. V. Iyer. 1997. Improved fashion buying with bayesian updates. Operations Research 45(6) 805–819.
- Eppen, G.D., FJ Gould, B.P. Pashigian. 1969. Extensions of the planning horizon theorem in the dynamic lot size model. *Management Science* **15** 268–277.
- Federgruen, A., M. Tzur. 1991. A Simple Forward Algorithm to Solve General Dynamic Lot Sizing Models with n Periods in 0 (n \log n) or 0 (n) Time. *Management Science* 37 909–925.
- Fisher, Marshall, Kumar Rajaram, Ananth Raman. 2001. Optimizing inventory replenishment of retail fashion products. *Manufacturing & Service Operations Management* 3(3) 230–241.
- Fisher, Marshall, Ananth Raman. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research* **44**(1) 87–99.
- Florian, M., JK Lenstra, AHG Rinnooy Kan. 1980. Deterministic Production Planning: Algorithms and Complexity. *Management Science* 26(7) 669–679.
- Gatica, G., LG Papageorgiou, N. Shah. 2003. Capacity planning under uncertainty for the pharmaceutical industry. *Chemical Engineering Research and Design* **81**(6) 665–678.
- George, Lesley. 2005. Investigational medicinal products optimizing the supply chain. Applied Clinical Trials 14(4) 42 – 48.
- Getz, K.A., A. de Bruin. 2000. Breaking the development speed barrier: Assessing successful practices of the fastest drug development companies. Drug Information Journal 34 p725 – 736.
- Girotra, K., C. Terwiesch, K.T. Ulrich. 2007. Valuing R&D Projects in a Portfolio: Evidence from the Pharmaceutical Industry. *Management Science* **53**(9) 1452–1466.
- Glickman, Seth W., John G. McHutchison, Eric D. Peterson, Charles B. Cairns, Robert A. Harrington, Robert M. Califf, Kevin A. Schulman. 2009. Ethical and scientific implications of the globalization of clinical research. N Engl J Med 360(8) 816–823. doi:10.1056/NEJMsb0803929.

- Graves, Stephen C. 1985. A multi-echelon inventory model for a repairable item with one-for-one replenishment. *Management Science* 31(10) 1247–1256.
- Hall, Catherine. 2008. Clinical supplies on the front lines: Adapting to trial demand. *Clinical Supply Management for Pharmaceuticals Conference*. Pfizer Global Research and Development, Philadelphia, PA.
- Iglehart, D.L., S. Karlin. 1962. Optimal policy for dynamic inventory process with nonstationary stochastic demands. *Studies in Applied Probability and Management Science* 127–147.
- Johnson, GD, HE Thompson. 1975. Optimality of Myopic Inventory Policies for Certain Dependent Demand Processes. *Management Science* **21**(11) 1303–1307.
- Katok, E., A. Lathrop, W. Tarantino, S.H. Xu. 2001. Jeppesen Uses a Dynamic-Programming-Based DSS to Manage Inventory. *Interfaces* **31**(6) 54–65.
- Katok, E., S.H. Xu. 2001. Managing Inventory in the Information Age: a system with product life-cycle uncertainty and stochastic time-varying demand. Tech. rep., Working Paper. Smeal College of Business Administration, The Pennsylvania State University. University Park, PA.
- Kornish, Laura J., Ralph L. Keeney. 2008. Repeated Commit-or-Defer Decisions with a Deadline: The Influenza Vaccine Composition. Operations Research 56(3) 527–541. doi:10.1287/opre.1070.0436.
- Krams, Michael, Kennedy R. Lees, Werner Hacke, Andrew P. Grieve, Jean-Marc Orgogozo, Gary A. Ford. 2003. Acute Stroke Therapy by Inhibition of Neutrophils (ASTIN): An Adaptive Dose-Response Study of UK-279,276 in Acute Ischemic Stroke. *Stroke* 34(11) 2543–2548. doi:10.1161/01.STR.0000092527.33910.89.
- Krishnan, V., KT Ulrich. 2001. Product development decisions: a review of the literature. Management Science 47(1) 1–21.
- Kumar, RS. 2008. Managing the clinical trial supply chain. Conference presentation. Clinical Trials Supply Management for Pharmaceuticals conference. Philadelphia, PA. June 25, 2008.
- Lau, H-S., A. H-L. Lau. 1997. A semi-analytical solution for a newsboy problem with mid-period replenishment. The Journal of the Operational Research Society 48(12) 1245–1253.
- Le Chevalier, T, D Brisgand, JY Douillard, JL Pujol, V Alberola, A Monnier, A Riviere, P Lianes, P Chomy, S Cigolari. 1994. Randomized study of vinorelbine and cisplatin versus vindesine and cisplatin versus vinorelbine alone in advanced non-small-cell lung cancer: results of a European multicenter trial including 612 patients. J Clin Oncol 12(2) 360–367.

- Lesko, L. J. 2007. Paving the critical path: How can clinical pharmacology help achieve the vision? *Clinical Pharmacology & Therapeutics* **81** 170–177.
- Levis, A.A., L.G. Papageorgiou. 2004. A hierarchical solution approach for multi-site capacity planning under uncertainty in the pharmaceutical industry. *Computers and Chemical Engineering* 28(5) 707–725.
- Li, Jian, Suresh Chand, Maqbool Dada, Shailendra Mehta. 2009. Managing Inventory Over a Short Season: Models with Two Procurement Opportunities. MAN-UFACTURING SERVICE OPERATIONS MANAGEMENT 11(1) 174–184. doi: 10.1287/msom.1070.0204.
- Lis, Frank, Phillip Gourley, Phillip Wilson, Martin Page. June 1, 2009. Global supply chain management. *Applied Clinical Trials Online*.
- Lowe, D. 2006. What you need to know about Adaptive Trials. *Pharmaceutical Executive*, *July*.
- Maravelias, C.T., I.E. Grossmann. 2004. Optimal resource investment and scheduling of tests for new product development. *Computers and Chemical Engineering* 28(6-7) 1021–1038.
- McEntegart, Damian, Barry O'Gorman. 2005. The Impact on Supply Logistics of Different Randomization and Medication Management Strategies Using Interactive Vocie Response Systems. *Pharmaceutical Engineering* September-October 36–46.
- Milner, Joseph M., Meir J. Rosenblatt. 2002. Flexible supply contracts for short lifecycle goods: The buyer's perspective. Naval Research Logistics 49(1) 25–45.
- Muckstadt, JA. 2005. Analysis and algorithms for service parts supply chains. Springer Verlag.
- Murray, Jr., George R., Edward A. Silver. 1966. A bayesian analysis of the style goods inventory problem. *Management Science* 12(11) 785–797.
- NeoRX, NeoRx. 10-Q Filed on May 9, 2000. Clinical manufacture and supply agreement. Poniard Pharmeceuticals (formerly NeoRX Inc.) 10-Q.
- Neuer, A. 2008. Study shows need for clinical trial supply chain improvement. *Bio-IT* World March 3.
- Papageorgiou, Lazaros G., Guillermo E. Rotstein, Nilay Shah. 2001. Strategic Supply Chain Optimization for the Pharmaceutical Industries. *Industrial & Engineering Chemistry Research* **40**(1).
- Parlar, Mahmut, Z. Kevin Weng. 1997. Designing a firm's coordinated manufacturing and supply decisions with short product life cycles. *Management Science* **43**(10)

1329 - 1344.

- Peck, C.C. 1997. Drug development: improving the process. Food and Drug Law Journal 52(2) 163–167.
- Peterson, Magnus, Bill Byrom, Nikki Dowlman, Damian McEntegart. 2004. Optimizing clinical trial supply requirements: simulation of computer-controlled supply chain management. *Clinical Trials* 1(4) 399–412. doi:10.1191/1740774504cn037oa.
- Pharmaceutical Research and Manufacturers of America. 2009. Pharmaceutical industry profile 2009. PhRMA, Washington, D.C.
- Pierskalla, W.P. 1969. An inventory problem with obsolescence. Naval Research Logistics Quarterly 16(2) 217–228.
- Pisano, G.P. 1997. The development factory: unlocking the potential of process innovation. Harvard Business School Pr.
- Porteus, E.L. 2002. Foundations of stochastic inventory theory. Stanford Business Books.
- Rogers, M.J., A. Gupta, C.D. Maranas. 2002. Real options based analysis of optimal pharmaceutical research and development portfolios. *Ind. Eng. Chem. Res* 41(25) 6607–6620.
- Rowland, C. 2004. Clinical trials seen shifting overseas. International Journal of Health Services **34**(3) p555 556.
- Sethi, S.P., F. Cheng. 1997. Optimality of (s, S) policies in inventory models with Markovian demand. Operations Research 45 931–939.
- Shah, N. 2004. Pharmaceutical supply chains: key issues and strategies for optimisation. Computers and Chemical Engineering 28(6-7) 929–941.
- Shaw, D.X., L.P.M. Wagelmans. 1998. An Algorithm for Single-Item Capacitated Economic Lot Sizing with Piecewise Linear Production Costs and General Holding Costs. *Management Science* 44(6) 831–838.
- Sherbrooke, Craig C. 1968. Metric: A multi-echelon technique for recoverable item control. *Operations Research* **16**(1) 122–141.
- Simchi-Levi, D., Y. Zhao. 2005. Safety Stock Positioning in Supply Chains with Stochastic Lead Times. Manufacturing & Service Operations Management 7(4) 295–318.
- Simchi-Levi, D., Y. Zhao. 2006. Three Generic Methods for Evaluating Stochastic Multi-Echelon Inventory Systems. Tech. rep., Working Paper. Rutgers University, Newark, NJ.

- Song, J.S., P. Zipkin. 1993. Inventory control in a fluctuating demand environment. *Operations Research* 41(2) 351–370.
- Song, J.S., P.H. Zipkin. 1996. Managing Inventory with the Prospect of Obsolescence. Operations Research 44(1) 215–222.
- Svoronos, A., P. Zipkin. 1991. Evaluation of one-for-one replenishment policies for multiechelon inventory systems. *Management Science* 37(1) 68–83.
- Terwiesch, Christian, Roger E. Bohn. 2001. Learning and process improvement during production ramp-up. International Journal of Production Economics 70(1) 1 19.
- Thiers, F.A., A.J. Sinskey, Berndt E.R. 2008. Trends in the globalization of clinical trials. *Nature Reviews Drug Discovery* **7** p13 15.
- Thomson CenterWatch. 2007. State of the clinical trials industry 2007. Http://www.centerwatch.com/.
- Tomas, G., M. Hult. 2003. A research agenda for the nexus of product development and supply chain management processes. *Journal of Product Innovation Management* 20(6) 427–428.
- Veinott Jr, AF. 1963. Unpublished Class Notes. Program in Operations Research, Stanford University, Stanford, California.
- Wagelmans, A., S. Van Hoesel, A. Kolen. 1992. Economic lot sizing: An O (n log n) algorithm that runs in linear time in the Wagner-Whitin case. *Operations Research* 40 S145–S156.
- Wagner, H.M., T.M. Whitin. 1958. Dynamic Version of the Economic Lot Size Model. Management Science 5 89–96.
- Weng, Z. Kevin. 2004. Coordinating order quantities between the manufacturer and the buyer: A generalized newsvendor model. *European Journal of Operational Research* 156(1) 148 – 161.
- Wright, T. 1936. Factors affecting the cost of airplanes. Journal of Aeronautical Sciences 3(4) 122–128.
- Yelle, Louis E. 1979. The learning curve: Historical review and comprehensive survey. *Decision Sciences* **10**(2) p302 – 328.
- Zangwill, W.I. 1969. A Backlogging Model and a Multi-Echelon Model of a Dynamic Economic Lot Size Production System–A Network Approach. *Management Science* 15(9) 506–527.
- Zipkin, P.H. 2000. Foundations of inventory management.

Vita

Adam Fleischhacker

| 1974 | Born March 6 in New York, NY |
|-----------|---|
| 1992 | Graduated from Clarkstown South Senior High School. |
| 1996 | B.S. in Mechanical Engineering, State University of New York at Buffalo. |
| 1998 | M.B.A., William E. Simon Graduate School of Business - University of Rochester. |
| 1998-2000 | Employed by i2 Technologies, Parsippany, NJ, as Product Manager. |
| 2000-2001 | Employed by TransportEdge Inc., Parsippany, NJ, as Director of Product Management. |
| 2001-2003 | Employed by World Fuel Services Inc., Edison, NJ, as Director of Front Office Solutions |
| 2004-2005 | Founded Category Captain Services. Self-employed. |
| 2005-2009 | Graduate work in Supply Chain Management, Rutgers University, Newark, NJ. |
| 2005-2009 | Teaching Assistantship, Dept. of Supply Chain Management |
| 2009 | Ph.D. in Management, Rutgers University, Newark, NJ. |
| | |